Kölner Zeitschrift für Soziologie und Sozialpsychologie

Cross-National Comparative Research—Analytical Strategies, Results, and Explanations

Hans-Jürgen Andreß · Detlef Fetchenhauer · Heiner Meulemann

Abstract This introductory article reviews the history of cross-national comparative research, discusses its typical research designs and research questions, and ultimately summarizes the contributions to this special issue with respect to two questions: (i) What are the methodological challenges of cross-national comparative research today? (ii) What typical effects of the national context have been identified up to now?

Keywords Multilevel analysis · Mixed effects models · Cross-sectional and longitudinal designs · Causality · Context effects

International vergleichende Forschung – Analysestrategien, Ergebnisse und Erklärungen

Zusammenfassung In diesem einleitenden Artikel wird die Geschichte der ländervergleichenden Forschung dargestellt, es werden die typischen Forschungsdesigns und Forschungsfragen erörtert und schließlich die Beiträge dieses Sonderhefts in Bezug auf zwei Fragen zusammengefasst: (i) Was sind die methodologischen Herausforderungen der ländervergleichenden Forschung heute? (ii) Welche typischen Auswirkungen des nationalen Kontexts wurden bisher festgestellt?

H.-J. Andreß (✉) · D. Fetchenhauer · H. Meulemann
Institut für Soziologie und Sozialpsychologie, Universität zu Köln
Universitätsstr. 24, 50931 Cologne, Germany
E-Mail: hja@wiso.uni-koeln.de

D. Fetchenhauer
E-Mail: detlef.fetchenhauer@uni-koeln.de

H. Meulemann
E-Mail: meulemann@wiso.uni-koeln.de

 Springer

Schlüsselwörter Mehrebenenanalyse · Mixed-Effects-Modelle · Querschnitte und Längsschnitte · Kausalität · Kontexteffekte

1 Introduction

"We love you. But we need Sweden." This sign was shown by refugees arriving at the Danish border in September 2015. At that time, Denmark had sought to reduce the influx of refugees by issuing only temporary residence permits, delaying family reunification, and slashing benefits. These policies were publicized by the Danish government through an international advertising campaign, and hence made Denmark a far less attractive destination country for refugees than Sweden, a country which has, for instance, granted permanent residence to all Syrian asylum seekers since 2013 (The Local 2015). The refugees unwittingly relied on a country effect, namely different immigration and asylum polices in Denmark and Sweden, to ask the Danish border police to let them board trains to Copenhagen, from where they wanted to move on to Sweden.

Cross-national comparative research (CNCR) is concerned by and large with observing social phenomena across countries, and with developing explanations for their similarities and differences. Numerous scholars have previously elaborated on different aspects of CNCR: on research methods used in CNCR (Minkov 2013; Hantrais 2008; Landman 2017), on problems of survey methodology (Harkness et al. 2003, 2010; Johnson et al. 2018), on the operationalization of concepts across country contexts (Hoffmeyer-Zlotnik and Wolf 2011), or on statistical procedures and their applications in CNCR (Davidov et al. 2014, 2018). Adding to this body of knowledge, this special issue focuses on the use of CNCR to study the effects of national and sub-national contexts on behaviors and attitudes of individual actors. Moreover, it is of interest how behaviors and attitudes at the individual level lead to national and sub-national outcomes at the meso and macro levels. How do immigration policies affect migrants' wellbeing? Does the number of divorcees in a country influence individual divorce risks? Are human values universal, or do they vary from one country to another? Under which conditions is political protest triggered, and when does it lead to revolutionary changes within society? These and other questions are typical of CNCR analyses that seek to ascertain how upper-level (macro, meso) contexts influence micro-level phenomena, and how outcomes at the individual level are reflected at the meso and macro levels (as was summarized in Coleman's (1990) macro-micro-macro scheme).

This approach needs empirical information (data) for several countries and at different levels, plus a methodology that is able to deal with multilayered data of this nature: multilevel analysis. The term multilevel analysis is often used for a specific statistical modeling strategy (mixed effects regression; see Sect. 2.4). In this introductory article, we define it in a more general way, and we use it as a term for analyses comparing micro-level units (for instance individuals) across different upper-level (meso, macro) contexts (these could be countries). The analysis focuses on individual (perhaps: wellbeing) and upper-level (to take the example of revolutions) outcomes, which are explained by individual (e.g., educational) and

upper-level (for example political) characteristics. In so doing, this approach goes beyond the macro-comparative approaches within CNCR, given that the latter focus solely on macro-level relationships.

The special issue will summarize the state-of-the-art of multilevel analysis. It consists of four parts: (i) an overview of analytical strategies, selected results, and explanations in this introductory article, (ii) a theoretical part summarizing social science theories linking micro- and macro-level characteristics, as well as potential research designs in order to study the macro-micro-macro link, (iii) a methodological part reviewing data problems and statistical methods of analyzing multi-level data, and (iv) a substantive part reviewing results from CNCR in a variety of societal arenas: in the economic sphere, in politics, in civil society, and in cultural issues. All contributors have been invited to summarize the state-of-the-art of research on their topic. The contributions have been extensively reviewed by the editors and external reviewers in order to give them a similar outline and focus. The special issue is accompanied by a website (CNCR 2019) providing additional material that can be used both for searching our database of multilevel analyses, and for designing teaching methods and results of multilevel analysis.

This introduction will briefly review the history of CNCR, discuss its typical research designs and research questions (Sect. 2 and 3), and summarize the contributions to this special issue with respect to two questions: "What are the methodological challenges facing cross-national comparative research today?" (Sect. 4) and "What effects of the national context have been identified up to now?" (Sect. 5). Sect. 6 concludes with a few remarks on the standards, the practice, and the analytical strategy of CNCR as presented in the contributions to this special issue.

2 Cross-National Comparative Research: a Brief Historical Overview

2.1 Macro-Comparative Research

CNCR has a long research tradition. If one defines CNCR as research that compares at least two countries based on data from these countries, one finds innumerable research articles and books. In a literature review focusing on the second half of the 1980s, Bollen et al. (1993) found 209 non-edited books reviewed in *Contemporary Sociology* and 85 articles in the three major sociological journals (*American Journal of Sociology, American Sociological Review, Social Forces*), and in *Comparative Studies in Society and History*. The authors credit this impressive research output in a relatively small observation period (1985–1990) to "the collapse of communism in the former Soviet Union and its satellites, the trend towards democratization worldwide, the continually growing political and economic importance of the Pacific rim, and an increasing awareness of the interdependence of nations". All these factors "challenge sociologists to think about social change at the macro level" (Bollen et al. 1993). And they still do so today, even if these days we have become more worried about tendencies towards undermining democracy.

Bollen et al. (1993) were interested in *macro*-comparative research, and hence include in their comparison studies that "involve global, aggregate, or individual-

level structure or process." This special issue has a more specific focus. As mentioned above, it asks how (macro or meso) contexts affect behaviors and attitudes of individual and collective actors at lower levels. Individual actors could be citizens or employees, whilst collective actors could be organizations such as political parties or businesses. The interest in behavior and attitudes at lower levels is rooted in the methodological individualism of many social science theories, i.e., the belief that social phenomena can be traced back to the motivations and actions of individual agents, either acting on their own or representing larger collectivities such as families, clans, or organizations. The prototype of such an individualistic explanation is Coleman's (1990) macro-micro-macro scheme (Coleman's "boat" or "bathtub"). Moreover, since the paper by Robinson (1950), social scientists have known that global or aggregate information may be quite misleading when it comes to such individual-level interpretations. Relationships observed at the macro level (Robinson's ecological correlations) may obscure those at lower levels (Robinson's individual correlations). In order to avoid this ecological fallacy (Freedman 2004), individual-level information is needed below the macro and meso levels.

Having identified how the context influences behaviors and attitudes at lower levels (the macro-micro link), an equally important second step should follow, showing how upper- (meso-, macro-)level outcomes result from behaviors and attitudes at the lower level. Although not often undertaken, only this micro-macro link would complete the "bathtub." And both steps together would explain what can be seen at the macro level, e.g., why and under what conditions economic downturns lead to political protest and collective mobilization (Opp 2009).

2.2 The Individualistic Turn

This individualistic turn can also be observed in Kohn's (1987) Presidential Address to the American Sociological Association. In his talk about cross-national research as an analytic strategy, he identified several types of cross-national research: (i) where the nation is the object of study (in modern parlance: country case studies), (ii) where the nation is the unit of analysis (to establish relationships among characteristics of nations in a sample of countries), or (iii) where nations are treated as components of larger international systems (Kohn 1987). An example of the first type is Gauthier's (1996) comparative analysis of family policies in industrialized countries. The second type comprises quantitative macrosociological analyses such as Bornschier's and Chase-Dunn's (1985) analysis of transnational corporations and underdevelopment, or Alderson's and Nielsen's (2002) work on inequality trends in OECD countries. Finally, Wallerstein's (2011a–d) analyses of the capitalist world system are a typical example of the third type. Having said that, Kohn's talk primarily focused on a fourth type of cross-national research in which (iv) the nation is the context of study and the units of analysis are individual actors. Kohn and his collaborative work with other scientists on the effects of social structure on personality in the US, Poland, and Japan (Kohn 2015) represent this fourth type of inquiry perfectly. He furthermore mentioned several classical writings that fit into this fourth category, such as Inkeles's "Industrial Man" (1960), Lipset's "Democracy and Work-

ing Class Authoritarianism" (Lipset 1959), or Treiman's "Occupational prestige in comparative perspective" (1977), to name but a few examples.

This notwithstanding, the main interest of these analyses of individual data in different countries was to test the generalizability of findings and interpretations regarding individual actors found in particular contexts. In other words, it was research on individual- (micro-)level relationships, e.g., whether working class individuals are more liberal than middle class individuals on economic issues but illiberal on issues of civil liberties and civil rights (Lipset 1959), and whether this observation is true in different countries. We will refer below to these studies which compare several countries as in-depth *comparative case studies* (see also Grunow 2019).

Some time passed until social scientists again became interested in the concrete effects exerted by country contexts.[1] Broadly speaking, two types of context effects can be distinguished: taking Lipset's research question as an example, (i) the endorsement of liberal views, for instance on economic issues, can differ between countries on *average*, and (ii) the *association* between social class and liberal views can be weaker in some countries and stronger in others. To analyze these two new research questions, social scientists have to develop theories that explain what makes countries so different, thus enabling them to observe different averages and different associations, and they certainly need larger country samples in order to draw statistically sound conclusions about the context effects (see Sect. 4). We will refer to these studies, which compare large numbers of countries by using country-level variables, as *multi-country studies*.

Context effects can be explained by a variety of theories (for more details see Sect. 5), many of them focusing on institutions (March and Olsen 1989; Hall and Taylor 1996; Meyer et al. 1997). In this theory tradition, attitudes and behaviors of actors are assumed to depend on formal and informal rules and norms (i.e., on institutions). These may be (local, group-related) rules and norms in the immediate vicinity, or (global, national) rules and norms that affect society as a whole. The institutional framework at a given point in time is assumed to be a result of historical processes comprising earlier actions and decisions on the part of the actors. This framework determines the present incentive structure for the behaviors and attitudes of individuals. Other explanations for context effects refer to the role of social structure, i.e., the distribution of certain individual characteristics in a context, and the role of networks, i.e., the relationships between individual actors within a context (Blau and Schwartz 1984; Pescosolido 2007). For example, research on marriage disruption shows that it makes a difference whether a married person divorces in a country where divorce is virtually unheard of, or where divorce is a frequent occurrence (Stavrova 2019). Or the experience of unemployment has been found to be different in countries with large family networks as compared to more individualistic countries (Gallie and Paugam 2000). Because such context explanations require variables, country (context) names have to be replaced by the theorized country

---

[1] The analysis of context effects is not only prominent in CNCR, but also in regional science and urban sociology. The 2014 special issue of the *Kölner Zeitschrift für Soziologie und Soziologie* discusses predominantly local contexts such as urban districts or other lower-level regional units (Friedrichs and Nonnenmacher 2014).

(context) characteristics that are supposed to make countries different (Przeworski and Teune 1970).

2.3 New Data

The individualistic turn in combination with the interest in context effects has been fueled by the advent of large cross-national comparative survey projects and the application of specialized statistical methods to deal with such hierarchical data structures comprising individuals nested in countries.

The first cross-national comparative survey project was the European Values Study (EVS), initiated by the European Value Systems Study Group in the late 1970s. It published its first wave of surveys in 1981, covering a total of nine countries. Since then, three additional waves have been published in 1990, 1999, and 2008, with the latest wave covering no fewer than 47 European countries/regions, ranging from Iceland to Azerbaijan and from Portugal to Norway (EVS 2019). The World Values Survey (WVS) builds on the EVS. While the EVS is limited to European societies, and hence largely developed countries, the WVS takes a global perspective. Ronald Inglehart played a leading role in extending these surveys to be carried out in countries around the world. Today, after seven waves of surveys, the WVS covers more than 60 countries (see www.worldvaluessurvey.org). Another example is the International Social Survey Programme (ISSP) which evolved out of pre-existing general social surveys. The responsible survey institutes from four countries (the USA, the United Kingdom, Germany, and Australia) founded the ISSP in 1984, and agreed to develop topical modules together on important social science topics, which were added as fifteen-minute supplements to the national social surveys. The first topical module focusing on the role of government came out in 1985, and a new topical module (or a replication of a previous module) has been surveyed every year since then (ISSP 2019). The European counterpart to the ISSP is the European Social Survey (ESS), which in 2005 won the Descartes Prize for Research and Science Communication, Europe's most prestigious science award. The ESS became part of the European Research Infrastructure (European Research Infrastructure Consortium, ERIC) in 2013. The first wave of surveys was collected in 2002, and a total of eight waves covering more than twenty European countries have followed since that time (ESS 2019). Nowadays, besides these and other[2] academic projects, there are numerous cross-national comparative surveys conducted on behalf of political institutions such as the European Commission (e.g., the Eurobarometer), or Statistical Offices such as Eurostat (e.g., the European Union Labor Force Survey or the European Union Statistics on Income and Living Conditions), or the World Bank (the Living Standards Measurement Study). All these surveys are assumed to be comparative because they use identical instruments and sampling procedures in each participating country.

There have also been attempts to post-harmonize existing surveys from different countries. The most prominent example is perhaps the Cross-National Equivalent

---

[2] The Survey of Health, Ageing and Retirement in Europe (SHARE 2019) and the Generations and Gender Survey (GGS 2019) are two such examples.

Table 1 Examples of genuine and aggregated micro-, meso-, and macrodata. Authors compilation

Type of variable	Macrodata	Mesodata	Microdata
Genuine	Type of political regime (federal vs. unitary)	Centralization of sectoral wage bargaining	Personal political attitudes
Aggregated	Gross domestic product	Sectoral unemployment rate	Total personal income

File (CNEF) of panel studies from eight countries: the United Kingdom, Australia, South Korea, the USA, Russia, Switzerland, Canada, and Germany (Frick et al. 2007). Other examples are Blossfeld's Globalife and Edulife projects (see Blossfeld et al. 2019).

Not only individual data from different countries are needed: from the viewpoint of contextual analysis, data on pertinent country characteristics are necessary too. These characteristics can be genuine macro (or meso) characteristics, or they can be aggregated data from lower levels. Genuine macro (or meso) characteristics are sometimes also called global or primary data, while aggregated data are also referred to as derived data (Lazarsfeld and Menzel 1969). Table 1 provides some examples and compares them with genuine and aggregated microdata. Such context information, measured at country and regional levels, is provided by Statistical Offices, governmental agencies, non-profit organizations, and academic projects. The ESS website provides a comprehensive overview of existing context data in different areas such as demography and geography, economy, health, education, crime, political institutions, immigration, and various composite measures (Context 2019). It also includes links to providers of these data.

2.4 Analytical Strategies

When these macro- (or meso-)data are merged with individual (micro-)data, a hierarchical data structure emerges with individuals (in the most complicated form) nested in regions, years, and countries (see Schmidt-Catran et al. 2019). Such hierarchical data have been analyzed in different forms in multi-country studies:[3]

First, Analyses of Aggregate Data Many prominent studies, such as Norris and Inglehart's (2004), work on social differentiation and secularization, whilst others such as Richard Wilkinson's (2006) study on inequality and health remain almost entirely at the macro level, and compare national aggregates (means, proportions, correlations, regression coefficients) across countries, mostly along a descriptive approach. Whilst they are insightful, such analyses are at risk of committing the ecological fallacy. Moreover, compositional differences between the countries compared may get in the way of the comparisons. Similar to analyses of the gender pay gap, which are criticized for not controlling for differences in human capital and type of employment, macro-comparative cross-national research can be criticized for not

---

[3] Nonnenmacher and Friedrichs (2013) review 22 articles using at least one of these different forms of multi-country studies to explain life satisfaction.

controlling for the different age, sex, and employment structures of the countries. Composition bias naturally increases the more disparate the country sample is.

Second, Two-Step Analyses Other studies go one step further, and use country estimates of means, proportions, correlations, or regression coefficients as dependent variables in regression models with country characteristics as explanatory variables. Guerin et al. (2001), for example, analyze individual and contextual determinants of recycling behavior by first estimating the individual determinants using Eurobarometer data in each of fifteen European Union countries. Contextual determinants are assessed in a second step by regressing the country-specific regression constants on various country characteristics, among them an indicator of the ecological mobilization in each country which turns out to be the most important contextual determinant of recycling. Other analyses in this direction use proportions or means as dependent variables (e.g., Cohen 2004; Kaltenthaler and Anderson 2001). Of course, applying more confirmatory procedures, such as regression analysis, to analyze contextual effects raises the question of how to deal with varying sample sizes of the surveys on which the country estimates are based and how to incorporate the estimates' standard errors into these two-step procedures (Lewis and Linzer 2005).

Third, Analyses of Disaggregated Data A third approach simply disaggregates the contextual information to the lower level and treats these macro- (and meso-)data as if they were microdata. For example, Welch et al. use data from 2667 adult Catholics surveyed as part of the Notre Dame Study of Catholic Parish Life to test the "moral communities" hypothesis, which assumes that "individuals residing in parish communities with high levels of religiosity [are] predicted to be less likely to commit deviant acts than their counterparts who reside in parish communities with lower levels of religiosity" (Welch et al. 1991). The authors merge average levels of religiosity within each parish with individual-level measures of religiosity and deviant acts. Multiple regression models were estimated based on all individuals with non-missing data. However, in this approach, statistical tests of the context effects will be incorrect because disaggregation implies that tests of the context effects are based on the number of units at the lowest level (usually large numbers), and not on the much smaller numbers at the macro- (or meso-)level. Hence, p-values are much too low, and context effects are overly significant.

Fourth, Mixed Effects Analyses The fourth approach takes the hierarchical nature of the data into account, and estimates individual and contextual effects simultaneously. It recognizes at which level each variable is measured and uses the correct sample size for each level. It controls for possible composition effects by using micro-level variables as part of the set of explanatory variables. And finally, it takes into account all[4] unobserved macro-level characteristics that make lower-level units (e.g., individuals) more similar within higher-level units (e.g., countries) than between them. A typical research question is then "How much of the between-country

---

[4] To be precise: it takes account of all unobserved heterogeneity that is *uncorrelated* with the explanatory variables.

heterogeneity (which at the same time reflects the degree of similarity within counties) can be explained by country characteristics?" Individual and contextual effects are estimated simultaneously in this approach by treating the lower-level regression coefficients as random variables, which are modeled as functions of upper-level variables (for a more detailed description see Sect. 4). These models are known by different names: random-effects models, mixed-effects models, or simply multilevel models. However, if one defines multilevel analysis—as we did at the outset—as any analysis (i) of nationally representative individual-level data for several countries (or large subunits of countries), which (ii) seeks to explain outcomes at the individual level by country characteristics, then the term multilevel models is not very precise because approaches 2 and 3 fall into the same category. The technical term *mixed-effects models* is more precise because it correctly describes the statistical model in which each regression coefficient is assumed to be a function of observed variables and unobserved heterogeneity. The former are called fixed effects and the latter is captured by random effects.

Mixed-effects models were developed in the 1970s and 1980s, long before the advent of cross-national comparative survey data. Models for random coefficients or for clustered data were first published in econometric (Swamy 1970) and biometric journals (Goldstein 1986). Treating regression coefficients explicitly as dependent variables has a history in econometrics (Saxonhouse 1976, 1977), political science (Boyd and Iversen 1979), and educational research (Bryk and Raudenbush 2002). They have been routinely applied in social science research since the turn of the millennium. One of the first edited volumes, with all contributions consistently applying mixed-effects models, was Meulemann's (2002) collection of analyses of the first ESS wave, focusing on social capital and its perception in various European countries. A cursory look at some major social science journals shows numerous publications applying this methodology to a multitude of research questions. A content analysis of all (2001–2014) publications in seven major social science journals provides more than one hundred articles using mixed-effects models.[5] Not surprisingly, given the large number of countries in Europe and the availability of many different comparative surveys, most of the articles are published in the *European Sociological Review*. Looking at all *European Sociological Review* (ESR) volumes from 1985 to 2014, a keyword search using the term "multilevel" in the ESR online search engine provides 191 pertinent publications (Schmidt-Catran and Fairbrother 2016). According to Schmidt-Catran et al. (2019; Fig. 1), the proportion of ESR publications applying mixed-effects models reached almost 50% in 2016. Similar developments can be observed for other social science journals (Giesselmann and Schmidt-Catran 2018).

---

[5] Data are available on request from the first author. The following journals were analyzed: *American Sociological Review, European Sociological Review, International Journal of Sociology, American Journal of Political Science, European Journal of Political Research, Political Research Quarterly,* and *Social Science Research.*

3 Typical Research Questions and Research Designs

CNCR has been conducted in a vast number of different fields. This special issue thus contains articles summarizing research on such diverse topics as context-level effects on immigrants' labor market outcomes (Careja), employment and its institutional contexts (Erlinghagen), paid and unpaid work (Grunow), policy effects on political engagement (Ziller), party competition and vote choice (Spies and Franzmann), political systems and electoral behavior (Schmitt-Beck), families and their institutional contexts (Hank and Steinbach), conditions and consequences of unequal educational opportunities (Blossfeld et al.), media use in cross-national perspective (Boomgaarden and Song 2019), cross-national differences in predictors and correlates of subjective wellbeing (Stavrova), the welfare state and health (Pförtner et al. 2019), national religious context and individual-level effects of religiosity (Siegers 2019), and values in life domains in cross-national perspective (Halman and Gelissen 2019).

CNCR has mainly been conducted in most of these fields in order to answer specific research questions that stem from these different fields, rather than with the aim of testing the validity of general theories that could be applied to several or even all these different phenomena in mind. One example of the attempt to use several areas of sociological research to test one single theory is the work of Stavrova (2019). She argues that individuals' life satisfaction is higher the closer their attitudes and behavior match the society in which they live. She empirically confirms this hypothesis with regard to the life satisfaction of lone mothers, the unemployed, political orientations, cohabiting couples, or religion. Thus, she explores different life domains (such as "family," "economy," or "religion") in order to test a general theory. Such forms of research should be applied much more frequently.

Yet the majority of CNCR studies reviewed here test hypotheses in just one life domain. Although there is a huge variety of approaches and methodologies (see Goerres et al. 2019), most of them can basically be regarded as examples of four types of research design.

The first design refers to cases in which researchers explore the general validity of theories across different countries, cultures, and contexts. This is necessary and important because certain nations and cultures are heavily overrepresented in social science research. Henrich et al. (2010) have criticized the fact that modern psychology mainly studies "weird" people (Western, educated and from industrialized, rich and democratic countries), and overgeneralizes these findings. What is more, a high percentage of studies published in psychology are based on undergraduate university students. Although things might be better in sociology, a similar form of (American) ethnocentrism can be observed there as well. For example, part of the textbook knowledge in sociology refers to the "robust" finding that high levels of religiosity are related to a high level of life satisfaction. However, most of the studies on which this "knowledge" is based have been conducted in the USA, where there is an unusually high level of religiosity when compared to other industrialized countries. And indeed, the relationship between religiosity and life satisfaction is much weaker in most other industrialized countries (see Diener et al. 2011).

Results gathered in one society thus only gain credibility if they can be confirmed across different societies—and this effect is more pronounced the more diverse the countries that are compared with each other are. When applying such a research strategy, it is also possible (and necessary) to control for composition effects of the countries that are investigated (e.g., the distribution of the age or education of their inhabitants).

The second design deals with the question of how contexts influence actors' behavior and attitudes at the micro level (cross-level main effects). For example, Ziller (2019) reviews studies that investigate the influence of social policies on individuals' acceptance of welfare state programs. Another example is research on the influence of a countries' wealth on individuals' life satisfaction, focusing on the so-called Easterlin paradox (Easterlin 1974). There is a strong positive correlation at the bottom half of all countries (i.e., poor to about average) between aggregate wealth (i.e., the gross domestic product of a given country) and life satisfaction, but no such correlation is found amongst the rich countries of the world. It is, however, important to clearly distinguish between country-level and individual-level wealth. On an individual level, there is a positive correlation between income and life satisfaction in both poor and rich countries (Diener and Oishi 2000).

The third design investigates how contexts influence the micro-level effects of individual characteristics on actors' behavior and attitudes (cross-level interactions). For example, Stavrova (2019) summarizes research demonstrating that general attitudes towards work and unemployment (country-level moderator) influence the relationship between being unemployed (individual-level independent variable) and personal life satisfaction (individual-level dependent variable). Another example is a study by Just and Anderson (2012, see Ziller 2019) showing that immigration policies (country-level moderator) influence the relationship between citizenship status (individual-level independent variable) and civic participation (individual-level dependent variable).

Finally, the fourth design deals with the question of how the behavior and attitudes of actors at the micro level bring about certain characteristics at the macro level. It is interesting to note that this question is not very prominent in CNCR, as most of the dependent variables are either individual behaviors or attitudes, or are simple means of such individual measures. It would nonetheless be worthwhile to more systematically investigate potential feedback loops between macro-level variables and to show how they are mediated through the respective variables at the micro level (e.g., investigate how citizens' attitudes motivate political parties to adopt certain policies, which in turn influence citizens' attitudes). When conducting such analyses, one could also investigate potential moderating influences of institutional arrangements (e.g., systems of majority voting versus proportional representation systems). Meuleman et al. (2019) give some examples of context-level outcomes and their analysis using multilevel structural equation modeling (MSEM).

These four kinds of research designs can be investigated in a number of different ways (see Goerres et al. 2019). One analytical strategy, referred to above as comparative case study, involves analyses of different studies in an (often limited) number of countries that are frequently post-harmonized (see, e.g., the contribution of Blossfeld et al. 2019). If possible, however, it is preferable to use large-scale

survey programs such as the ISSP, the ESS, or the WVS, which conduct (mostly) identical surveys in many countries and carry out what we have called a multi-country study (for examples, see most substantive contributions in this special issue). As has already been mentioned, the progress that has been achieved by CNCR in recent decades would not have been possible without the existence of these survey programs.

These different kinds of research designs applied in CNCR partly resemble the logic of the famous bathtub model of sociological explanations by James Coleman (1990). In this model, correlations between two macro variables are explained by the influence that a macro variable 1 has on the "definition of the situation" by individual actors (micro variable 1), which determines individual actors' behavior (micro variable 2), which in turn determines macro variable 2 by simple or complex rules of aggregation. As an example, Coleman models Weber's theory about the Protestant work ethic along these lines. The prevalence of Protestantism in a given society (macro variable 1) leads actors to a belief in the sanctity of hard work and an ascetic lifestyle (micro variable 1; Coleman 1990, Chapter 1). This work ethic leads to certain behaviors (economic activities, working long hours, high rates of reinvestment of earned income; micro variable 2), which ultimately lead to an accelerated development of technology and productivity in a given society (macro variable 2).

On closer inspection, however, only few studies within the general framework of CNCR apply Coleman's bathtub model to its fullest extent. On the one hand, there is often no clear distinction between cognitive and behavioral variables on the micro level; and individual-level dependent variables often refer to attitudes rather than to behavior. On the other hand, the last step of the bathtub model (i.e., the link between micro variable 2 and macro variable 2) is seldom explicitly modeled or empirically investigated.

4 New Opportunities and Challenges of Cross-national Comparative Research

In the same way as the objects of sociology, that is societies, result from people's actions, sociology must deal with individuals as well. For this reason, multilevel analysis is a genuine sociological perspective. Until now, however, it has been used mostly as a cross-sectional research design. Yet societies change. In order to examine change, the analysis must be broadened by introducing a longitudinal design. Such a design, in turn, opens up new opportunities to ascertain causality, and poses the challenge of following up and explaining societal developments, in other words social change. How a longitudinal perspective in multilevel analysis may help to identify causality will be explained briefly, and what it can contribute to the analysis of social change will be elaborated upon more extensively.

4.1 Causality at the Macro Level

Cross-sectional CNCR describes correlations and therefore has two weaknesses. It cannot determine the direction of causality inherent in the correlations. And as every country is observed only once, CNCR does not control for time-constant unobserved heterogeneity, and this fact may bias correlations. Longitudinal CNCR can, however, overcome both weaknesses. It makes it possible to disentangle causal directions. And it controls for time-constant, unobserved heterogeneity.

As theories of social change contend an impact of one societal development on another, their examination of necessity requires one to take a longitudinal perspective. In the simplest case of two timepoints and two macro variables, which naturally vary over time, it constitutes a simple path model which provides coefficients for the stability of each of the two variables and, across the two variables, for the causal impact of each one on the other between the first and the second points in time (referred to as an autoregressive cross-lagged panel model). A comparison between the latter two coefficients therefore allows an assessment of the relative strengths of their causal impacts. As an extension, the effects of time-constant independent variables on both variables at the first point in time can be estimated such that unobserved heterogeneity is further reduced. Such a model can be applied to samples of any entity that is observed at least twice—be it persons, collective actors, or societies. And in each case, it can be analyzed with the same statistical technique, namely panel analysis (Andreß et al. 2013). How the causality that is hypothesized in theories of social change can be examined in a longitudinal multilevel analysis will be shown with a substantive example from modernization theory in the remaining paragraphs of this section.

4.2 Modernization Theory as a Common Denominator of Societal Developments

Since its start in the 19th century, sociology has regarded modernization as a scale of development. Modernization theory defines a set of societal developments and inserts them into a causal chain between driving forces and goals on the level of societies. It encompasses many societal tendencies: industrialization in terms of the increase in the percentage of manufacturing firms of gross domestic product per capita (GDPpc), urbanization in terms of the increase in the percentage of people living in cities, tertiarization in terms of the increase in the percentage of the labor force working in the service sector, educational expansion in terms of the increase in the percentage of the population holding a high school diploma, etc. For all of them, it postulates a common driving force, social differentiation, and a common goal, namely upward movement to a greater adaptive capacity of societies (Parsons 1964; Zapf 1994; Halman and Gelissen 2019). Secularization theory is a more specific example that is often seen as a further strand of modernization. It expects a decrease in religious belief and practice to occur as a consequence of social differentiation and cultural pluralization in societies (Meulemann 2017).

Extensive databases obtained from public censuses and administrative sources contain timeseries capturing these macro tendencies for many European countries, some of them from the 19th century up to today (e.g., Flora 1983 and 1986). Yet

they are restricted to demographic indicators of family and occupational statuses, the individual records of which cannot be recovered; more importantly, they do not record people's everyday actions and opinions. Only with the start of large-scale internationally comparative macro surveys in 1981 was it possible to compare tendencies between countries and follow up within them, that is for individual persons and for a broad range of attitudes and behaviors. And after replications up to today, they cover almost four decades. To exhaust their potential, the two-level methodology focused on in CNCR so far must be extended to three levels. What a cross-sectional, two-level analysis achieves, and how its achievements are surpassed by a longitudinal, three-level analysis, will be briefly outlined.

4.3 Two-Level Analysis: Controlling for Distributional Differences Between Countries

Theories of societal developments, such as modernization theory, propose that macro social properties should reflect countries' developmental stages. Secularization theory, for example, contends that advances in social differentiation decrease religiosity, that is, they cause secularization (Norris and Inglehart 2004). A causal hypothesis such as this can obviously only be examined when some antecedents are correlated with a particular outcome for many countries. However, such a macro correlation is subject to the ecological fallacy (see Sect. 2). Moreover, it cannot be understood as a macro process because it may have been produced by actors on the individual level. Individual-level variables are most often distributed differently between countries and can affect the development under scrutiny differently between countries. Furthermore, behind the correlation of the two macro variables, there is a multitude of further variables at work, both time-constant as well as time-varying. The unobserved heterogeneity referred to above can never be completely controlled for in cross-sectional terms but only when the same countries are observed repeatedly, that is longitudinally. Take again the example of secularization: country differences in religiosity depend on a myriad of country characteristics which can never be completely controlled for in cross-sectional designs. Yet following up one and the same country controls for all of its characteristics, be they its denominational composition, its legal regulation of relations between the state and the church, its representation of churches in party politics, or its cultural pluralization—all of which may be time constant or change over time. And as we later argue, all (observed and unobserved) time-constant country characteristics are easily controlled for by focusing on the over-time ("within") variance only.

Because different distributions as well as different effects of individual-level determinants may distort country-level causality, a *two-level analysis* which controls for compositional differences between countries and examines the equality of individual-level impacts is already required for the cross-sectional explanation of country differences as genuinely produced by country-level properties. These are the real targets of the macro analysis. In the process of further analysis, they must be traced back to different country-level variables, such as wealth or inequality, which in turn affect the macro goal variable.

In its simplest form, such a two-level analysis consists of two regression equations: Firstly, the micro-level dependent variable is regressed on one or more micro-level-independent variable(s) in the totality of all country samples. The intercept of this regression is the mean value in all countries; if one detects that it varies strongly between countries, it is worthwhile analyzing this variation as a random variable at the country level, depending on country characteristics. Secondly, therefore, this random intercept is regressed on one or more country property or properties. Just as the variance on the micro level will not be fully explained by the chosen individual-level predictors, the variance of the means on the country level will not be explained by the chosen country-level predictors, such that each equation will have its own error term. As the dependent variable of the second equation is the random intercept of the first equation (varying between countries), the second equation can be inserted into the first instead of the intercept. The resulting single-regression equation then contains micro- and country-level predictors and two error terms, one for the micro-level dependent variable, and one for its country means.

Let us take secularization theory as an example. Some secularization indicators—such as church attendance—should, according to secularization theory, be caused by social differentiation, indicated by for instance gross domestic product per capita (GDPpc). In order to prove this, church attendance must be regressed not only on GDPpc but also on individual-level determinants of church attendance, for example age. If older people attend church more often than younger people do, and if the mean age of countries increases with their advancing modernization, measured by their GDPpc, then age must be controlled for in order to ascertain at which stage of secularization the countries find themselves; without such controls, one would attribute the effects of different population distributions between countries to differences between countries in global characteristics (see Sect. 2). So far, the regression contains two error terms: for church attendance and mean church attendance.

In a more complicated form, a two-level analysis is extended by a third regression equation: the country-specific slope of an individual-level independent variable is regressed on one or more country properties. But, as a rule, the variance of the slopes will not also be fully explained. This more complicated form of two-level analysis therefore contains a third error term for the slopes. And if the individual-level regression equation contains more than one predictor, their slopes can be treated in the same manner. Let us take again secularization theory as an example. If the slopes—the effects of age upon church attendance—vary widely between countries and increase with their advancing modernization, then measured again by GDPpc, they must be regressed on the countries' advancement and a third error term for them must be introduced into the regression equation.

These analytical strategies can already be applied when a sufficient number of countries have been surveyed in a cross-sectional design at a specific point in time. As the contributions in this volume show, they give correct information about country differences and their—potentially causal—correlates, that is, they control for distributional differences between countries. But they do not tackle the question of causality head on. As both country-level and micro-level data are measured at the same points in time, the analysis remains cross-sectional. However, in studies

of societal developments, the most fundamental requirement to secure causality is measurements for at least two points in time.

4.4 Longitudinal Multilevel Analysis: Separating Within-Country from Between-Country Effects

As many cross-national surveys have been repeated since 1981, data which are *simultaneously cross-sectional and longitudinal* are available. The most important requirement to truly, that is causally, test developmental societal theories is therefore fulfilled. Time effects can then be explained in exactly the same manner as country differences by macro indicators, since the names of countries as well as points in time can be substituted by properties (Przeworski and Teune 1970). Moreover, by focusing on the over-time ("within") variance only, repeated surveys of the same countries can control for every time-constant property, and so far solve the problem of unobserved time-constant heterogeneity. However, such a test brings with it some challenges for methodological as well as substantive research (Schmidt-Catran et al. 2019; Hosoya et al. 2014; Meuleman et al. 2018).

First, it extends the analysis from two to three levels: persons, within country time points, within countries; if every country is surveyed at each point in time, it may also be specified as a cross-classified design—countries by timepoints—at the second level. Second, it requires a corresponding specification of the random part with three error terms (Meuleman et al. 2018, p. 189). Third, it requires a specific parametrization for time—either by time dummies, or by linear and higher-order functions of time, or by "societal growth curve models" (Hosoya et al. 2014; Meuleman et al. 2018). Fourth, and most importantly, it requires separating the cross-sectional comparison between countries from following up a development within countries where causality is at stake. The cross-sectional differences are estimated by the means of the predictor variables of each country over the points in time, the developments by the within-country differences between these means, and the time-specific values over all countries.

Let us take again the example of differentiation driving secularization. In a cross-sectional perspective, differences between countries on a scale of secularization at a given point in time are comparable to a photo finish of a race; they may reflect further or lower advances on a differentiation scale, just as the positions of the runners in a photo finish result from different training efforts and talents. In a longitudinal perspective, an advance on a scale of secularization within countries may result from an advance on a scale of differentiation within countries, that is, a correlation between a dependent time-varying variable and an independent time-varying variable, while controlling for time-constant variables—just as increased training efforts may grant a given runner a better position in the photo finish of the next race, while controlling for time-constant conditions such as genetic endowment. Thus, the effect of differentiation—measured by GDPpc for each point in time—on secularization—measured as church attendance for each point in time—can be split up into one effect that is due to differences between countries and another that is due to differences within countries over time (for details see Schmidt-Catran et al. 2019). Only the latter, namely the within-country differences, truly pertain to developmental

theories. Even in a longitudinal research design, failing to distinguish between the between-country and the within-country effects can lead to an overestimation of the developmental effect and to premature acceptance of the developmental theory.

The development of multilevel models with country-level and timepoint indicators is a major challenge for future methodological research. Its statistical complexities notwithstanding, it is needed in order to test substantive theories of societal developments, such as modernization and secularization theory, which until today have either been taken for granted or disputed on merely conceptual grounds. The advent of multilevel modeling and its extension over time opens up the possibility of subjecting such theories to stringent testing.

5 The Effects of Contexts

CNCR deals with the question of how behaviors and attitudes of citizens are formed by the contexts in which they live—which is best exemplified by nations. The nation is seen as a context in which citizens are embedded. But, most of the time, nations are entities that are remote from the lives of their citizens, and they differ in many ways. How they affect behaviors and attitudes must therefore be attributed to some analytical property which all nations share (Przeworski and Teune 1970), and which is sufficiently present in the lives of their citizens. To justify these properties and their reality in citizens' everyday lives is one of the main challenges of multilevel analysis. The substantive articles in this special issue implicitly suggested two steps to address this challenge.

5.1 Nations and Indicators

First, the specific domain of social life, that is, the pertinent behaviors and attitudes of the citizens to be regulated, must be identified. This is exemplified in this special issue for the labor market and employment opportunities by Erlinghagen and Careja, for the welfare system by Hank and Steinbach and Pförtner et al., for the family and family legislation by Grunow and Hank and Steinbach, and for the electoral and political party system and for political voting by Spies and Franzmann and Schmitt-Beck. Domain and behavior need not be always so close to each other as they are in these cases; they can be somewhat distant as well. Thus, for example, the fact of the welfare system providing social and personal security might reduce the need for religion (Schmidt-Catran et al. 2019). Yet in all the cases above, "domain" is understood as a complex of institutions, that is, rules for specific actions that are informally established or laid down in some form of legislation, which "by structuring opportunities and constraints, create expectations and incentives" (Schmitt-Beck 2019). This still leaves open the question of which opportunities and constraints are at work.

Second, therefore, the notion of a specific institutional context in a nation which directs the actors' actions and beliefs "in" the context must be specified by some measurable indicator. Ideally, therefore, such an indicator must indeed capture the orientation provided by the institution to its clientele; it cannot be an aggregate mea-

sure of individual-level properties, but must be a genuine, global characteristic of the societal sector (see Sect. 2). Let us take a few examples from this special issue. First, a higher level of the Gender Empowerment Measure (GEM), the Gender Development Index (GDI), or the Gender Inequality Index (GII) indicate legal regulations that are less or more incisive in order to handle conflicts between occupational and family careers; they may support person-level "agency" and facilitate women's labor force participation as well as men's housework involvement (Grunow 2019). Second, social welfare expenditure as a percentage of GDPpc indicates social security, which shields everybody against the risks of life (Norris and Inglehart 2004) and reduces the need to give them a religious explanation, at least for some—thus boosting secularization (Schmidt-Catran et al. 2019). Third, higher levels of the Index of Citizenship Rights for Immigrants (ICRI) and a low unemployment rate among natives indicate a "welcome culture," and may instigate the immigration and integration of new citizens (Careja 2019; Ziller 2019). Fourth, high values of the index of employment protection legislation (EPL) or the index of active labor market policy measures (LMP) indicate better opportunity structures and should increase personal employment (Erlinghagen 2019). Finally, there is a plethora of established indicators of the political system and of party competition which have been widely tested (and often confirmed) as positive or negative effects on voter turnout and voting (Schmitt-Beck 2019, Table 1; Spiess and Franzmann 2019, Table 1).

There is obviously no shortage of indicators of analytical properties of nations, and cross-national multilevel research has, by and large, successfully related them to sectors of a nation on the one hand, and to personal agency on the other. The nation is more than the statistical aggregate of the citizens living within its boundaries or sharing its passport. As our reviews show, it affects and guides the actions of its citizens across almost every domain of social life. But how is it that the aggregate gains power over its constituent elements? How does the context become a point of orientation for action?

5.2 Contexts as Aggregates and Points of Action Orientation

In seeking to answer this question, it is useful to look at different levels of contexts and examine whether and why they have the capacity to serve as points of action orientation. There are many contexts, that is levels of aggregation, above and below the nation: from family and neighborhood, through political and religious communities, networks, firms and plants, school classes and schools, to nations and transnational political units. But not every one of these regulates actions. What gives some of them this privilege? Two criteria suggest themselves:

The first stems from Weber's (1980, pp. 698–707) definition of a "*Verband*," a collectivity. According to him, a collectivity is a group of actors (1) devoted to a specific form of action or relationship, (2) which is represented by a leader speaking and acting in the name of all, and (3) whose members are oriented to a specific constitution, that is, a set of rules implicitly acknowledged, even when violated, by every member and potentially explicitly stated. In modern parlance, a collectivity becomes more than a random collection of persons once it is represented by a "collective actor" (Coleman 1990). A collective actor, of course, need not be a natural

person; indeed, in most cases, it is a legal person defined by the constitution adopted by the collectivity, that is, a president or government, a chief executive or a team captain. The collective actor sets the rules which orient the actions of its members. By its existence, what has been merely a statistical aggregation gains life in social reality.

Taking Weber's definition as a yardstick, not every context is a relevant frame of a person's thoughts and actions. This can be illustrated by examples on the lowest and highest levels of aggregation. Indeed, a nation, a community, and a parish certainly do constitute such a frame; all of them are built upon a specific form of action, led by a collective actor, and subject to a constitution. But a city neighborhood precinct, as delineated by the census bureau, has none of these. And a union of nation states has no collective actors of its own at the beginning but may construct these when implementing its genuine constitution, as the European Union is in the process of doing. As far as its powers reach, it can be a frame for the actions of all the individuals living on its territory. However, grouping nation states into Eastern and Western, that is, capitalist and former socialist states, or according to their "conservative" or "liberal" welfare "regimes" (Schröder 2019; Kroneberg 2019) constitutes a creation by the researcher. No collective actor is responsible for *the group* of these nations or welfare systems.[6] Rather, citizens follow the regulations and demands of their respective national welfare systems. "Regimes"—that is nations clustered according to similar property profiles—are analytical constructs which should not be reified.

Second, the distribution of the members' characteristics and the relations between them in a context—"social structure" in the distributional and relational sense, as a set of aggregate parameters and as a network (Meulemann 2013, pp. 275–287)—operates as a profile of personal opportunities which functions in favor of or counter to the life plans of each individual member, without being explicitly taken into account by those who are subject to it (Friedrichs and Nonnenmacher 2014, p. 4). Examples of the social structure as a distributional parameter are as follows: the gender ratio in a society skews the chance of marrying in favor of either men or women; the relative sizes of economic sectors in a society precondition the choice of occupational training; the unemployment ratio in an economy circumscribes the employment opportunities among the unemployed and engenders fear of unemployment among the employed; and a policy of educational expansion in a country increases university graduates' chances of finding an adequate lifetime position. Examples of the social structure as a network are: weak and strong ties within family, kin, and work furnish avenues to find a job, a marriage partner, a business opportunity, or a consumer bargain. Moreover, an ego-centered network is even more closely woven into people's life-world, and may affect their decisions

---

[6] Of course, countries which have the same welfare regime may install councils in order to learn from each other—as the Scandinavian welfare states did. If such councils attain power over their constituent countries, they can become a collective actor in their own right, and the borderline from aggregation to social reality will be transgressed—just as in the case of the European Union. Furthermore, such councils are examples of the interaction between collective actors, which is beyond the purview of CNCR. International relations may be a complementary research arena to cross-national comparison.

even more than would a non-personal "total" network of a community. In conceptual terms, it moves down from a macro to a micro property.

The characteristics whose distributions in a context constitute an opportunity profile are not restricted to demographic properties such as gender, education, and employment. They may also refer to norms guiding individual-level attitudes and behavior. The more some personal quality is in accord with the norm in a country, the more it will contribute to personal wellbeing. For example, the more a country is highly religious on average, the more personal religiosity will increase wellbeing (Stavrova 2019). In such cases, the distribution empirically operates as a behavioral model which need not be literally formulated as a norm and incorporated into a constitution.

Individuals may be aware of the advantages or disadvantages that are granted to them in their context and respond to them, or they may simply follow its predesigned tracks, such that the orienting capacity of the context may be more or less reflected in a search for orientation on the part of its subjects. It goes without saying that even though a context regulates the actions of its subjects through a collective actor, this does not preclude it operating as an opportunity structure as well. For example, decisions on a life career can be preformed by family policies as well as by the opportunity structure given by demographic variables (Grunow 2019; Hank and Steinbach 2019). It is also self-evident that a specific opportunity structure may even be more effective on context levels below the country level. For example, the gender ratio may more strongly affect marriage opportunities in city neighborhoods or in cities than it does at the country level, and the unemployment ratio may exert a stronger influence on the employment opportunities in a district than is the case at the country level (further examples in Friedrichs and Nonnenmacher 2014, p. 8). In this special issue, Careja reviews several studies which identified neighborhood characteristics conditioning opportunities for immigrants.

In summary, a context is no more than a statistical aggregate. Yet it can *become* a point of orientation for its members if there is a collective actor which demands contributions and grants support in specific life domains; or it can operate as an opportunity structure inadvertently affecting life decisions in these very domains—family, education, employment, politics, and others. Yet it is not clear from the outset that a given context has orienting power over its members, nor in which ways it operates as an opportunity structure. It is worthwhile to ask and examine how it attains such capacities.

6 Conclusion

There are many ways to compare societies. And there are many ways to distinguish between the levels of a society. Yet there are not so many ways to compare societies across their constituent levels with a single predefined method that is applicable in any societal domain, i.e., in an integrative perspective. Multilevel analysis provides such an approach. It presupposes a hierarchy of societal levels, such as citizens in nations, political parties in parliaments, or firms in economic sectors, comprising many units at the lower level and an adequate number at the higher level. Given that

data corresponding to the different levels exist, multilevel analysis can be applied to solve any substantive question. Such analysis uses a specific type of regression analysis, the so-called mixed models which combine equations for each level of the hierarchy and assume a corresponding structure of multiple error terms (see Sect. 2 and 4).

The first part of this special issue treats issues of research strategy typically encountered in multilevel analyses and the statistical models on which they rest. As for research strategies, the contributions discuss whether mechanisms mediating between citizens and nations require a third, meso, level in a multilevel design (Kroneberg), whether typologies are adequate to capture country groups and their differences (Schröder), or what the pitfalls and potential gains of case and context selection are (Goerres et al. 2019). As for statistical models, rules to define the error structure are developed (Schmidt-Catran et al.; for additional discussion on this topic, see also Meuleman et al. 2018) and the use of multiple indicators for a concept is advocated, although these are rarely implemented in multilevel analysis (Cieciuch et al.). It goes without saying that there are other strategic questions and developments of statistical modeling, and the ones presented here are not representative. But they do prove that multilevel analysis is a branch of *methodological* research in its own right, following its own dynamics, and open to any substantive application with the appropriate hierarchical data. As such a tool has now been available for a couple of decades, it seemed worthwhile to ask what has been achieved with it in specific applications.

With this question in mind, the second part of the special issue includes contributions on a vast array of research questions concerning the economy, politics, civil society, and culture—domains which may be rightly considered to make up the backbone of every modern society. Yet in each of these domains, we cannot pretend to address all the questions or even the most important ones. We were for instance unable to gather any contributions on criminal behavior or on leisure activities (beyond media use by Boomgarden and Song). We hope that the choice of topics reflects the state of the art rather than our predilections. The intention for each contribution was to synthesize widespread results, generated with a single instrument from the methodical toolbox of social science, into some conclusive answers. Based on these contributions, a few concluding remarks may be ventured concerning the standards, the practice, and the analytical strategy.

First, **the relative explanatory weight of the macro and micro level**: Measures of explained variance by country effects, such as the intra-class correlation coefficient (ICC), are not cited in quite a few of the summarized studies, and where they are cited, they are rather low. It is regrettable that the ICC is not presented, as the latter allows a rough evaluation to be carried out of the homogeneity or heterogeneity of the country sample. If it is heterogeneous—as in the WVS—then there should be ample room for context effects to operate and to detect large ICCs. If it is fairly homogeneous—as in the EVS or ESS—then the ICCs should be low. For example, the main result of a comparison of micro and macro determinants of civil engagement in the ESS countries was the "similarity of countries and diversity of people" (Meulemann 2002): civil engagement does not differ widely from one country to another, but it varies strongly with personal characteristics—and more or

less equally so within countries. Thus, small ICCs might reflect the homogeneity of country samples as well as weak country impacts. Quite apart from the statistical ascertainment of relative variances, the studies summarized in the reviews of this special issue seldom report the degree to which the addition of a second, macro, level has changed known results at the micro level. However, in a first review of a technique's performance, it is the yield rather than the surplus that should be reported.

Second, **improvement of measurements**: Low explained variances at both the macro as well as the micro levels may result from imperfect measurement. The small percentage of country level variance, that is the low ICCs, can—apart from the homogeneity of country samples—result from the deficient operationalization of country characteristics. Furthermore, the mechanisms that underlie context effects may operate at different levels (e.g., policies have to be enacted at local levels) and may hence be difficult to detect—and the more so the more distantly they are measured from the individual actors.

At the macro level, the measurement of country characteristics can be improved, and the mechanisms of their operation could be studied in greater detail. Unfortunately, doing what historians call "*Quellenkritik*" (evaluation of sources) often does not make its way into highly ranked journal publications, but more critical evaluations of the variables currently used to measure country effects are definitely required. Furthermore, in order to demonstrate how and by which intermediate steps context effects operate at the individual level, and how context characteristics emerge from individual behaviors and attitudes, more qualitative analyses should be performed, using case studies and process tracing.

At the micro level, measurement could be improved as well. Objective personal data should be ascertained along with subjective survey responses. For example, not only self-rated health data should be acquired from survey participants, but it should be provided from medical reports. In this vein, Stavrova (2015) used the 18 waves of the US General Social Survey National Death Index dataset and showed that the influence exerted by participants' religiosity on their longevity (measured by the occurrence of death as recoded in the dataset) was moderated by a country's average level of religiosity. Furthermore, many of the theoretical constructs analyzed in multilevel analyses are measured only with few items, and indeed sometimes with only one (an exception is Cieciuch et al. 2019). For example, the measurement of generalized trust with WVS data is based on one single survey question with only two response options.

Third, **causality**: Most of the multilevel findings are based on cross-sectional analyses, which are plagued by unclear causality directions and unobserved heterogeneity, especially at the country level. Many contributions therefore call for more longitudinal research, and use panel and event history data (such as Blossfeld et al., Careja, or Grunow). Schmidt-Catran et al. show how this could be done with repeated cross sections from the comparative survey projects which are readily available today (see also Sect. 4). To explore questions of causality, multilevel analyses based on many countries and large population-wide surveys can be complemented with country case studies using multi-item questionnaires (perhaps focusing on extreme or theoretically interesting cases; see Goerres et al. 2019).

Fourth, **broader theoretical integration**: The majority of the analyses presented in this special issue did not aim at theoretical generalization beyond the domains that were addressed. Given our goal of a *current stage* synthesis of domain-specific research, however, theoretical connections between or generalizations over contributions and the life domains treated therein probably could not yet be expected as a rule. Future researchers should nevertheless embark on research programs which systematically test overarching theories in a variety of domains of social life. Such an approach would not only test the explanatory power of our theories, but would also set the stage for broader theoretical integration.

All in all, by covering a broad range of life domains, this special issue aims to demonstrate that multilevel analysis is an over-arching means to compare societies and their constituents—an integrative perspective which does not presuppose theoretical generalizations, but may well stimulate them. At least we know of no other where results could be presented on such a broad range of life domains and questions as covered in this special issue.

References

Alderson, Arthur S., and Francois Nielsen. 2002. Globalization and the great U-turn: Income inequality trends in 16 OECD countries. *American Journal of Sociology* 107:1244–1299.

Andreß, Hans-Jürgen, Katrin Golsch and Alexander W. Schmidt. 2013. *Applied panel analysis for economic and social surveys*. Berlin: Springer.

Blau, Peter M., and Joseph E. Schwartz. 1984. *Crosscutting social circles: Testing a macrostructural theory of intergroup relations*. Orlando: Academic Press.

Blossfeld, Hans-Peter, Nevena Kulic, Jan Skopek, Moris Triventi, Elina Kilpi-Jakonen, Daniela Vono de Vilhena, and Sandra Buchholz. 2019. Conditions and consequences of unequal educational opportunities in the life course: Results from the Cross-National Comparative eduLIFE Project. In *Cross-*

national comparative research – analytical strategies, results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00595-w.

Bollen, Kenneth A., Barbara Entwisle and Arthur S. Alderson. 1993. Macrocomparative research methods. *Annual Review of Sociology* 19:321–351.

Boomgaarden, Hajo G., and Hyunjin Song. 2019. Media use and its effects in a cross-national perspective. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00596-9.

Bornschier, Volker, and Christopher K. Chase-Dunn. 1985. *Transnational corporations and underdevelopment*. New York: Praeger.

Boyd, Lawrence H., and Gudmund R. Iversen. 1979. *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth.

Bryk, Anthony S., and Stephen W. Raudenbush. 2002. *Hierarchical linear models: Applications and data analysis methods*. Second edition. Newbury Park: Sage.

Careja, Romana. 2019. Immigrants' labor market outcomes: Contributions from multilevel studies. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00597-8.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt and René Algesheimer. 2019. How to obtain comparable measures for cross-national comparisons. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00598-7.

CNCR. 2019. Cross-national Comparative Research. KZfSS Special Issue—(website). http://eswf.uni-koeln.de/cncr/ (Date of access: 8 Feb. 2019).

Cohen, Jeffrey E. 2004. Economic perceptions and executive approval in comparative perspective. *Political Behavior* 26:27–43.

Coleman, James S. 1990. *Foundations of social theory*. Cambridge, MA: Belknap Press of Harvard University Press.

Context. 2019. ESS Multilevel Data. https://www.europeansocialsurvey.org/data/multilevel/guide/about.html (Date of access: 29 Jan. 2019).

Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt and Jaak Billiet. 2014. Measurement equivalence in cross-national research. *Annual Review of Sociology* 40:55–75.

Davidov, Eldad, Peter Schmidt, Jaak Billiet and Bart Meuleman. 2018. *Cross-cultural analysis: Methods and applications, second edition*. New York: Taylor & Francis.

Diener, Ed, and Shigehiro Oishi. 2000. Money and happiness: Income and subjective well-being across nations. In *Culture and subjective well-being*, eds. Ed Diener and Eunkook Mark Suh, 185–218. Cambridge: The MIT Press.

Diener, Ed, Louis Tay and David G. Myers. 2011. The religion paradox: If religion makes people happy, why are so many dropping out? *Journal of Personality and Social Psychology* 101:1278–1290.

Easterlin, Richard A. 1974. Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth: Essays in honor of Moses Abramovitz*, eds. Paul A. David and Melvin W. Reder, 89–125. New York: Academic Press.

Erlinghagen, Marcel. 2019. Employment and its institutional contexts. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00599-6.

ESS. 2019. European Social Survey. https://www.europeansocialsurvey.org (Date of access: 29 Jan. 2019).

EVS. 2019. European Values Study. https://europeanvaluesstudy.eu (Date of access: 29 Jan. 2019).

Flora, Peter. 1983 and 1986. *State, economy, and society in Western Europe*. Frankfurt a. M.: Campus.

Freedman, David A. 2004. The ecological fallacy. In *The SAGE Encyclopedia of Social Science Research Methods, Volume 1*, eds. Michael S. Lewis-Beck, Alan Bryman and Tim Futing Liao, 293. Thousand Oaks: Sage.

Frick, Joachim R., Stephen P. Jenkins, Dean R. Lillard, Oliver Lipps and Mark Wooden. 2007. The Cross-National Equivalent File (CNEF) and its member country household panel studies. *Schmollers Jahrbuch* 127:627–654.

Friedrichs, Jürgen, and Alexandra Nonnenmacher. 2014. Die Analyse sozialer Kontexte (An analysis of social contexts). In *Soziale Kontexte und soziale Mechanismen (Kölner Zeitschrift für Soziologie Sonderheft 54/2014)*, eds. Jürgen Friedrichs and Alexandra Nonnenmacher, 1–16. Wiesbaden: Springer.

Gallie, Duncan, and Serge Paugam. Eds. 2000. *Welfare regimes and the experience of unemployment in Europe*. Oxford: Oxford University Press.

Gauthier, Anne Hélène. 1996. *The state and the family: a comparative analysis of family policies in industrialized countries*. Oxford: Clarendon Press.

GGS. 2019. Generations and Gender Survey. https://www.ggp-i.org (Date of access: 29 Jan. 2019).

Giesselmann, Marco, and Alexander W. Schmidt-Catran. 2018. Getting the within estimator of cross-level interactions in multilevel models with pooled cross-sections: Why country dummies (sometimes) do not do the job. *Sociological Methodology* Online First, https://doi.org/10.1177/0081175018809150.

Goerres, Achim, Markus B. Siewert and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Goldstein, Harvey. 1986. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73:43–56.

Grunow, Daniela. 2019. Comparative analyses of housework and its relation to paid work: Institutional contexts and individual agency. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00601-1.

Guerin, Daniel, Jean Crete and Jean Mercier. 2001. A multilevel analysis of the determinants of recycling behavior in the European countries. *Social Science Research* 30:195–218.

Hall, Peter A., and Rosemary C. R. Taylor. 1996. Political science and the three new institutionalisms. *Political Studies* 44:936–957.

Halman, Loek, and John Gelissen. 2019. Values in life domains in a cross-national perspective. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS.

Hank, Karsten, and Anja Steinbach. 2019. Families and their institutional contexts: The role of family policies and legal regulations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00603-z.

Hantrais, Linda. 2008. *International comparative research: Theory, methods and practice*. Basingstoke: Palgrave Macmillan.

Harkness, Janet A., Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith. 2010. *Survey methods in multinational, multiregional, and multicultural contexts*. New York: Wiley.

Harkness, Janet A., Fons J. R. Van De Vijver and Peter Ph. Mohler. Eds. 2003. *Cross-cultural survey methods*. Hoboken: Wiley.

Henrich, Joseph, Steven J. Heine and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33:61–135.

Hoffmeyer-Zlotnik, Jürgen H. P., and Christof Wolf. 2011. *Advances in cross-national comparison: A European working book for demographic and socio-economic variables*. New York: Springer.

Hosoya, Georg, Tobias Koch and Michael Eid. 2014. Längsschnittdaten und Mehrebenenanalyse (Longitudinal data and multi-level analyses). In *Soziale Kontexte und soziale Mechanismen (Kölner Zeitschrift für Soziologie Sonderheft 54/2014)*, eds. Jürgen Friedrichs and Alexandra Nonnenmacher, 189–219. Wiesbaden: Springer.

Inkeles, Alex. 1960. Industrial man: The relation of status to experience, perception, and value. *American Journal of Sociology* 66:1–31.

ISSP. 2019. International Social Survey Programme. https://www.issp.org (Date of access: 29 Jan. 2019).

Johnson, Timothy P., Beth-Ellen Pennell, Ineke A. L. Stoop and Brita Dorer. 2018. *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*. New York: Wiley.

Just, Aida, and Christopher J. Anderson. 2012. Immigrants, Citizenship and Political Action in Europe. *British Journal of Political Science* 42:481–509. https://doi.org/10.1017/S0007123411000378.

Kaltenthaler, Karl C., and Christopher J. Anderson. 2001. Europeans and their money: Explaining public support for the common European currency. *European Journal of Political Research* 40:139–170.

Kohn, Melvin L. 1987. Cross-national research as an analytic strategy: American Sociological Association, 1987 Presidential Address. *American Sociological Review* 52:713–731.

Kohn, Melvin L. 2015. *Change and stability: A cross-national analysis of social structure and personality*. New York: Taylor & Francis.

Kroneberg, Clemens. 2019. Theory development in comparative social research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00604-y.

Landman, Todd. 2017. *Issues and methods in comparative politics. An introduction. 4th edition*. London and New York: Routledge.

Lazarsfeld, Paul, and Herbert Menzel. 1969. On the relation between individual and collective properties. In *A sociological reader on complex organization*. 2nd edition, ed. Amitai Etzioni, 422–440. New York: Holt, Rinehart and Winston.

Lewis, Jeffrey B., and Drew A. Linzer. 2005. Estimating regression models in which the dependent variable is based on estimates. *Political Analysis* 13:345–364.

Lipset, Seymour Martin. 1959. Democracy and working-class authoritarianism. *American Sociological Review* 24:482–501.

March, James G., and Johan P. Olsen. 1989. *Rediscovering institutions*. New York: Free Press.

Meuleman, Bart. 2019. Multilevel Structural Equation Modeling for Cross-National Comparative Research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00605-x.

Meuleman, Bart, Eldad Davidov and Jaak Billiet. 2018. Modeling multiple-country repeated cross-sections. A societal growth curve model for studying the effect of the economic crisis on perceived ethnic threat. *Methoden, Daten, Analysen* 12:185–210.

Meulemann, Heiner. 2002. Ed. *Social capital in Europe: Similarity of countries and diversity of people? Multi-level analyses of the European Social Survey 2002*. Leiden: Brill.

Meulemann, Heiner. 2013. *Soziologie von Anfang an. Eine Einführung in Themen, Ergebnisse und Literatur. 3rd ed.* Wiesbaden: Springer.

Meulemann, Heiner. 2017. The secularization theory—not disconfirmed, yet rarely tested. *Analyse und Kritik* 39:325–355.

Meyer, John W., John Boli, George M. Thomas and Francisco O. Ramirez. 1997. World society and the nation-state. *American Journal of Sociology* 103:144–181.

Minkov, Michael. 2013. *Cross-cultural analysis: The science and art of comparing the world's modern societies and their cultures*. Los Angeles: Sage.

Nonnenmacher, Alexandra, and Jürgen Friedrichs. 2013. The missing link: Deficits of country level studies. A review of 22 articles explaining life satisfaction. *Social Indicators Research* 110(3):1221–1244.

Norris, Pippa, and Ronald Inglehart. 2004. *Sacred and secular. Religion and politics worldwide*. Cambridge: Cambridge University Press.

Opp, Karl-Dieter. 2009. *Theories of political protest and social movements. A multidisciplinary introduction, critique and synthesis*. London: Routledge

Parsons, Talcott. 1964. Evolutionary universals in society. *American Sociological Review* 29:339–357.

Pescosolido, Bernice. 2007. Sociology of social networks. In *21st century sociology: A reference handbook, Volume 1*, eds. Clifton D. Bryant and Dennis L. Peck, 208–217. Thousand Oaks: Sage.

Pförtner, Timo-Kolja, Holger Pfaff and Frank J. Elgar. 2019. The role of welfare state characteristics for health and inequalities in health from a cross-national perspective: A critical research synthesis. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00606-w.

Przeworski, Adam, and Henri Teune. 1970. *The logic of comparative social inquiry*. New York: Wiley-Interscience.

Robinson, William S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 15:351–357.

Saxonhouse, Gary R. 1976. Estimated parameters as dependent variables. *The American Economic Review* 66:178–183.

Saxonhouse, Gary R. 1977. Regressions from samples having different characteristics. *The Review of Economics and Statistics* 59:234–237.

Schmidt-Catran, Alexander W. and Malcolm Fairbrother. 2016. The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review* 32:23–38.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialspychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schmitt-Beck, Rüdiger. 2019. Political systems and electoral behavior: A review of internationally comparative multilevel research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Schröder, Martin. 2019. Varieties of capitalism and welfare regime theories: Assumptions, accomplishments, and the need for different Methods. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00609-7.

SHARE. 2019. Survey of Health, Ageing and Retirement in Europe. https://www.share-project.org (Date of access: 29 Jan. 2019).

Siegers, Pascal. 2019. Is the influence of religiosity on attitudes and behaviors stronger in less religious or more religious societies? A review of theories and contradictory evidence. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00610-0.

Spies, Dennis Christopher, and Simon T. Franzmann. 2019. Party competition and vote choice. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00611-z.

Stavrova, Olga. 2015. Religion, self-rated health, and mortality: Whether religiosity delays death depends on the cultural context. *Social Psychological and Personality Science* 6:911–922.

Stavrova, Olga. 2019. How much do sources of happiness vary across countries? A review of the empirical literature. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00612-y.

Swamy, Paravastu A. V. B. 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38:311–323.

The Local. 2015. Danish police: Refugees can travel on to Sweden. TT/AFP/The Local, September 10, 2015. https://www.thelocal.se/20150910/danish-police-allow-refugees-to-continue-on-to-sweden (Date of access: 15 Jan. 2019).

Treiman, Donald J. 1977. *Occupational prestige in comparative perspective*. New York: Academic Press.

Wallerstein, Immanuel Maurice. 2011a. *The modern world-system III: The second era of great expansion of the capitalist world-economy, 1730s–1840s*. Berkeley and Los Angeles: University of California Press.

Wallerstein, Immanuel Maurice. 2011b. *The modern world-system IV: Centrist liberalism triumphant, 1789–1914*. Berkeley and Los Angeles: University of California Press.

Wallerstein, Immanuel Maurice. 2011c. *The modern world-system II: Mercantilism and the consolidation of the European world-economy, 1600–1750*. Berkeley: University of California Press.

Wallerstein, Immanuel Maurice. 2011d. *The modern world-system I: Capitalist agriculture and the origins of the European world-economy in the sixteenth century*. Berkeley and Los Angeles: University of California Press.

Weber, Max. 1980. Soziologische Grundbegriffe (Basic sociological terms). In *Max Weber Schriften 1894–1922*, ed. Dirk Käsler, 653–716. Stuttgart: Kröner (originally published 1920).

Welch, Michael R., Charles R. Tittle and Thomas Petee. 1991. Religion and deviance among adult Catholics: A test of the "moral communities" hypothesis. *Journal for the Scientific Study of Religion* 30:159–172.

Wilkinson, Richard G. 2006. *The impact of inequality: How to make sick societies healthier*. New York: New Press.

Zapf, Wolfgang. 1994. *Modernisierung, Wohlfahrtsentwicklung und Transformation. Soziologische Aufsätze 1987–1994*. Berlin: Edition sigma.

Ziller, Conrad. 2019. Policy Effects on Political Engagement. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00613-x.

Hans-Jürgen Andreß 1952, Prof. i.R. Dr., Univ.-Prof. for Empirical Social and Economic Research at the Institute for Sociology and Social Psychology, University of Cologne. Publications: Applied panel data analysis for economic and social surveys. Berlin 2013 (with K. Golsch and A.W. Schmidt); In search of a frame: Challenges and opportunities for sampling immigrant minorities. Special Issue of Comparative Migration Studies 2018 (with R. Careja); Is material deprivation decreasing in Germany? A trend analysis using PASS data from 2006 to 2013. Journal for Labour Market Research 52, 2018.

Detlef Fetchenhauer 1965, Univ.-Prof. for Economic and Social Psychology at the Institute for Sociology and Social Psychology, University of Cologne. Publications: Belief in scientific-technological progress and life satisfaction: The role of personal control. Personality and Individual Differences 2016 (with O. Stavrova and D. Ehlebracht); Married and cohabiting parents' well-being: The effects of a cultural normative context across countries. Journal of Social and Personal Relationships 58, 2014 (with O. Stavrova); Solidarity in the absence of external sanctions: A cross-cultural study of educational goals and fair-share behavior. In: Solidarity and prosocial behavior. New York 2006 (with R. Wittek; eds. D. Fetchenhauer, A. Flache, B. Buunk, and S. Lindenberg).

Heiner Meulemann 1944, Prof. em. Dr., Univ.-Prof. for Sociology at the Institute for Sociology and Social Psychology, University of Cologne. Publications: Soziologie von Anfang an. Eine Einführung in Themen, Ergebnisse und Literatur. 3rd. revised edition. Wiesbaden 2013; Ohne Kirche leben. Säkularisierung als Tendenz und Theorie in Deutschland, Europa und anderswo. Wiesbaden 2018; Lebensdeutung und Lebensplanung in der Lebensmitte. Vom Gymnasium bis zur Planung des Ruhestands. Wiesbaden 2017 (with K. Birkelbach); Bin ich meines Glückes Schmied? Kausalattribution des persönlichen Berufserfolgs und der sozialen Mobilität zwischen früher und später Lebensmitte. Zeitschrift für Soziologie 45, 2016; The secularization theory—not disconfirmed, yet rarely tested. Analyse und Kritik 39, 2017.

Media Use and Its Effects in a Cross-National Perspective

Hajo G. Boomgaarden · Hyunjin Song

Abstract This contribution provides a critical reflection on the state-of-the-art of cross-national media use and media effect studies. Increasing availability of data sources, advances in theorizing and facilitation of international research collaboration have contributed to an increasing application of cross-national perspectives in communication research. Contingencies of media use and media effects brought about by national media systems or sociopolitical and cultural contexts of media use have become a central tenet of such research. The paper starts out by discussing the need for cross-national comparative perspectives in communication research. It then goes on to review the generally problematic nature of "media use" measurement, in particular in a comparative perspective, followed by an introduction to media systems and information environments as among the central macrolevel concepts in media use and media effect studies. In its core, the contribution reviews multilevel studies examining media use and the impact had by media, most of which stem from the realm of news use and its effects in politics. The article then discusses whether and to what degree these cross-national studies have contributed to further theory building. It concludes by discussing and providing an outlook for the future of comparative communication research.

Keywords Media use · Media effects · Comparative communication research · Measurement

H. G. Boomgaarden (✉) · H. Song
Department of Communication, University of Vienna
Rathausstraße 19/1/9, 1010 Vienna, Austria
E-Mail: hajo.boomgaarden@univie.ac.at

H. Song
E-Mail: hyunjin.song@univie.ac.at

🖄 Springer

Mediennutzung und ihre Auswirkungen in einer länderübergreifenden Perspektive

Zusammenfassung Dieser Beitrag liefert eine kritische Reflexion über den Stand der länderübergreifenden Studien zu Mediennutzung und Medienwirkung. Die zunehmende Verfügbarkeit von Datenquellen, Fortschritte in der Theoretisierung und Erleichterung der internationalen Forschungszusammenarbeit haben zu einer zunehmenden Anwendung länderübergreifender Perspektiven in der Kommunikationsforschung beigetragen. Kontingenzen der Mediennutzung und Medieneffekte, die durch nationale Mediensysteme oder gesellschaftspolitische und kulturelle Kontexte der Mediennutzung hervorgerufen werden, sind zu einem zentralen Grundsatz dieser Forschung geworden. Das Papier beginnt mit der Diskussion über die Notwendigkeit länderübergreifender vergleichender Perspektiven in der Kommunikationsforschung. Anschließend wird die allgemeine Problematik der Messung von „Mediennutzung", insbesondere in vergleichender Perspektive, untersucht, dem folgt eine Einführung in Mediensysteme und Informationsumgebungen als zentrale Konzepte der Mediennutzungs- und Medienwirkungsstudien auf Makroebene. Im Kern werden in dem Beitrag Mehrebenenstudien betrachtet, die die Mediennutzung und die mediale Wirkung untersuchen, von denen die meisten aus dem Bereich der Nachrichtennutzung und ihrer Auswirkungen in der Politik stammen. Darüber hinaus wird in dem Artikel erörtert, ob und inwieweit diese länderübergreifenden Studien zur weiteren Theoriebildung beigetragen haben. Abschließend wird ein Ausblick auf die Zukunft der vergleichenden Kommunikationsforschung gewährt sowie darüber diskutiert.

Schlüsselwörter Mediennutzung · Medienwirkung · Vergleichende Kommunikationsforschung · Messung

1 Introduction

The media have taken central stage as drivers and facilitators of a great range of social, psychological or political processes (Bryant and Oliver 2009). They for instance appear as a main source of information for sociopolitical engagement (Anduiza et al. 2012; Beck et al. 2002; McLeod et al. 1999), and thereby contribute to the (mal)functioning of democracy (Curran 2011; Delli Carpini and Keeter 1993). While being a main source of one's social identities (Slater 2007, 2015), they are considered a threat to social cohesion (Putman 2000), as well as a factor producing social capital (Shah et al. 2001; Campbell and Kwak 2010). They are seen as a major source of pleasure, relaxation, and gratification (Zillmann 1988; Vorderer et al. 2004), but are also associated with decreased well-being (such as loneliness and depression) and problematic behaviors (Becker et al. 2013; Kim et al. 2009). Hence a systematic engagement with questions about why it is that people are using media, and what the consequences of media use are for cognitions, attitudes and behaviors, has been characterizing empirical studies in communication science.

Broadly classified as transmitters of information between senders and receivers, media are broken down into those types with a purely technological function (such

as the Internet or a telephone) and those which are involved not only with the transmission, but more importantly with the selection and packaging of information (e.g., traditional mass media outlets, social media). Interactivity has become more prominent in recent decades, with the boundaries between senders and receivers of information becoming increasingly blurred (Neuman 2016). While the sheer usage of different types of media is the focus of a rich research tradition (Althaus and Tewksbery 2000; Katz et al. 1973; Sears and Freedman 1967) including various motivations, needs, and predispositions that underlie such media usage patterns (e.g. Donohew et al. 1987; Wang et al. 2012), at least an equally prominent tradition is formed by research looking at the consequences of such media usage for a variety of cognitive, emotional, attitudinal or behavioral outcomes and their contingency conditions (e.g. Bryant and Oliver 2009; Nabi and Oliver 2009; Potter 2011; Valkenburg and Peter 2013). These two research traditions, taking individual media use as either an outcome or as an explanatory variable, are at the heart of this contribution.

Media use has been largely treated as an individual-level construct, with individuals more or less consciously and voluntarily deciding to turn their attention toward any types of mediated communication. Factors influencing individuals' media diets relate to information seeking, motivations and needs, or personal predispositions (Bartsch et al. 2006; Blumler 1979; Ruggiero 2000). Media use, then, has been shown to affect a range of "individual" outcomes along the continuum of cognition, affect, attitudes, and behaviors (e.g. Bryant and Oliver 2009). Such individual media use does not however take place in a vacuum, but is hosted in a range of contextual levels, including families, neighborhoods, or media markets (Slater et al. 2006). While one's microsocial settings—namely interpersonal influences—during communication, or in the pre- and postcommunicative phase have been acknowledged and subject to empirical investigation (e.g. Boomgaarden 2014; Schmitt-Beck 2003; Southwell and Yzer 2007), larger contexts and environments and how they interact with individual media use have been somewhat neglected for quite some time (Pan and McLeod 1991): "communication continues to be dominated by research at the individual level of analysis" (McLeod et al. 2010, p. 183).

It could be argued that such an individual, microlevel focus on communication (a) neglects a larger contextualization of research findings, (b) thereby ignores debates of universal applicability and generalizability versus the context dependency of empirical findings, and as a consequence, (c) misses out on the opportunities for further theorizing and theory refinement regarding media use and its effects. Considering media use as embedded in higher-level structures will thus enable a more comprehensive, encompassing, and arguably the theoretically enhanced understanding of the role of media in contemporary societies. "The lack of theoretical and empirical connections between levels has produced explanations of communication phenomena that center on internal mental states rather than on social, economic and political conditions" (McLeod et al. 2010, p. 184). Such a line of argument has certainly gained in relevance, given the rapidly changing contexts in which communication takes place, in terms of advancements of communication technologies and ever-growing distribution of such technologies, and in terms of economic and cultural global interconnectedness as a consequence of communication technologies (McLuhan and Powers 1989; Castells 2011). As recently argued, "scholars have be-

come sensitive to the contexts of democratic development, sociocultural influences and economic (de)regulation and eventually drastic changes in information environments to shape citizens' media and political roles" (Pfetsch and Esser 2014). This contribution addresses the degree to which a systematic contextualization of media practices indeed takes place in the current empirical literature.

In a seminal paper, McLeod and Blumler (1987) provide three major reasons for serious consideration of the macrolevel in media and communication studies: (a) generating a more comprehensive understanding of media processes, (b) making the field relevant for public policy making, and (c) recognizing that economic, social and technological contexts are theoretically important factors in media use and their effects. With increasing trends of globalization and transnationalization of media, "it is no longer plausible to study a phenomenon in one country without asking whether it is common across the globe or distinctive to that specific context" (Esser and Hanitzsch 2012, p. 3). And if the latter is true, it is indeed imperative to understand why that may be the case, addressing the specificity and generalizability of one's findings across different geographical, national, or cultural contexts (Livingstone 2003). Consequentially, communication has been devised as a cross-level "variable" field, rather than a single "level" field (Paisley 1984).[1]

Following Edelstein's (1982, p. 14) definition, we conceptualize comparative communication (or comparative media) research as "a study that compares two or more nations with respect to some common activity" with the theoretical focus of such comparison of "common activity" being media use and effects. At the very basic level, comparative research on media processes, if it takes seriously the interdependencies of individual and contextual variables, "creates a need to think structurally, to conceptualize in macro terms, to stretch vertically across levels and horizontally across systems" (Blumler et al. 1992, p. 8). Context can be close to communication phenomena (such as media systems or information environments, see further below), or more remote but still bearing theoretical relevance (e.g. differences in culture, values, or political systems). Following McLeod et al. (2010, p. 192), contexts "are broadly defined as properties of macro-units that operate as constraints, shaping individual-level (or lower-level) phenomena through incentives or reducing patterns by deterrents or sanctions." Comparative communication research therefore "attempts to reach conclusions beyond single systems or cultures, and explains differences and similarities between objects of analysis against the backdrop of their cultural conditions" (Esser and Hanitzsch 2012, p. 5). Ultimately it is about understanding how characteristics of the contextual environment in which individual media users are situated shape their communication processes, and how such processes vary across different settings (Blumler and Gurevitch 1995).

Truly comparative communication research has been a marginal field for a long time, and even today the spread of comparative approaches is rather uneven across different subfields of the discipline. Most common, if comparative perspectives are taken into account, are studies that would be classified as "comparative case studies"

---

[1] According to Paisley (1984), "variable" fields tend to focus on a single "variable" (e.g. communication) across all levels of analysis, whereas "level" fields tend to fix their levels of analysis and focus on all relevant variables within such a single level.

in which at most two (more or less) identical studies in two different contexts are compared (or single cases are contrasted to prior empirical observations in the framework of prototypical or deviant cases). These are what Vliegenthart (2012) would coin as descriptive, or basic explanatory, comparative designs that rely at most on a qualitative comparison between different cases, in light of the logic of most similar or most different system designs (Teune and Przeworski 1970). Studies that formally compare more than two contexts from an (explicit) *cross-national* perspective are somewhat less common. Such studies appear to constitute the standard for comparative research in many areas of media use and media effect studies (Esser and Hanitzsch 2012, p. 13), often taking the form of "comparison of relations" (Vliegenthart 2012, p. 487), albeit such applications are still rare compared to a merely descriptive comparative design.

Large-N, multilevel comparative approaches and applications of *cross-national* perspectives to individual-level media processes have been increasing in the past two decades. Sparked to some degree by studies in the subfields of journalism (Hanitzsch et al. 2011) and political communication (Esser and Pfetsch 2004), and given great impetus by a special issue of Human Communication Research (Slater et al. 2006), such comparative perspectives have entered the stage in several subfields (see the collection by Esser and Hanitzsch 2012, for an extensive review). Such studies have drawn on large-scale cross-national research projects such as *World of Journalism, Providing an Infrastructure for Research on Electoral Democracy in the European Union* (PIREDEU), *Cooperation in Science & Technology* (COST), *Action Populist Political Communication in Europe, World Values Survey* (WVS), *European Election Studies* (EES), or the *European Social Survey* (ESS). Collaborative efforts on the part of international networks of scholars such as the *Network of European Political Communication Scholars* (NePoCS) or the *Comparative National Election Project* (CNEP) also have played a critical role in increasing the availability of appropriate data sources for research on media use and their effects. These have in turn brought about advances when it comes to theorizing the contingencies of media use and their effects triggered by national media systems or sociopolitical and cultural contexts.

This contribution focuses on reviewing the state-of-the-art of the latter category of comparative media use and media effects studies, considering only those empirical contributions that, at the very least, employ multilevel models explaining either media use or looking at the effects of individual media use while taking into account country-contextual variation. Within such an explicit focus on multilevel modeling studies, it appears that these largely stem from the broad subfield of political communication, which is arguably due to the nature of the data sources in use. Before reviewing the individual contributions from the field, it appears imperative to briefly discuss two central variables and their measurement problems in a comparative perspective, on individual-level media use (or media exposure) and on the contextual level of media systems and information environments. As will subsequently become evident, while the large-N, multilevel comparative approaches using multilevel linear modeling (MLM) literature on media effects have gained some momentum over the past decade, literature on explanations of media use across countries and taking contextual variation into account is still in its infancy. We will conclude with a discussion of potential future avenues and challenges to such comparative endeavors.

2 Conceptual and Empirical Considerations

2.1 Comparative Measurement of Media Use

The measurement of media use is central to answering individual-level questions regarding explanations of time investment into media reception or the consequences of the media. We speak of media use and media exposure as synonyms here, the former being more of a tradition in reception research (with the idea of using media for a given purpose), the latter more of a tradition in effects research. *Media use is a quantitative measure of self-reported time spent on using media in general, or certain subcategories of media or media outlets, or of the frequency of use of such media in a given timeframe* (e.g. per week or month). While it is nowadays often included in different kinds of survey research (e.g. election studies, social surveys), the measurement of media use is by no means unproblematic (Slater 2004). A vibrant debate revolves around the potential of overreporting media use (e.g. Prior 2009a, 2009b), alternative measurements (Dilliplane et al. 2013; Prior 2013) and possible ways ahead to come to theoretically useful and methodologically valid measures (de Vreese and Neijens 2016).

It could be argued that the measurement problem is considerably enlarged in the case of comparative cross-country surveys. It is important here to distinguish between media use in more general terms, and genre- or outlet-specific media use. The former for instance asks how many hours per week an individual spends watching TV, how many days per week (s)he reads a printed newspaper, or how many minutes (s)he spends per day using social media. These categories are more or less readily comparable between countries or media systems. From an analytical point of view, however, such "overall time spent" measures appear somewhat less useful given the very generic and content-agonistic nature of the measurement instruments. In contrast, the latter (outlet-specific) media use measures would for instance probe the readership of a particular newspaper title, watching a particular television show, or following a certain actor on social media. For media effects research in particular, in which knowledge about the particular contents to which people are exposed matters, such measurement is often to be preferred, allowing for linkage approaches integrating media content and exposure measures (de Vreese et al. 2017). Such categories are however much more difficult to validly measure and standardize across different national contexts in a comparative perspective, given the particularities of national media systems. The literature hence faces a friction between the robustly comparable measurement of broader categories, and more detailed, analytically more useful, but certainly less comparable and more demanding, measures of detailed, outlet-specific exposure.

International survey programs reflect this diversity of approaches. The *European Social Survey,* for instance, has always used rather generic questions on media use, probing the total time spent watching television, and the time spent watching news and current affairs on television (until wave 7), or the total time spent watching, reading or listening to news about politics and current affairs, and the time spent using the Internet (wave 8 onwards). A similar focus on political information is seen in the *European Values Study* (2008), which asks whether respondents actively

follow politics in the news on television, on the radio and in the newspapers, all in one question. The *World Values Survey* (2010–2014) also takes a fairly generic approach, and asks about the frequency of using different types of media, such as newspapers, magazines, television news or the Internet. The *European Election Study* (2009, 2014) includes an intense battery of media use questions, probing the frequency of use of particular media outlets in the different countries, both television and newspapers. While such outlet-specific measurement, which naturally is limited to a number of outlets per country, may not capture all the details of every individual's media use, it is analytically valuable in connection with measures of media content. Interestingly, the literature has not so far paid sufficient attention to estimations or evaluations of the measurement equivalence of media use measures across countries (Hanitzsch and Esser 2012).

2.2 Nations as Units: Media Systems and Information Environments

Much comparative research, in particular in the realm of political communication, relies on the macrolevel classification of nations and countries along the lines of different media systems or information environments, sometimes broadly termed as "opportunity structures" for media exposure.[2] As is further discussed below, the organizational structure of media is supposed to influence individual media use or structure relationships between media use and other outcome variables. While it is increasingly contended that the nation itself may not be the most appropriate unit of analysis (Livingstone 2003), the idea of classifying countries originated from the work aiming to understand differences in how the press are organized (Siebert et al. 1956), identifying four ideal types (libertarian, social responsibility, authoritarian and Soviet-type models). Extending this line of work, and taking up extensions proposed by Blumler and Gurevitch (1975), the seminal work of Hallin and Mancini (2004) paved the way for a strong acknowledgement of media systems in comparative communication research. The authors develop a typology of media systems based on four empirical dimensions: (1) degree of state interventionism, (2) degree of press-party parallelism (Seymour-Ure 1974), (3) professionalization of journalism, and (4) degree of commercialization, resulting in three ideal types: (1) liberal model, (2) democratic corporatist model, and (3) polarized pluralist model. Such media systems then "describe typical patterns of how journalism cultures, media policy, media markets and media use are connected in a given society" (Brüggemann et al. 2014, p. 1038). The model has been recently extended to non-Western contexts (Hallin and Mancini 2012), and subject to rigorous empirical investigations, resulting in a somewhat revised typology (Brüggemann et al. 2014). Implementation of such macrolevel, media system structures into comparative research strongly depends on the available data, which are either combined from a variety of sources and sometimes imperfect in their operationalization (Brüggemann et al. 2014), or gathered through indirect sources such as expert interviews (Popescu et al. 2011).

---

[2] Sometimes these are not necessarily identical to country units, such as two different media systems in Belgium (due to language differences) or smaller media markets (designated market area, or DMA) in the U.S., where such different opportunity structures exist *within* a single country.

Beyond the structural characteristics of media systems, recent research increasingly acknowledges the importance attached to information environments as macrolevel characteristics[3] influencing microlevel media consumption *processes*. Such information environments consider the total information available to people within the boundaries of media systems (Pfetsch and Esser 2014). Information environments are therefore empirical observations of outputs of media sources routinely available to national audiences (Esser et al. 2012). As such, they may determine individual media use and their effects by offering opportunity structures of encountering certain types of information. It is argued that information environments are tied to media systems in the sense that certain structural characteristics would favor or be biased towards certain types of information flows (Aalberg et al. 2010). Comparative research then focuses on identifying and isolating a certain characteristic of this information environment, aggregated to the country level. Measurement of such characteristics is, however, demanding since it requires actual media output data from a great range of country contexts. While this used to be only realizable via large-scale projects in which media analysts working in a variety of languages are employed (e.g. PIREDEU), nowadays there is an increasing potential for (semi-) automated procedures to analyze media outputs from a variety of languages and sources (e.g. Proksch et al. 2018; Lind et al. n.d.).

3 Comparative Research on Media Use

A long research tradition on predictors and correlates of media use has generally pointed to individual-level factors such as socioeconomic status and education (Livingstone 1998), interest and motivations (Blumler 1979; Knobloch 2003; Knobloch-Westerwick and Kleinman 2012; Papacharissi and Rubin 2000; Price and Zaller 1993), or demographic characteristics such as age, gender or race (Lauf 2001; Livignstone 1998; Roe 2000). These approaches very much reflect theorizing that focused on intrinsic motivations and gratifications that individuals sought from media consumption in order to explain why certain individuals consume certain media or genres. Literature addressing this topic from a comparative perspective however generally holds that, as previously acknowledged, "communication not only reflects one's individual predispositions but also the nature of one's social environment" (Cho 2011, p. 434). Motivated by such a principle, indeed many cross-national comparative media studies explicitly aim to understand the degree to which, how and why people prefer to see certain media or genres over others as a function of some national, cultural, or media system-level differences across nations (Althaus et al. 2009; McLeod and Lee 2012). Below, we review and highlight a fairly comprehensive, but not exhaustive, set of exemplary studies on the topic of explanations and correlates of media use. Table 1 gives a broad overview in alphabetical order of the studies reviewed in the following sections, with their key constructs regarding cross-national differences in media use and their effects.

---

[3] This is not to be confused with "individual-level" information environments, which denote the total diet of information consumed by individuals (Jerit et al. 2006).

Table 1 Key major studies of comparative media use and media effects and their methodological details

Study	System-level variables and/or country-case selection criteria	Individual-level dependent variables	Survey sources (N of upper units)
Comparative studies on media use			
Aalberg et al. (2013)	Media system (per Hallin & Mancini) measured as national contexts	Internet use, time used for newspaper reading	ESS (31 countries)
[a] Althaus et al. (2009)	Aggregated demographic characteristics (% of white, foreign-born, 4-year college degree, age 65 or older, etc.) Median income & population density of DMA Market size Market complexity	Average days per week of self-reported news exposure	NAES (206 US DMAs)
Bright (2018)	Ideological distance of political parties **DV**: *Fragmentation of party twitter network*	–	EU Parliament election 2014 (28 EU countries)
Castro et al. (2018)	Majoritarian/proportional electoral system Media-party parallelism Presence of public service broadcasting	Cross-cutting news media exposure	EES 2009 (27 countries)
Elvestad and Blekesaune (2008)	Media system (per Hallin & Mancini) and in Eastern bloc (measured as national contexts) Newspaper circulation and reach Average use of TV, radio, Internet Aggregated demographic characteristics (population density, unemployment rates, average income and education, etc.) Average interest, trust in parliament, trust in press	Newspaper reading	ESS (23 countries)
Esser et al. (2012)	The availability of news and current affairs programs	–	NEPOCS (13 countries)
Fletcher and Nielsen (2017)	Media system (per Hallin & Mancini) **DV**: *Audience fragmentation vs. duplication*	–	2016 Reuters Institute Digital News Report (6 countries)
Goldman and Mutz (2011)	Newspaper—party parallelism TV—party parallelism	Perceived favoritism of media (cross-cutting media use)	CNEP (11 countries)

Table 1 (Continued)

Study	System-level variables and/or country-case selection criteria	Individual-level dependent variables	Survey sources (*N* of upper units)
Iyengar et al. (2010)	Average percentage of hard vs. soft news, domestic vs. international news	Frequency of exposure to national newscasts	US, UK, Denmark, & Finland
Kalogeropoulos (2018)	Media system (per Hallin & Mancini)	Frequency of online news video use, use of video in social media platforms and the websites of news organizations	2016 Reuters Institute Digital News Report (6 countries)
Kalogeropoulos et al. (2017)	Media system (per Hallin & Mancini)	Online engagement with the news	2016 Reuters Institute Digital News Report (6 countries)
Lizardo and Skiles (2009)	For-profit vs. non-profit orientation of the national TV production field	Consumption of musical genres & popular TV programming	2001 Eurobarometer (15 countries)
Loveless (2015)	GDP per capita, World Bank governance score Laws and regulations influence media content, economic influence over media, political pressure and control over media, media trust	Media use prior to European elections	EES 2009 (27 countries)
Nielsen and Schrøder (2014)	Media system (per Hallin & Mancini)	Social media as sources of news (relative to other sources)	2013 Reuters Institute Digital News Report (8 countries)
Papathanassopoulos et al. (2013)	Media system (per Hallin & Mancini), media culture, political histories, measured as national contexts	Exposure to news media	Media System, Political Context and Informed Citizenship (11 countries)
Sehl et al. (2018)	Media system (per Hallin & Mancini)	The adaption of digital newsroom in the public service media (PSM)	Semi-structured interviews (6 countries)
Shehata and Strömbäck (2011)	Media system (per Hallin & Mancini) Newspaper vs. television centrism	Reading news about politics in newspapers Television news consumption about politics	ESS 2002, 2004, & 2006 (16 countries)

Table 1 (Continued)

Study	System-level variables and/or country-case selection criteria	Individual-level dependent variables	Survey sources (N of upper units)
Comparative studies on media effects			
Barnidge et al. (2018)	Freedom of expression index	Political expression on social media	20 countries
Clark (2014)	The percentage of positive and negative EU messages	EU-level and national-level political knowledge	Eurobarometer 61.0 EES 2009 (25 countries)
Elenbaas et al. (2014)	The number of different media sources (information saturation)	Acquisition of EU-level political information	2009 EP election survey (Denmark, Germany, Netherlands, UK)
Fraile and Iyengar (2014)	The availability of more hard news-oriented news sources	EU political knowledge	EES 2009 (27 countries)
Hameleers et al. (2018)	Unemployment rates per country The size of populist parties on the Left and Right	Political engagement	Comparative experiment across 16 countries
[a] Jerit et al. (2006)	The total amount of media attention (newswire, TV, newspaper) devoted to an issue	Political knowledge	Pooling of 41 public opinion surveys
Lelkes (2016)	Party-media parallelism Electoral disproportionality Economic development	Institutional trust Satisfaction with democracy	ESS 2006–2010 (28 countries)
Nir (2012)	Shares of readership of the largest newspaper Viewership of the most watched news program News visibility and reporting depth within country Number of effective parties	Political interest and knowledge	EES voter study and content analyses 2009 (13 countries)
Nisbet et al. (2012)	Freedom House's democracy rating Human development index Internet penetration	Citizens' demand for democracy	2008 Afrobarometer & 2006–2008 Asian Barometer (28 countries)
Schoonvelde (2014)	Country-level media freedom The audience share of public broadcasting channels The number of daily newspaper subscribers per capita Media market competitiveness	Political knowledge	CSES (31 country-year election results)

Table 1 (Continued)

Study	System-level variables and/or country-case selection criteria	Individual-level dependent variables	Survey sources (*N* of upper units)
Schuck et al. (2013)	The level of democratization	Political cynicism about the EP election campaign	2009 EP election survey (21 countries)
Schuck et al. (2016)	The presence of conflict framing in news regarding the EU and the EP elections	Voter mobilization	PIREDEU (27 countries)
Tsfati and Ariely (2014)	Country-level post-materialism index Level of economic development Media ownership structure	Media trust	WVS 2005–2008 (44 countries)
van Spanje and de Vreese (2014)	Visibility and evaluation of Eurosceptic actors Media evaluation of EU actors	Eurosceptic voting	2009 EP election survey (21 countries)
Wilson and Hobolt (2014)	Negative media tone Party polarization	Attribution of responsibility	EES 2009 (27 countries)

Studies denoted with ([a]) examine regional variations within U.S. System

System-level dependent variables are denoted in italics.

ESS European Social Survey, *NAES* The National Annenberg Election Survey, *NEPOCS* Network of European Political Communication Scholars, *CNEP* Comparative National Election Project, *CSES* The Comparative Study of Electoral Systems, *PIREDEU* Providing an Infrastructure for Research on Electoral Democracy in the European Union, *WVS* World Value Survey

Aalberg et al. (2013; also see Blekesaune et al. 2010) looked at European Social Survey data from 31 European countries, and found that different political regime classifications of countries—specifically Anglo-Saxon and Eastern regimes where a greater choice of television channels and programming is available—have significantly stronger media consumption gaps in terms of time spent watching television in general, versus consumption of news on politics and current affairs. The difference between the time spent watching television in general, and the time spent on news and current affairs, is therefore more/less pronounced in countries with a wider variety of television channels and programming. Papathanassopoulos et al. (2013) observation of 11 countries on four continents across the globe somewhat directly echoes this observation, documenting that complexities of information markets, such as high penetrations of broadband Internet, have a direct bearing on exposure to news at the individual level. Consistent with Prior's (2007) observation, they found that increasing media choices might have led to an increasing gap between exposure to entertainment and news consumption, but to different degrees depending on the availability of access to different information sources.

Explicitly formalizing how contextual differences across U.S. media markets might stimulate patterns of individual-level exposure, Althaus et al. (2009) examine relative contributions made by individual- and contextual-level predictors when it comes to explaining individual-level local and national news exposure. They found that market-level demographic compositions (such as the percentage of the white population, median household income, etc.), market size and differentiation, and the spatial structure of media markets (e.g. the spillover of broadcast signal propagation from neighboring media markets) uniquely influence individual-level TV news exposure. Within the context of Europe, and focusing on newspaper reading behavior, Elvestad and Blekesaune (2008) also found significant influences of certain national-level variations such as unemployment rates, population density, and most importantly media system differences (Northern European or democratic corporatist countries, per Hallin and Mancini's (2004) classifications) in explaining higher newspaper consumptions across Europe. Goldman and Mutz's (2011) investigation examined the extent of cross-cutting exposure from media (operationalized as the distance between perceived political bias of the media and respondents' own political viewpoint), and found that cross-cutting exposure through newspaper and television news programs is negatively related to higher levels of press-party parallelism at the national level. Overall, the studies reviewed above point to the importance of taking into account media system-related variations, as these directly affect consumption patterns across and within media genres.

While most of the aforementioned studies have generally shifted the theoretical focus from individual-level predictors to contextual/system-level predictors, Shehata and Strömbäck's (2011) observation regarding news consumption in 16 European countries represents another stream of thought in this tradition. In that vein, their analysis sheds light on methods to conceptualize and statistically model complex interactions between individual- and contextual-level factors in predicting one's media consumption. Utilizing a set of cross-level interactions, they provide insights into how contextual factors (e.g. a country's newspaper- vs. television-centrism) can further condition the influence of one's motivations and interests, with these factors

playing a stronger role in countries with more newspaper-centric media environments (such as northern European countries) than television-centric countries (as in Southern European countries). Using the 2009 European Election Study across 25 countries, Loveless' (2015) research also does a good job of exemplifying a typical theoretical setup of treating individual-level "media use" as a dependent variable, being explained by certain contextual variations focusing on system-level characteristics (e.g. media-system differences) and their interaction with individual-level correlates. His study finds some indication that individual-level political interest and national-level journalistic autonomy and professionalism significantly interact with each other in predicting media use. Similarly, in a recent study by Castro, Nir, and Skovsgaard (2018), the impact of an individual's political interest on his or her exposure to (politically) cross-cutting media contents is found to be contingent upon the relative strength of public service broadcasting. They found that the presence of strong public service broadcasting minimizes the opportunities for selectively avoiding non-like-minded contents, echoing the view that one's use of media is shaped by both opportunities (from contextual factors) and motivations (from individual factors). Similar approaches and findings are provided by, for example, Iyengar et al. (2010). Hence, in addition to media system characteristics bearing direct relevance as opportunity structures that would explain individual consumption, these studies emphasize the conditioning role of context in structuring the importance of individual-level factors, above all motivation and interest. In sum, this body of literature stresses the need to investigate media use patterns beyond the peculiarities of national media systems in order to come to a comprehensive understanding.

While the aforementioned studies of cross-national comparison of (individual-level) media use have constituted the majority of the research tradition, it is indeed not uncommon to find studies that go beyond simple exposure to traditional media, examining interactive, digital narrowcast (social) media. With the advent of alternative forms of media content consumption driven by technological advancements, in particular the spread of digital and mobile media, and by organizational changes in newsrooms across the globe that aim to facilitate news production for different media platforms, there has been a substantial increase in academic attention to the issue of social media and audiovisual consumption (Bright 2018; Kalogeropoulos 2018; Nielsen and Schrøder 2014), along with audience engagements in such new media platforms (Kalogeropoulos et al. 2017). Furthermore, spatial and temporal structural changes in media ecosystems—such as digital newsroom integration into public service media (Sehl et al. 2018), or changes in opportunity structures of information environment (Esser et al. 2012)—are increasingly subject to cross-national comparative analysis. Some recent work on this area has also examined the extent of audience fragmentation across different news media platforms (Fletcher and Nielsen 2017), or day-to-day media diets such as television programming (Lizardo and Skiles 2009) through the lens of comparative, cross-national research. For instance, Fletcher and Nielsen (2017) have found that, while online news audiences are not more fragmented than offline news audiences, there seems to be a higher degree of audience fragmentation in countries with media organizations that offer more diverse content with a high proportion of hard news (e.g. the United Kingdom and Denmark, compared to Spain and the United States), largely because of the dom-

inance of very strong, powerful sources with a very high reach. Similarly, Lizardo and Skiles (2009) have found that, in countries with relatively more commercialized, profit-oriented market systems (such as England, France, or Germany), consumers with a more highbrow taste are less likely to report having watched a broader range of television programming than those who reside in less commercialized markets (such as in Austria, Finland, and Denmark).

4 Comparative Research on Media Effects

Along with the study of media use and its correlates, studies of media "effects" are another fundamental aspect of communication research (for a broad review, see Bryant and Oliver 2009). While the bulk of traditional media effect studies tries to identify why and how certain media use (or exposure) would produce observed affective, cognitive, or behavioral outcomes, comparative studies of media effects tend to focus on their boundary conditions across different national-contextual differences. When and in which contextual circumstances media effects occur, or in which contexts the presence of media effects would diminish, are central questions in this area of research. Theoretically, this inclusion of context often takes the form of focusing on different "opportunity structures" represented by the contextual variations of, for instance, media systems, press-government relations, or the supply of certain types of media programming or contents across countries, on the one hand, and how such variations may in turn produce or further condition the effects of given media exposure on the outcome variables in question, on the other. Methodologically, this is often achieved by linking traditional survey data across countries with respective media content data, sometimes on both the aggregate and on individual levels ("linkage studies": de Vreese et al. 2017). Here, we also include any studies conceptualizing "media exposure" or "media use" as moderating factors of other theoretically relevant variables (such as political ideology) in explaining one's political cognitions or attitudes in this classification, given its theoretical focus of media consumption predicting a dependent variable of interest.

A considerable portion of prior work on this topic has examined the impact of one's mass media use on political knowledge or political engagement and, furthermore, how and why contextual variation across different national contexts can influence this link. These works are motivated at least in part by the observation that, while media exposure in general is related to political learning and, as a result, to political engagements (Carpini and Keeter 1996; Eveland 2001), it appears that the extent of such a relationship is highly variable across different geographical contexts, presumably contingent upon what is actually transmitted by different media (Fraile and Iyengar 2014). It therefore logically follows that any contextual variations in terms of journalistic norms, the degree of commercialization, or the sheer number of choices between news vs. entertainment—all of which can systematically affect the content and frequency of news programming—may affect the relationship between news media exposure and political knowledge and/or engagement.

Using 2009 European Election Study data from 13 EU Member States, Nir (2012) documents that the higher the country-level media fragmentation (operationalized

as the circulation/viewership share of the largest newspaper/TV news) across different EU Member States, the higher the knowledge gaps are between the top and bottom socioeconomic quartiles. This approach is directly echoed in Elenbaas et al. (2014) research, where they found the more "saturated" (i.e. the level of cross-media diffusion of the same information across different media) media environment additionality moderate the impact of one's motivation on political learning from media exposure. Yet they also have found that this moderating relationship does not appear to be strictly monotonic. Fraile and Iyengar (2014) also found that the availability of more hard news-oriented news sources (i.e. public broadcasting and broadsheet newspapers) increases the impact of weekly media exposure on citizens' level of political knowledge. Focusing on institutional factors across 31 countries around the globe, Schoonvelde (2014) found that the less governments interfere in media environments, the higher the levels of citizens' political knowledge as a function of personal educational attainments. Similarly, using Eurobarometer data, Clark (2014) observed that more media coverage of European politics at the national level increases citizens' knowledge about EU-related issues, thus confirming the expectation that variations of information "availability" or "opportunity" significantly shape personal knowledge above and beyond any individual-level predictors of political knowledge. Applying a similar logic yet focusing on the U.S. context, Jerit et al. (2006; also see Jerit 2009) also show that contextual variations of available information regarding a given issue positively predict one's knowledge, as does the gap in knowledge between the highly educated vs. the least well educated (for a similar application but in a nonpolitical domain issue, see Hwang and Southwell 2009). What becomes apparent here is that the combination of macrolevel indicators of media systems or information opportunity structures and individual-level characteristics such as media exposure or education levels provides for an opportunity to reconcile microlevel theories on learning from the media (e.g. Eveland 2001) with macrolevel theorizing related to the knowledge gap hypothesis (Tichenor et al. 1970).

European elections have been described as a natural "playground for comparative research" (de Vreese and Boomgaarden 2012, p. 328) and a "laboratory" for comparative social sciences, in particular for electoral research (van der Eijk and Franklin 1996). The fact that the same event takes place at the same time and at regular intervals in, as of 2004, more than 24 different political and media systems, makes a systematic engagement with the interplay of contextual and individual-level effects possible and necessary. Given that European election studies have been conducted ever since the 1999 election, including a comprehensive battery of media use questions, and given that these surveys have been supplemented in some years by large-scale content analyses of media coverage in national newspapers and television (de Vreese et al. 2006; Schuck et al. 2011) important insights have been generated regarding the effects of media use and their conditionality.

Some contributions in this area examine the interplay between characteristics of the national information environment and individual exposure to certain types of media contents, relying on a linkage approach. Regarding the effect of media on political engagement and voting behavior, Schuck et al. (2016) investigation provides a fairly comprehensive and innovative strategy to model contextual variations in the effect of media exposure on individuals' voting behavior, here turnout. Combining

a large-scale media content analysis of a sample of national news media across 27 EU Member States with representative panel surveys conducted in each country, they show that the level of conflict framing in media coverage of campaign news significantly varies across countries, and more importantly, the effect of conflict framing exposure on voter turnout significantly depends on EU polity evaluations in a given country (i.e. the tone toward the EU adopted in the national media as a whole). Public satisfaction with democracy was influenced by individual exposure to positive news about the democratic function of the EU, and this effect was found to be stronger in countries in which there was a dominantly positive message flow about the EU, so individual media exposure and the country information environment reinforced each other (Desmet et al. 2015). Once more utilizing a linkage technique, van Spanje and de Vreese (2014) found that citizens across the 27 EU Member States are more likely to support Eurosceptic parties when they are exposed to more negative coverage of the EU. Further, they found that, especially when mainstream parties of a given nation hold highly divergent stances with regard to European integration, the media exposure of the benefit ensuing from the EU significantly lowered the likelihood of voting for Eurosceptic parties.

Studies have also looked beyond the distribution of political knowledge and patterns of voting behavior across different contexts. Based on the seminal work of Jamieson and Cappella (1997), which argues in favor of a relationship between strategy framing in the news and political cynicism among the public, Schuck et al. (2013) have argued that this effect likely depends on the functioning of a country's political system. They argue that citizens' level of cynicism towards a European Parliament election campaign is shaped by both individual-level and contextual-level factors. They found that exposure to strategic, game-framed news (i.e. news articles depicting politics as strategic games among political actors) decreased political cynicism among the more politically engaged. At the same time, they also found that exposure to strategic, game-framed news significantly increased cynicism towards EP elections, especially within political contexts that are characterized by high levels of democratization and higher-quality governance (for a similar finding, see also Desmet et al. 2015). Moreover, while attitudes towards EU membership for Turkey were not driven by individual exposure to EU news coverage, it was shown that the effect of individuals' attitudes towards immigration on support for EU membership for Turkey was stronger in those countries in which the mass mediated information flow about the EU was predominantly negative (Azrout et al. 2012).

Yet another study drawing on European election data, carried out by Wilson and Hobolt (2014), examined citizens' attribution of responsibility between national vs. EU-wide governmental bodies using the 2009 European Election Study (EES) Voter Survey, coupled with the EES media study and their own expert survey. They found that, while political knowledge is positively correlated the with "correct" attribution of responsibility (measured as the discrepancy between citizens' attribution and expert opinions) regarding certain EU issues, a negative media tone and party polarization—which represents the politicization of topics and, therefore, opportunities for learning about such issues—increases the correct allocations of responsibility to the EU over national political actors. In sum, this set of studies, drawing as it does on data collection in the framework of PIREDEU for the 2009 European

Elections, demonstrates the importance of relying on multiple data sources beyond survey data. In particular, the fact that PIREDEU data included a systematic media content analysis in all EU Member States at that time, and that such content data could either be used to operationalize national information environments or to link up to individual media use measures, was an important impetus for multilevel work on the consequences of media for a range of political outcome variables.

Additional studies examine the consequences of media use on attitudes towards political or media systems in a comparative perspective. Tsfati and Ariely (2014) have found that trust towards media is significantly shaped by general trust, exposure to television news and to newspapers across 44 countries, while a post-materialistic culture, or the aggregate-level political culture stressing higher order needs such as individual freedom and self-expression over physical security or economic endurance, tends to lower people's trust in media. Similarly, Lelkes (2016)—using ESS data across 28 countries—reveals that the degree of press-party parallelism of media significantly conditions the effect of media consumption on trust vis-à-vis political institutions, and of whether an individual has supported the current governing coalition ("winner vs. loser").

Few studies have turned towards the consequences of using new media. Focusing on Sub-Saharan Africa and on Asia, Nisbet, Stoycheff, and Pearce (2012) have examined the impact of Internet use on attitudes toward regime legitimacy ("demand for democracy"). They found that, while individual-level Internet use increases citizens' demand for democracy, this relationship is more pronounced in countries with higher democratization and with higher Internet penetration rates at the contextual level. In a similar vein, Barnidge et al. (2018) found that the effects of individuals' social network heterogeneity in predicting political expression on social media are much higher for countries with less freedom of expression.

Lastly, a novel and promising aspect in the comparative communication literature is the conduct of identical experiments at the same time, in different national contexts. Taking such an approach, Hameleers et al. (2018) investigated the effects of populist communication messages on political engagement, and found that in particular anti-elitism cues in populist messages led to greater engagement in those countries in which the unemployment rate was high. While investment in and coordination of such comparative experimental designs are demanding, including the proper specification of stimulus material, they may be important avenues to move forward in terms of the generalizability of the causal structures demonstrated by experimentation.

5 Conclusion: Ways and Challenges Ahead

Communication, in particular explanations and consequences of media use, should be studied in conjunction between microlevel processes and macrolevel influences. While this was already acknowledged by communication scholars decades ago (McLeod and Blumler 1987; Paisley 1984; Pan and McLeod 1991), a systematic and statistical engagement with the interplay between these different levels of influence, based on large-N study designs, is still only a marginal (albeit growing)

phenomenon in the literature, as described by our review of the state-of-the-art. While the insights that have been generated offer some degree of theoretical convergence, and hence important insights into theorizing of communication processes, much work remains to be done. In what follows, we first discuss the generated empirical insights in a wider context, followed by a critical assessment of the state-of-the-art. We then conclude by highlighting the challenges faced by the comparative study of media use and media effects, and how the field could move forward.

A general conclusion from the findings sketched above is that context does indeed matter; it either structures individual cognitions, attitudes, and behaviors directly in a predictable way, or it further illuminates the boundary conditions of individual influence processes. In particular, the literature on antecedents and the consequences of media use—especially for political learning—consistently highlights the role played by media system characteristics. Media systems, when characterized by a high degree of journalistic professionalism, lesser press-party parallelism, or low commercialization of media markets (e.g. a stronger public broadcasting system) appear to lead directly to high levels of news consumption, or to knowledge gains in the realm of politics. By contrast, in media systems with higher commercialization or market segmentations, people are less likely to follow news and current affairs programming or learn about political processes. Such system characteristics are conceptualized and theorized as opportunity structures that facilitate (or disinhibit) exposure or learning (Aalberg et al. 2010; Curran et al. 2009; Fraile and Iyengar 2014). In some instances, such opportunities not only raise media use and learning processes directly, but also influence individuals' motivations or interests, which in turn trigger these outcomes. The substantial amount of studies in this particular area therefore converges towards a consistent picture of opportunity structures as an important driver of political learning from media. It might be possible to trace a consolidation of the theoretical underpinning of such theorizing to Hallin and Mancini's (2004) seminal discussion of national media systems.

Empirical evidence is more scattered in other areas, and arguably does not lead to more general theorizing. This particularly applies to studies that look at information environments, i.e., the concrete outputs of national media systems as conditioning the effects of individual media consumption on various types of "attitudes" (rather than learning). In such studies, the theorizing of the boundary conditions appears to tend less towards converging on a more general idea of why different information environments matter. Do they matter because of the fact that dominant characteristics of such information environments are competing against alternative sources of information such as interpersonal communication (Boomgaarden 2014; Schmitt-Beck 2003; Southwell and Yzer 2007), and hence interfere with the information that individuals encounter directly? Or do they matter because they make information that individuals receive directly more or less important, so that it can stand out among the dominant information flows? It seems that theorizing in such studies is more ad hoc, driven by the particular research interests or data availability. Based on the evidence, inferring a general picture of how and why information environments matter for media effects in more general terms—especially for its attitudinal consequences—seems a daunting task so far, and the field would do well to invest

more heavily in the conceptual and theoretical underpinnings that would apply more generally.

An important point to add here relates to the observation that the literature reviewed above, including much of the early calls for more comparative communication research that specifically deals with media use and its effects, mostly emanated from the subfield of political communication. While acknowledging that we might have neglected important prior works from other subfields of communication in the review process, it appears that political communication scholars have particularly embraced the use of multilevel models as a methodological toolkit for systematic comparative research of media use and media effects (see Schmidt-Catran et al. 2019 in this special issue). Our assumption is that this is largely due to the types of data that are available to this particular subdiscipline (often facilitated by cross-national collaborations of comparative election studies) and the theoretical impetus put forward by a number of seminal scholars in the field, most of whose intellectual origins can be traced back to the field of comparative politics (also see Schmitt-Beck 2019 in this special issue). Surely there has been an attempt to embrace systematic comparative inquiries in other subfields of communication science as well (see Esser and Hanitzsch 2012). It is arguably in its earlier stage—at least for large-N, multilevel comparative approaches. Our hope would be to see other areas of communication science following suit, areas in which contextual factors may be equally important, e.g. health communication, organizational communication, or media economics.

As in many other areas of comparative research, data availability appears to constitute a central hurdle to the further development of comparative studies of communication processes, and this issue has both genuine practical and conceptual/theoretical implications. Starting with the latter, the logic of case selection remains pretty much driven by (post hoc) data availability, and not by theoretical concerns that reveal important insights regarding the regularities and its contingencies of communication processes. This is for instance evident even when it comes to looking at media systems. On a global scale, media systems would offer much greater variation than is utilized in typical empirical studies in the field, which primarily—except for a few cases—look at European countries cross-sectionally. If the aim is to understand the boundary conditions imposed by different media systems on media use patterns or learning effects, it would be theoretically desirable to include a much wider range of media systems across a much broader geographical context, and also over time. This would allow researchers to examine the impact of (more meaningful) variations in the characteristics of interest at national levels. But such data is usually gathered externally (such as in the case of the European Social Survey or the European Value Survey), or driven by certain geographically confined events (such as in the case of European elections), which limits the types of macro–micro dynamics assessments to these limited cases only. Also, media systems data (not to speak of information environment data) are often more readily available for certain regions of the world than for others, which again imposes practical constraints.

But even if comparative data are available, the type of measurement does not always speak to the questions that communication scholars may be interested in addressing. Media use batteries in large-scale (comparative) surveys are oftentimes

much more limited than one hopes that they might be (Prior 2009b; Slater 2004), and often do not fully capture the types of usage that would be needed in order to investigate explanations or effects of type, content, or genre-specific consumption (Prior 2013). This imperfect measurement of media exposure is in particular true for media effects studies which wish to link exposure to specific media outlets and content data (i. e. linkage analysis: De Vreese et al. 2017), where it is often argued that it produces a severe downward bias in estimated media effects (Scharkow and Bachl 2017). Assessing outlet-specific or content-specific exposure requires tediously long and detailed measurements, and existing data collection efforts often do not allocate the appropriate space to do so. The few examples that have done such in relation to European elections have, however, generated important insights. Another aspect that comes into play here is the availability of media content data to either operationalize information environments, or to link up with individual media use. Conducting large-scale content analyses across multiple countries and languages, which would allow for multilevel modeling, have long been a labor-intensive and costly task, and except for the PIREDEU project, such an endeavor is very rare at best. However, the recent developments in automated content analysis and text-as-data approaches (Proksch et al. 2018; Lind et al. n.d.) are starting to provide more accessible alternatives in this regard.

A further aspect that also requires greater attention than it has hitherto received is the matter of conceptual equivalence, measurement equivalence, and (study) administration equivalence. *Conceptual equivalence* (Hui and Triandis 1985) denotes a representation of common, established and shared knowledge between the (proposed) theory and empirical phenomenon in question. While this constitutes a very basic condition necessary for making cross-cultural or cross-national comparisons, the comparability and equivalence of concepts and their interpretations have always been a problematic endeavor (see Wirth and Kolb 2004). This issue is sometimes discussed within the context of *meaning (semantic) equivalence* (i. e. whether the concept in question is interpreted similarly across different national contexts), within the context of *measurement equivalence* (regarding the operationalization and measurement of the concepts across contexts: also see Cieciuch et al. 2019 in this special issue), or in terms of the equivalence of study administration (regarding the standardization of research designs, instruments, and instructions). While all of these considerations aim to maximize the theoretical and empirical comparability of the findings at every stage of research, it is ironic that such multiple considerations often work in opposite directions. For instance, it is now a very common practice (at least for large-scale comparative studies) to impose identical wordings and study administrative procedures, yet imposing such restrictions does not necessarily yield or aid meaning and conceptual equivalence across all contexts. Many of the large-N, multilevel comparative studies that have been published to date are rather silent about this issue in evaluating their findings due to the (largely) secondary-analysis nature of the study. While this issue is arguably highly context- and research question-specific, it is our observation that a more thorough and systematic evaluation of this matter is needed.

In order to move the field forward, two additional aspects stand out. First, standardization of data collection and data sharing, both in terms of survey data, but

even more so relating to detailed media content data must be addressed. If individual projects, for instance national election studies, were to adhere to standardized, highly calibrated approaches to measure media use in their surveys and abide by these approaches in consecutive waves, combining national surveys into comparative data sets would offer a far greater potential beyond long-standing cross-national survey projects. Also, the latter should invest more in listening to the needs of communication scholars. Second, international cooperation now seems more and more imperative, with funding opportunities not only for networking (such as in COST action), but also for data collection. While there is still much to gain from large-N, multilevel comparative communication research, as has become evident from the review provided here, the road ahead is, at best, bumpy. Placing stronger emphasis on both the theoretical need to invest in comparative studies in order to establish the boundary of existing theorizing, in combination with establishing stronger infrastructures to enable this type of work to be done, would be needed in order to advance the field of comparative communication.

References

Aalberg, Toril, Peter van Aelst and James Curran. 2010. Media systems and the political information environment: A cross-national comparison. *The International Journal of Press/Politics* 15:255–271.

Aalberg, Toril, Arild Blekesaune and Eiri Elvestad. 2013. Media choice and informed democracy: Toward increasing news consumption gaps in Europe? *The International Journal of Press/Politics* 18:281–303.

Althaus, Scott L., and David Tewksbury. 2000. Patterns of Internet and traditional news media use in a networked community. *Political Communication* 17:21–45.

Althaus, Scott L., Anne M. Cizmar and James G. Gimpel. 2009. Media supply, audience demand, and the geography of news consumption in the United States. *Political Communication* 26:249–277.

Anduiza, Eva, Eva Anduiza Perea, Michael James Jensen and Laia Jorba. 2012. *Digital media and political engagement worldwide: A comparative study*. Cambridge University Press.

Azrout, Rachid, Joost van Spanje and Claes de Vreese. 2012. When news matters: Media effects on public support for European Union enlargement in 21 countries. *JCMS: Journal of Common Market Studies* 50:691–708.

Barnidge, Matthew, Brigitte Huber, Homero Gil de Zúñiga and James H. Liu. 2018. Social media as a sphere for "risky" political expression: A twenty-country multilevel comparative analysis. *The International Journal of Press/Politics* 23:161–182.

Bartsch, Anne, Roland Mangold, Reinhold Viehoff and Peter Vorderer. 2006. Emotional gratifications during media use – An integrative approach. *Communications* 31:261–278.

Beck, Paul Allen, Russell J. Dalton, Steven Greene, and Robert Huckfeldt. 2002. The social calculus of voting: Interpersonal, media, and organizational influences on presidential choices. *American Political Science Review* 96: 57–73.

Becker, Mark W., Reem Alzahabi and Christopher J. Hopwood. 2–13. Media multitasking is associated with symptoms of depression and social anxiety. *Cyberpsychology, Behavior, and Social Networking* 16:132–135.

Blekesaune, Arild, Eiri Elvestad and Toril Aalberg. 2012. Tuning out the world of news and current affairs—An empirical study of Europe's disconnected citizens. *European Sociological Review* 28:110–126.

Blumler, Jay. 1979. The role of theory in uses and gratifications studies. *Communication Research*, 6: 9–36.

Blumler, Jay G., and Michael Gurevitch. 1975. Towards a comparative framework for political communication research. In *Political communication: Issues and strategies for research,* edited by Steven H. Chaffee, 165–193. Beverly Hills, CA: Sage.

Blumler, Jay G., and Michael Gurevitch. 1995. *The crisis of public communication*. London, UK: Routledge.

Blumler, Jay G., Jack M. McLeod and Karl Erik Rosengren. 1992. *Comparatively speaking: Communication and culture across space and time*. Newbury Park, CA: Sage

Boomgaarden, Hajo G. 2014. Interpersonal and mass mediated political communication. In *Political communication*, ed. Carsten Reinemann, 469–488. Berlin: Walter de Gruyer.

Bright, Jonathan. 2018. Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication* 23:17–33.

Brüggemann, Michael, Sven Engesser, Florin Büchel, Edda Humprecht and Laia Castro. 2014. Hallin and Mancini revisited: Four empirical types of western media systems. Journal of Communication 64:1037–1065.

Bryant, Jennings, and Mary Beth Oliver. Eds. 2009. *Media effects: Advances in theory and research*. New York: Routledge.

Campbell, Scott W., and Nojin Kwak. 2010. Mobile communication and social capital: An analysis of geographically differentiated usage patterns. *New Media & Society* 12:435–451.

Carpini, Michael X. Delli, and Scott Keeter. 1996. *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Castells, Manuel. 2011. *The rise of the network society* (2nd ed.). Chichester, UK: John Wiley & Sons.

Castro, Laia, Lilach Nir and Morten Skovsgaard. 2018. Bridging gaps in cross-cutting media exposure: The role of public service broadcasting. *Political Communication*, online first.

Cho, Jaeho. 2011. The geography of political communication: Effects of regional variations in campaign advertising on citizen communication. *Human Communication Research* 37:434–462.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt and René Algesheimer. 2019. How to obtain comparable measures for cross-national comparisons. In Cross-national comparative research – analytical strategies, results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00598-7.

Clark, Nicholas. 2014. The EU's information deficit: Comparing political knowledge across levels of governance. *Perspectives on European Politics and Society* 15:445–463.

Curran, James. 2011. *Media and democracy*. New York: Routledge.

Curran, James, Shanto Iyengar, Anker Brink Lund and Inka Salovaara-Moring. 2009. Media system, public knowledge and democracy: A comparative study. *European Journal of Communication* 24:5–26.

De Vreese, Claes H., Susan A. Banducci, Holli A. Semetko and Hajo G. Boomgaarden. 2006. The news coverage of the 2004 European Parliamentary election campaign in 25 countries. *European Union Politics* 7:477–504.

De Vreese, Claes, and Hajo G. Boomgaarden. 2012. Comparing news on Europe: Elections and beyond. In *The Handbook of Comparative Communication Research*, ed. Frank Esser and Thomas Hanitzsch, 327–340. London, UK: Routledge.

De Vreese, Claes H., and Peter Neijens. 2016. Measuring media exposure in a changing communications environment. *Communication Methods and Measures, 10*: 69–80.

De Vreese, Claes H., Mark Boukes, Andreas Schuck, Rens Vliegenthart, Linda Bos and Yph Lelkes. 2017. Linking survey and media content data: Opportunities, considerations, and pitfalls. *Communication Methods and Measures* 11:221–244.

Delli Carpini, Michael X., and Scott Keeter. 1993. Measuring political knowledge: Putting first things first. *American Journal of Political Science* 37:1179–1206.

Desmet, Pieterjan, Joost van Spanje and Claes de Vreese. 2015. Discussing the democratic deficit: Effects of media and interpersonal communication on satisfaction with democracy in the European Union. *International Journal of Communication* 9:3177–3198.

Dilliplane, Susanna, Seth K. Goldman and Diana C. Mutz. 2013. Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science* 57:236–248.

Donohew, Lewis, Philip Palmgreen, and Jay D. Rayburn. 1987. Social and psychological origins of media use: A lifestyle analysis. *Journal of Broadcasting & Electronic Media* 31:255–278.

Edelstein, Alex S. 1982. *Comparative communication research*. Beverly Hills, CA: Sage.

Elenbaas, Matthijs, Claes de Vreese, Andreas Schuck and Hajo Boomgaarden. 2014. Reconciling passive and motivated learning: The saturation-conditional impact of media coverage and motivation on political information. *Communication Research* 41:481–504.

Elvestad, Eiri, and Arild Blekesaune. 2008. Newspaper readers in Europe: A multilevel study of individual and national differences. *European Journal of Communication* 23:425–447.

Esser, Frank, and Thomas Hanitzsch. 2012. On the why and how of comparative inquiry in communication studies. In *The handbook of comparative communication research*, ed. Frank Esser and Thomas Hanitzsch, 3–22. London, UK: Routledge.

Esser, Frank, and Barbara Pfetsch. 2004. Meeting the challenges of global communication and political integration: The significance of comparative research in a changing world. In *Comparing political communication: Theories, cases, and challenges*, ed. Frank Esser and Barbara Pfetsch, 384–410. New York: Cambridge University Press.

Esser, Frank, Claes H. de Vreese, Jesper Strömbäck, Peter Van Aelst, Toril Aalberg, James Stanyer, Günther Lengauer, Rosa Berganza, Guido Legnante, et al. 2012. Political information opportunities in Europe: A longitudinal and comparative study of thirteen television systems. *The International Journal of Press/Politics* 17:247–274.

Eveland, William P., Jr. 2001. The cognitive mediation model of learning from the news: Evidence from nonelection, off-year election, and presidential election contexts. *Communication Research* 28:571–601.

Fletcher, Richard, and Rasmus Kleis Nielsen. 2017. Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication* 67:476–498.

Fraile, Marta, and Shanto Iyengar. 2014. Not all news sources are equally informative: A cross-national analysis of political knowledge in Europe. *The International Journal of Press/Politics* 19:275–294.

Goldman, Seth K., and Diana C. Mutz. 2011. The friendly media phenomenon: A cross-national analysis of cross-cutting exposure. *Political Communication* 28:42–66.

Hallin, Daniel. C., and Paolo Mancini. 2004. *Comparing media systems: Three models of media and politics.* Cambridge, UK: Cambridge University Press

Hallin, Daniel. C., and Paolo Mancini. 2012. *Comparing media systems beyond the Western world.* New York, NY: Cambridge University Press.

Hameleers, Michael, Linda Bos, Nayla Fawzi, Carsten Reinemann, Ioannis Andreadis, Nicoleta Corbu, Christian Schemer et al. 2018. Start spreading the news: A comparative experiment on the effects of populist communication on political engagement in sixteen European countries. *The International Journal of Press/Politics* 23:517–538.

Hanitzsch, Thomas, and Frank Esser. 2012. Challenges and perspectives of comparative communication inquiry. In *The handbook of comparative communication research*, ed. Frank Esser and Thomas Hanitzsch, 501–516. London, UK: Routledge.

Hanitzsch, Thomas, Folker Hanusch, Claudia Mellado, Maria Anikina, Rosa Berganza, Incilay Cangoz, Mihai Coman, Basyouni Hamada, María Elena Hernández, Christopher D. Karadjov, Sonia Virginia Moreira, Peter G. Mwesige, Patrick Lee Plaisance, Zvi Reich, Josef Seethaler, Elizabeth A. Skewes, Dani Vardiansyah Noor and Edgar Kee Wang Yuen. 2011. Mapping journalism cultures across nations. *Journalism Studies* 12:273–293.

Hui, C. Harry, and Harry C. Triandis. 1985. Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology* 16:131–152.

Hwang, Yoori, and Brian G. Southwell. 2009. Science TV news exposure predicts science beliefs: Real world effects among a national sample. *Communication Research* 36:724–742.

Iyengar, Shanto, James Curran, Anker Brink Lund, Inka Salovaara-Moring, Kyu S. Hahn and Sharon Coen. 2010. Cross-national versus individual-level differences in political information: A media systems perspective. *Journal of Elections, Public Opinion and Parties* 20:291–309.

Jamieson, Kathleen Hall, and Joseph N. Cappella. 1997. Setting the record straight: Do ad watches help or hurt? *Harvard International Journal of Press/Politics* 2:13–22.

Jerit, Jennifer. 2009. Understanding the knowledge gap: The role of experts and journalists. *The Journal of Politics* 71:442–456.

Jerit, Jennifer, Jason Barabas and Toby Bolsen. 2006. Citizens, knowledge, and the information environment. *American Journal of Political Science* 50:266–282.

Kalogeropoulos, Antonis. 2018. Online news video consumption: A comparison of six countries. *Digital Journalism* 6:651–665.

Kalogeropoulos, Antonis, Samuel Negredo, Ike Picone and Rasmus Kleis Nielsen. 2017. Who shares and comments on news? A cross-national comparative analysis of online and social media participation. *Social Media+ Society* 3:1–12.

Katz, Elihu, Jay G. Blumler and Michael Gurevitch. 1973. Uses and gratifications research. *The Public Opinion Quarterly* 37:509–523.

Kim, Junghyun, Robert LaRose and Wei Peng. 2009. Loneliness as the cause and the effect of problematic Internet use: The relationship between Internet use and psychological well-being. *CyberPsychology & Behavior* 12:451–455.

Knobloch, Silvia. 2003. Mood adjustment via mass communication. *Journal of Communication* 53:233–250.

Knobloch-Westerwick, Silvia, and Steven B. Kleinman. 2012. Preelection selective exposure: Confirmation bias versus informational utility. *Communication Research* 39:170–193.

Lauf, Edmund. 2001. Research note: The vanishing young reader: Sociodemographic determinants of newspaper use as a source of political information in Europe, 1980–98. *European Journal of Communication* 16:233–243.

Lelkes, Yphtach. 2016. Winners, losers, and the press: the relationship between political parallelism and the legitimacy gap. *Political Communication* 33:523–543.

Lind, Fabienne, Jakob-Moritz Eberl, Tobias Heidenreich and Hajo Boomgaarden. 2018. Multilingual dictionary construction: Comparing different approaches for the measurement of migration frames. *Paper presented at the 68th Annual Conference of the International Communication Association*, Prague, Czech Republic.

Livingstone, Sonia. 1998. Mediated childhoods: A comparative approach to young people's changing media environment in Europe. *European Journal of communication* 13:435–456

Livingstone, Sonia. 2003. On the challenges of cross-national comparative media research. *European Journal of Communication* 18:477–500.

Lizardo, Omar, and Sara Skiles. 2009. Highbrow omnivorousness on the small screen? Cultural industry systems and patterns of cultural choice in Europe. *Poetics* 37:1–23.

Loveless, Matthew. 2015. Contextualizing media behavior: Media environments and individuals' media use in the European Union. *Central European Journal of Communication* 8:112–131.

McLeod, Jack, and Jay Blumler. 1987. The macrosocial level of communication science. In *Handbook of communication science*, ed. Charles R. Berger and Steven H. Chaffee, 271–322. Newbury Park, CA: Sage

McLeod, Jack M., and Nam-Ji Lee. 2012. Comparative research designs: Toward a multilevel approach. In *The handbook of comparative communication research*, ed. Frank Esser and Thomas Hanitzsch, 430–444. London, UK: Routledge.

McLeod, Jack M., Gerald M. Kosicki and Douglas M. McLeod. 2010. Levels of analysis and communication Science. In *The handbook of communication science* (2nd ed), ed. Michael E. Roloff, David R. Roskos-Ewoldsen and Charles R. Berger, 183–200. Thousand Oaks, CA: Sage Publications.

McLeod, Jack M., Dietram A. Scheufele and Patricia Moy. 1999. Community, communication, and participation: The role of mass media and interpersonal discussion in local political participation. *Political Communication* 16:315–336.

McLuhan, Marshall, and Bruce R. Powers. 1989. *The global village: Transformations in world life and media in the 21st century.* Oxford, UK: Oxford University Press

Nabi, Robin L., and Mary Beth Oliver. 2009. *The SAGE handbook of media processes and effects.* Thousand Oaks, CA: Sage.

Neuman, W Russell. 2016. *The digital difference: Media technology and the theory of communication effects.* London, UK: Harvard University Press.

Nielsen, Rasmus Kleis, and Kim Christian Schrøder. 2014. The relative importance of social media for accessing, finding, and engaging with news: An eight-country cross-media comparison. *Digital Journalism* 2:472–489.

Nir, Lilach. 2012. Public space: How shared news landscapes close gaps in political engagement. *Journal of Broadcasting & Electronic Media* 56:578–596.

Nisbet, Erik C., Elizabeth Stoycheff and Katy E. Pearce. 2012. Internet use and democratic demands: A multinational, multilevel model of Internet use and citizen attitudes about democracy. *Journal of Communication* 62:249–265.

Paisley, William. 1984. Communication in the communication sciences. In *Progress in the communication sciences*, ed. Brenda Dervin and Melvin Voigt, 1–43. Norwood, NJ: Ablex.

Pan, Zhongdang, and Jack M. McLeod. 1991. Multilevel analysis in mass communication research. *Communication Research* 18:140–173.

Papacharissi, Zizi, and Alan M. Rubin. 2000. Predictors of Internet use. *Journal of Broadcasting & Electronic Media* 44:175–196.

Papathanassopoulos, Stylianos, Sharon Coen, James Curran, Toril Aalberg, David Rowe, Paul Jones, Hernando Rojas and Rod Tiffen. 2013. Online threat, but television is still dominant: A comparative study of 11 nations' news consumption. *Journalism Practice* 7:690–704.

Pfetsch, Barbara, and Frank Esser. 2014. *Political communication in comparative perspective: Key concepts and new insights* (Vol. 18), 97–105. Berlin: De Gruyter Mouton.

Popescu, Marina, Gabor Toka, Tania Gosselin and Jose Santana Pereira. 2011. *European media systems survey 2010: Results and documentation.* Research report. Colchester, UK: Department of Government, University of Essex. http://www.mediasystemsineurope.org (Date of access: 20 Nov. 2018).

Potter, W. James. 2011. Conceptualizing mass media effect. *Journal of Communication* 61:896–915.

Price, Vincent, and John Zaller. 1993. Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly* 57:133–164.

Prior, Markus. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections.* New York: Cambridge University Press.

Prior, Markus. 2009a. Improving media effects research through better measurement of news exposure. *The Journal of Politics* 71:893–908.

Prior, Markus. 2009b. The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly* 73:130–143.

Prior, Markus. 2013. The challenge of measuring media exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication* 30:620–634.

Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle and Stuart Soroka. 2018. Multilingual sentiment analysis: A new approach to measuring conflict in parliamentary speeches. *Legislative Studies Quarterly.* https://doi.org/10.1111/lsq.12218

Putman, Robert D. 2000. *Bowling alone: The collapse and revival of American community.* New York: Simon & Schuster.

Roe, Keith. 2000. Adolescents' media use: a European view. *Journal of Adolescent Health* 27:15–21.

Ruggiero, Thomas E. 2000. Uses and gratifications theory in the 21st century. *Mass Communication & Society* 3:3–37.

Scharkow, Michael, and Marko Bachl. 2017. How measurement error in content analysis and self-reported media use lead to minimal media effect findings in linkage analyses: A simulation study. *Political Communication* 34:323–343.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In Cross-national comparative research – analytical strategies, results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie.* Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-006907-9.

Schmitt-Beck, Rüdiger. 2003. Mass communication, personal communication and vote choice: The filter hypothesis of media influence in comparative perspective. *British Journal of Political Science* 33:233–259.

Schmitt-Beck, Rüdiger. 2019. Political systems and electoral behavior: A review of internationally comparative multilevel research. In Cross-national comparative research – analytical strategies, results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie.* Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Schoonvelde, Martijn. 2014. Media freedom and the institutional underpinnings of political knowledge. *Political Science Research & Methods* 2:163–178.

Schuck, Andreas R., Hajo G. Boomgaarden and Claes H. de Vreese. 2013. Cynics all around? The impact of election news on political cynicism in comparative perspective. *Journal of Communication* 63:287–311.

Schuck, Andreas R., Rens Vliegenthart and Claes H. de Vreese. 2016. Who's afraid of conflict? The mobilizing effect of conflict framing in campaign news. *British Journal of Political Science* 46:177–194.

Schuck, Andreas R., Georgios Xezonakis, Matthijs Elenbaas, Susan A. Banducci and Claes H. de Vreese. 2011. Party contestation and Europe on the news agenda: The 2009 European Parliamentary Elections. *Electoral Studies* 30:41–52.

Sears, David O., and Jonathan L. Freedman. 1967. Selective exposure to information: A critical review. *Public Opinion Quarterly* 31:194–213.

Sehl, Annika, Alessio Cornia, Lucas Graves, and Rasmus Kleis Nielsen. 2018. Newsroom integration as an organizational challenge: Approaches of European public service media from a comparative perspective. *Journalism Studies*, online advanced publication. https://doi.org/10.1080/1461670X.2018.1507684

Seymour-Ure, Colin. 1974. *The political impact of mass media.* Beverly Hills, CA: Sage.

Shah, Dhaven V., Nojin Kwak and R. Lance Holbert. 2001. "Connecting" and "disconnecting" with civic life: Patterns of Internet use and the production of social capital. *Political Communication* 18:141–162.

Shehata, Adam, and Jesper Strömbäck. 2011. A matter of context: A comparative study of media environments and news consumption gaps in Europe. *Political Communication* 28:110–134.

Siebert, Fred, Theodore Bernard Peterson, Theodore Peterson and Wilbur Schramm. 1956. *Four theories of the press: The authoritarian, libertarian, social responsibility, and Soviet communist concepts of what the press should be and do*. Chicago, IL: University of Illinois press.

Slater, Michael D. 2004. Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism & Mass Communication Quarterly* 81:168–183.

Slater, Michael D. 2007. Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory* 17:281–303.

Slater, Michael D. 2015. Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. *Media Psychology* 18:370–395.

Slater, Michael D., Leslie Snyder and Andrew F. Hayes. 2006. Thinking and modeling at multiple levels: The potential contribution of multilevel modeling to communication theory and research. *Human Communication Research* 32:375–384.

Southwell, Brian G., and Marco C. Yzer. 2007. The roles of interpersonal communication in mass media campaigns. *Communication Yearbook* 31:419–462.

Teune, Henry, and Adam Przeworski. 1970. *The logic of comparative social inquiry*. New York: Wiley.

Tichenor, Phillip J., George A. Donohue and Clarice N. Olien. 1970. Mass media flow and differential growth in knowledge. *Public Opinion Quarterly* 34:159–170.

Tsfati, Yariv, and Gal Ariely. 2014. Individual and contextual correlates of trust in media across 44 countries. *Communication Research* 41:760–782.

Valkenburg, Patti M, and Jochen Peter. 2013. The differential susceptibility to media effects model. *Journal of Communication* 63:221–243.

Van der Eijk, Cees, and Mark N. Franklin. 1996. *Choosing Europe? The European electorate and national politics in the face of union*. Ann Arbor, MI: University of Michigan Press.

Van Spanje, Joost, and Claes de Vreese. 2014. Europhile media and Eurosceptic voting: Effects of news media coverage on Eurosceptic voting in the 2009 European parliamentary elections. *Political Communication* 31:325–354.

Vliegenthart, Rens. 2012. Analyzing comparative data: Opportunities and challenges. In *The handbook of comparative communication research*, ed. Frank Esser and Thomas Hanitzsch, 486–500. London, UK: Routledge.

Vorderer, Peter, Christoph Klimmt and Ute Ritterfeld. 2004. Enjoyment: At the heart of media entertainment. *Communication Theory* 14:388–408.

Wang, Zheng, John M. Tchernev and Tyler Solloway. 2012. A dynamic longitudinal examination of social media use, needs, and gratifications among college students. *Computers in Human Behavior* 28:1829–1839.

Wilson, Traci L., and Sara B. Hobolt. 2014. Allocating responsibility in multilevel government systems: Voter and expert attributions in the European Union. *The Journal of Politics* 77:102–113.

Wirth, Werner, and Steffen Kolb. 2004. Designs and methods of comparative political communication research. In *Comparing political communication: Theories, cases, and challenges*, ed. Frank Esser and Barbara Pfetsch, 87–111. New York: Cambridge.

Zillmann, Dolf. 1988. Mood management through communication choices. *American Behavioral Scientist* 31:327–341.

Hajo G. Boomgaarden 1977, Ph.D, Prof., Dean of the Faculty of Social Sciences and Professor of Methods in the Social Sciences with a focus on Text Analysis at the Department of Communication, University of Vienna. Areas of research: the coverage and effects of political information on citizens' cognitions, attitudes and behaviors in various domains of media and politics, and developments in automated content analysis techniques. Publications: One bias fits all? Three types of media bias and their effects on party preferences. Communication Research 44, 2015 (with J-M. Eberl and M. Wagner).

Hyunjin Song 1983, Ph.D, University Assistant, Department of Communication, University of Vienna. Areas of research: the influence of interpersonal discussion on political engagement, advanced computational methods such as statistical modeling of social networks. Publications: The dynamics of message selection in online political discussion forums: Self-segregation or diverse exposure? Communication Research 2018 (with J. Cho and G. A. Benefield).

Check for
updates

Political Systems and Electoral Behavior: a Review of Internationally Comparative Multilevel Research

Rüdiger Schmitt-Beck

Abstract A significant body of internationally comparative research has been accumulated over the past decade that seeks a better understanding of electoral behavior by combining survey data on voters with system-level data on their countries in complex multilevel designs. The current paper offers a state-of-the-art review of this rapidly evolving landscape of research, which thus far has been mainly guided by two questions: (i) Which conditions promote voter turnout that is high and more egalitarian, thus giving citizens an equal say in politics? And (ii) Which conditions promote electoral choices that are in line with voters' own interests, thus enhancing the role of elections as instruments for holding governments to account? While some studies primarily help to consolidate the field of comparative electoral behavior by taking more appropriate methodological approaches, others demonstrate with unprecedented clarity how individuals' voting behavior is systematically moderated by institutional as well as socioeconomic features of country contexts. Whether and how you select among electoral candidates depends critically on where you live—this is the powerful message of cross-national multilevel research on voting behavior. By identifying important sources of heterogeneity in voters' decision-making, this line of research profoundly questions the homogeneity assumption that has been a hallmark of electoral studies for decades.

Keywords Cross-nationally comparative electoral research · Electoral studies · Multilevel analysis · Turnout · Vote choice

R. Schmitt-Beck (✉)
Chair of Political Science—Political Sociology, University of Mannheim
68131 Mannheim, Germany
E-Mail: schmitt-beck@uni-mannheim.de

Springer

Politische Systeme und Wählerverhalten: eine Diskussion der international vergleichenden Mehrebenenforschung

Zusammenfassung In den letzten zehn Jahren wurde eine beträchtliche Anzahl international vergleichender Studien vorgelegt, die versuchen, durch komplexe Mehrebenendesigns, welche Daten aus Bevölkerungsumfragen mit Systemattributen auf Länderebene kombinieren, ein besseres Verständnis des Wählerverhaltens zu erreichen. Der Aufsatz stellt diese sich rasch entwickelnde Forschungslandschaft dar und diskutiert ihre Erträge. Bislang standen zwei Fragen bei dieser Forschung im Vordergrund: Welche Bedingungen fördern eine hohe und egalitäre, die Gleichheit politischer Mitsprache sichernde Wahlbeteiligung? Welche Bedingungen begünstigen Wahlentscheidungen, die mit den Interessen der Wähler in Einklang stehen, und stärken dadurch die Rolle von Wahlen als Instrument, um die Verantwortlichkeit von Parteien und Regierungen gegenüber den Wählern zu gewährleisten? Während einige Studien durch verbesserte methodische Ansätze in erster Linie zur Konsolidierung der international vergleichenden Wahlforschung beigetragen haben, zeigen andere mit größerer Klarheit als frühere Untersuchungen, wie das individuelle Entscheidungsverhalten bei Wahlen systematisch durch institutionelle sowie sozioökonomische Merkmale von Länderkontexten beeinflusst wird. Ob und wie man wählt, hängt entscheidend davon ab, wo man lebt – dies ist die starke Botschaft der international vergleichenden Mehrebenenforschung zum Wählerverhalten. Indem sie wichtige Quellen der Heterogenität in der Entscheidungsfindung der Wähler identifiziert, stellt sie die Homogenitätsannahme infrage, die seit Jahrzehnten ein Kennzeichnen von Wahlstudien gewesen ist.

Schlüsselwörter International vergleichende Wahlforschung · Wahlstudien · Mehrebenenanalyse · Wahlbeteiligung · Wahlentscheidung

1 Introduction

All the world's democracies are representative systems where governments are determined by means of general elections at which parties and candidates compete for citizens' votes. In all countries, participating in the selection of leaders is the most important way—and in some countries the only way—in which citizens can directly and effectively influence the course of politics and public policies. Besides referenda, elections are the sole event where democratic governance in a literal sense manifests itself as "government by the people". The fixed duration of electoral cycles is to make sure that governance also fulfills the criteria of being "government of" and "government for the people." Because of this central role for democratic politics, political scientists have always accorded special attention to elections.

Voting behavior has not been among the pioneer applications of the emerging scholarly world of cross-nationally comparative research on how system attributes affect individuals' attitudes and behaviors. Compared to other areas of study in citizen politics (Dalton 2014), such as political culture (Almond and Verba 1963) or political participation (Barnes et al. 1979), electoral research took longer to move to-

ward genuinely internationally comparative approaches (Schmitt-Beck 2015). This can be explained to some extent by peculiarities of its object, which consists of choices taking place according to their own national schedules and concerning entities that, by definition, exist only in one country. However, after a somewhat late start, the literature in this field has virtually "exploded in recent years" (Franklin et al. 2014, p. 400).

This development was made possible by the emergence of a rich reservoir of cross-national survey programs and the development of sophisticated statistical tools for analyzing data that are hierarchically nested into two or more levels of observation, facilitated by the high level of intellectual integration of scholarship on elections and spurred by recent improvements in the theoretical understanding of the contextual embeddedness of political behavior. As a result, numerous studies have been published during the past decade that applied techniques of multilevel modeling to investigate voters' turnout and decision-making. A significant body of internationally comparative research has meanwhile been accumulated that seeks a better understanding of electoral behavior by combining survey data on voters with system-level data on their countries in complex multilevel designs.

The present paper attempts a state-of-the-art review of this rapidly evolving research landscape. Since it needs to cover a lot of ground, its view on this landscape must necessarily be a high-altitude view. The picture it paints is one of broad strokes with little detail. The studies it covers are mainly guided by two questions: (i) Which conditions promote turnout that is high and more egalitarian, thus giving citizens an equal say in politics? And (ii) Which conditions promote electoral choices that are in line with voters' own interests, thus strengthening the role of elections as instruments for holding governments to account? The knowledge generated during the past decade has improved the scholarly understanding of the democratic role of elections in several ways. To be sure, not all the results are substantively surprising. Some studies primarily contribute to consolidating the field with more appropriate methodological approaches. This is of course a welcome development. More exciting are those studies that demonstrate in unprecedented clarity how individual electoral behavior is moderated by institutional as well as socioeconomic features of countries as contexts of voters. By identifying important sources of heterogeneity in voters' decision-making, such studies profoundly question the homogeneity assumption that has been a hallmark of electoral studies for decades (Rivers 1988). The following review starts with a general discussion of the development, promises, and achievements of cross-national multilevel research on electoral behavior. The subsequent sections turn more specifically to the state of research on turnout and voters' decision-making.

2 Cross-National Multilevel Research into Electoral Behavior: an Emerging Field of Study

2.1 Towards Cross-National Multilevel Approaches in Electoral Research

Why are cross-nationally comparative multilevel approaches an attractive perspective for the study of electoral behavior? Turnout and vote choice are the main dependent variables of research into citizens' behavior at elections. Its main purpose is to improve our understanding of the mechanisms by which these decisions come about. Since these decisions are taken by individual voters and then aggregated into election results according to the principle of "one person, one vote," the *micro level* is the appropriate level for approaching electoral behavior.

However, casting votes is a highly institutionalized form of political behavior. First off, it presupposes the existence of two institutions: periodic competitive elections and the franchise. More importantly, by definition, it can only occur in ways that are preconfigured by institutions. The most obvious ones are the institutional architecture of *systems of government* that determine for which offices elections are held in the first place; *electoral laws* that regulate the conditions under which individuals are allowed to vote, how votes are cast, and how elective offices are allocated on the basis of election outcomes; and *party systems* that structure the supply of competitors among whom voters can choose (specifically on the latter, see also Spies and Franzmann 2019 in this special issue). These and many other potentially relevant institutional preconditions vary widely across the world's democracies. Therefore, it seems almost trivial to expect electoral behavior to depend heavily on the institutional settings within which it takes place.

Nevertheless, although occasionally paying lip service to the institutional embeddedness of voting behavior, electoral studies have been dominated for most of their history by single-country studies that worked with purely individualistic models of turnout and vote choice. However, an awareness has grown during the past two decades that this "aggregate-psychological" approach (Esser 1999, pp. 419–421) can only provide an incomplete understanding. Conceiving of turnout and electoral choice as micro-level phenomena pure and simple entails the risk of "individualistic fallacies," i.e., erroneous inferences from micro-analytical evidence to the level of political systems (Scheuch 1969).

An important stimulus for this growing awareness of the *relevance of contexts* came from studies of social network effects on individuals' political behavior. They challenged the notion of citizens as independent, autonomous actors, instead advocating a perspective envisaging individuals as socially embedded and interdependent (Eulau 1986; Huckfeldt and Sprague 1995). Echoing a classic but neglected motive of sociological thinking (Esser 1999, pp. 415–462), proponents of this "social logic of politics" (Zuckerman 2005) argued that in order to fully comprehend individuals' attitudes and behaviors, researchers need to take note of the experiences to which these persons are exposed by their social environments. As an eye-opener, this research contributed significantly to raising scholarly awareness of citizens being "nested" (Anderson 2009) within contexts of relevance for vote choice and turnout (McClurg et al. 2017). However, its understanding of context is rather nar-

row, restricted to the structuring of interactions within socio-spatially circumscribed environments (Huckfeldt and Sprague 1995, p. 9).

Internationally comparative electoral research requires a wider conception of context. To address not only the "social logic" but also what might be called the "institutional logic" of politics, the notion of context must be conceived in a more abstract way, as encompassing social and spatial arenas of experience into which individuals are embedded and that, by structuring opportunities and constraints, create expectations and incentives for these individuals' political behavior (Anderson 2007b, 2009; Dalton and Anderson 2011b; Friedrichs and Nonnenmacher 2014; Lubbers and Sipma 2017; Schoen et al. 2017). As highlighted by Anderson (2009, p. 323), this understanding nicely accommodates the notion of political institutions as formally (or informally, for that matter) imposed constraints on human behavior. Cross-nationally comparative research into voting behavior is thus by necessity primarily interested in how political institutions influence people's processes of deciding about whether to vote and who to choose at an election. But to a lesser extent, it also takes account of the potentially influential role of socioeconomic features of contexts. It is thus located at the intersection of the study of political behavior and comparative politics, promising new insights for both sub-disciplines of political science (Anderson 2009). *Obviously, a multilevel perspective that simultaneously takes the micro level of individual voters and the macro level of political systems into account is thus the appropriate way to fully comprehend citizens' behavior at elections.*

While awareness of the contextual embeddedness of electoral behavior has thus gradually been growing, methodological progress provided tools that made it possible to cash in on the promise of this new perspective. During the past two decades, a veritable *"globalization" of survey research* (Heath et al. 2005; Norris 2009) has turned a situation of scarcity of individual-level data into one of abundance (Kaase 2015). Since studying national elections in an internationally comparative framework poses special challenges, it cannot come as a surprise that the pioneer project focused instead on elections for a supranational legislative body: the European Parliament (*European Election Study*, ESS). Heroic efforts were also undertaken to bring to life hitherto latent potentials of existing national election studies by harmonizing and pooling voter surveys from different countries (Thomassen 2005). Most importantly, since the mid-1990s, the *Comparative Study of Electoral Systems* (CSES) has compiled a treasure trove of specially collected (post-election) survey data in about 60 countries (Klingemann 2009). The CSES quickly turned into the single most important source of data for the cross-national study of electoral behavior.[1]

To be brought to fruition for cross-national comparisons that follow Przeworski and Teune's (1970, p. 8) maxim of replacing country names by variables, survey data must be amended with *macrodata* on system attributes. While researchers sometimes use custom-tailored instruments especially designed for their research, most studies rely on tried and tested standard indices taken from the toolbox that has been developed in comparative politics in recent decades. Measures of features of party systems, electoral systems, and governmental systems such as Laakso and Taagespera's (1979) effective number of parties, Dalton's (2008) index of party–system

---

[1] All studies discussed in the next two sections are based on CSES data unless explicitly stated otherwise.

Table 1 Methodological details of most important studies: turnout

Study	System variables	Moderator variables	Surveys (number of countries)
Brockington (2009)	Ideological spread of party system	–	CSES (28)
Chen (2011)	Typology of electoral systems	–	CSES (~30)
Córdova and Rangel (2017)	Compulsory voting	Gender	CSES (44)
Gallego (2010)	Compulsory voting Voter registration system Effective number of parties	Education	Pooled ESS and CSES (28)
Gallego (2015)	Compulsory voting Ballot structure Coalition governments Media system classification Public TV market share Trade union density Gini index	Education	CSES (36)
Singh (2011)	Proportionality of electoral system	Left–right proximity	CSES (19)
Singh (2015)	Compulsory voting	Age Political knowledge Income Efficacy Partisanship	CSES (35)
Solt (2008)	Gini index	Income	Pooled Eurobarometer, ISSP, CSES, and ESS (23)

CSES Comparative Study of Electoral Systems, *ESS* European Social Survey, *ISSP* International Social Survey Programme

polarization, Gallagher's (1991) index of the disproportionality of electoral systems, and Lijphart's (1999) indices of majoritarian versus consensus democracy are widely used in multilevel studies of electoral behavior. Tables 1 and 2 provide overviews of the indicators used by the most important publications reviewed in the following sections.

The concomitant development of sophisticated procedures *for analyzing data that are hierarchically nested* into two or more levels of observation created the tools for efficiently and adequately dealing with the statistical challenges posed by such data. Although two-step approaches (Jusko and Shively 2005) and pooled interactive regression analysis (Franzese 2005) found a fair number of applications, hierarchical modeling (Steenbergen and Jones 2002; Lubbers and Sipma 2017; Schmidt-Catran 2019) has, over time, become the most widely used technique for cross-national multilevel investigations of electoral behavior (see also Spies and Franzmann 2019). This method is uniquely suited for dealing with two kinds of research questions: whether and how properties of higher-level units, such as countries, influence the means and distributions of individual-level dependent variables, over and beyond the effects of persons' own attributes (random intercept models in the language of hierarchical modeling), and whether and how such higher-order phenomena af-

Table 2 Methodological details of the most important studies: vote choice

Study	System variables	Moderator variables	Surveys (number of countries)
Vote for incumbent			
Becher and Donnely (2013)	Gross domestic product Unemployment	–	Pooled NES (18)
Fischer and Hobolt (2010)	Coalition government	Government performance evaluation	CSES (33)
Giger (2011)	Clarity of responsibility Campaign agenda	Government performance evaluation on social policy	CSES (19)
Hellwig (2011)	Distribution of government responsibility Party system polarization	Government performance evaluation Left–right proximity	CSES (35)
Kayser and Peress (2012)	Local and global Gross domestic product Local and global unemployment	–	CSES (18)
Marinova (2016)	Party instability	Evaluations of government economic performance	CSES (27)
Party choice			
Duch and Stevenson (2008)	Unemployment Inflation Gross domestic product Trade openness Size of state sector Index of corporatism Index of regulation density Parliamentary vs. presidential system	Parties' share of administrative responsibility	Pooled NES (18)
Gingrich (2014)	Welfare spending mix	Left–right proximity Redistribution preferences	CSES (21) and ISSP (18)
Kedar (2009)	Single-party cabinets Effective number of parliamentary parties District size Control over parliamentary agenda	Left–right proximity	Pooled CSES and NES (13)
Klingemann and Weßels (2009)	Index of differentiation of political supply Index of effectiveness of electoral institutions	Left–right proximity Party-liking Candidate-liking	CSES (37)

Table 2 (Continued)

Study	System variables	Moderator variables	Surveys (number of countries)
Kroh (2009)	Effective number of parties in parliament Effective number of parties in government Party system polarization Ideological concentration of parties	Left–right proximity Political knowledge	CSES (30)
Wagner (2013)	Effective district size	Left–right proximity Party size	CSES (35)
Weßels and Schmitt (2014)	Consensus vs. majoritarian democracy Clarity of parties' policy positions	Left–right proximity	CSES (23)
Propensities to vote			
Lachat (2008)	Party system polarization	Left–right proximity Political sophistication Partisanship	ESS (13)
Van der Brug et al. (2007)	Unemployment Inflation GDP Concentration of policymaking responsibility Consensus vs. majoritarian democracy Welfare type	Party attributes (incumbency, left–right, size) Voter sophistication Welfare dependence National electoral cycle Length of incumbency	EES (15)
Other aspects of vote choice			
Blais and Gschwend (2011)	Disproportionality of electoral system	Party size	CSES (27)
Lau et al. (2014)	Age of democracy Density of media environment Candidate vs. party vote in electoral system Effective number of parties Polarization of party system Clarity of responsibility	Age Efficacy	CSES (35)
Singh (2010)	District size Ideological dispersion of parties Effective number of parties Compulsory voting Freedom House political rights and civil liberties score	–	CSES (42)

Note: In most studies the number of system-level cases is larger than the number of countries because several elections from the same countries are included and the models effectively operate on country-election cases

CSES Comparative Study of Electoral Systems, *ESS* European Social Survey, *NES* National Election Studies

fect associations between micro-level variables in cross-level interactions (so-called random slope models).

2.2 Analytical Promises

In an attempt to systematize the research agenda of comparative multilevel electoral research, Anderson (2007b, pp. 594–597; 2009; Anderson and Dalton 2011) suggested a typology of ways in which contextual conditions can affect individual-level political behavior. He speaks of "direct effects" when political behavior is immediately affected by contextual conditions, and of "indirect effects" when system attributes affect some intervening variable, which in turn influences the behavior of interest. These mediating factors may include beliefs and attitudes that lead to certain behaviors, but also meso-level phenomena such as parties' differential responses to varying contextual conditions, for instance styles of voter mobilization under different electoral systems. "Contingent effects" pertain to situations where the impact of contextual conditions on political behavior is moderated by citizens' personal attributes, or where the effects of micro-level predictors on dependent variables are strengthened or weakened by context-related attributes.

- From a methodological point of view, the notion of *direct effects* corresponds to the logic of random intercept models. In electoral research, many recent analyses of this type can in essence be qualified as sophisticated, methodologically sounder *replication studies*. Although sample surveys have been the dominant source of electoral data for half a century, macro analyses of aggregate data have remained popular in some areas of electoral research. In particular, this concerns broadly comparative studies of turnout that encompass large numbers of elections and countries (for reviews see Cancela and Geys 2016; Stockemer 2017). Regarding vote choice, aggregate data have traditionally played a prominent role in the economic voting literature (Lewis-Beck and Stegmaier 2000). Both fields share an early interest in institutional effects on electoral behavior. In addition, economic voting is concerned with the electoral implications of a phenomenon that by its nature is a constant at any single election, and thus not readily amenable to survey studies: the state of the economy. Hence, for lack of alternatives, aggregate data analysis appeared the most appropriate way to go.

  Studies of this kind obviously run the risk of falling into the trap of the ecological fallacy, that is drawing erroneous inferences from aggregate to individual-level relationships (Goodman 1953). Against this background, multilevel methods offer the unmistakable advantage of modeling the assumed processes of electoral decision-making in methodologically sound ways. Accordingly, much of the multilevel research that has been carried out does not necessarily aim to unearth innovative findings, but rather to validate known findings from extant macro research by using more adequate methods. This literature is not necessarily a constant source of surprises, but it places existing knowledge on firmer ground.

- The notion of direct effects is arguably an artificial one. Referring to the meta-theoretical maxims of methodological individualism, it can be argued that all effects of contextual conditions on individuals' electoral behavior are necessarily medi-

ated by individual perceptions, beliefs, and attitudes (Coleman 1990, pp. 1–23). From that perspective, demonstrating direct effects of contextual conditions on individual behavior amounts to an incomplete modeling of *indirect* effects that skips the mediating factors. This is illustrated by research on turnout. It is often hypothesized that contextual conditions, such as registration laws, influence voters' likelihood to participate in elections by increasing or decreasing the effort associated with the act of voting (see below). Such macro-micro-micro *chains of mediation* are, however, typically theorized, but not demonstrated. Hence, much of the literature suffers from a rather wide gap between the theoretical propositions tested and the actual modeling of the surmised processes.

It can be argued that such mechanisms can in principle always be demonstrated, although it requires the right data and adequate modeling tools. Ambitious studies that move beyond demonstrating direct effects and aim to clarify processes of mediation, thus demonstrating how direct effects are, in essence, indirect effects, are still rare. This may have something to do with problems relating to data availability, especially the scarcity of individual-level indicators of the effort involved in voting, but also the methodological challenges posed by analyses of this kind. Such research ideally requires methods that make it possible to decompose the total effect of a macro predictor on the final outcome variable into a mediation effect that runs through a micro-level variable which intervenes between the macro phenomenon in question and the outcome variable, and a direct effect of the macro predictor on the individual-level outcome variable that captures other intervening mechanisms (Becher and Donnelly 2013).

- Investigations of *contingent effects* are concerned with cross-level interactions. Most turnout studies of this type refer to the well-known problem of socioeconomically unequal participation (Verba and Nie 1972). They enquire as to whether, and if so how, contextual conditions open or close gaps in turnout between social groups, thereby increasing or decreasing the extent to which social inequality translates into political inequality (Gallego 2015). In studies of vote choice, contingent effects are not a sideline, but constitute the dominant phenomenon of interest. Most cross-national multilevel studies of electoral choice are not interested in explaining the outcome variable itself, but in how voters reach these decisions. This literature is primarily concerned with establishing whether and how the extent to which voters refer to certain values, attitudes, or beliefs depends on contextual conditions (Klingemann and Weßels 2009). In essence, these studies explore random slopes between known predictors of vote choice and the resulting decisions. They show how system characteristics strengthen or attenuate the relevance of factors whose universal importance for voting decisions, on the not very reliable basis of single-country research, is often too easily taken for granted.

This line of research implies a completely new way of thinking about electoral behavior, and has major potential for breaking entirely new ground. Ultimately, it is concerned with the question of whether voters in different systems construct their electoral choices in ways that are similar, regardless of context, or that are different. It can thus be seen as a potentially particularly consequential contribution toward overcoming the homogeneity assumption that was a hallmark of electoral research for decades. Whereas this conception rests on the (rarely explicated)

premise that one and the same model explains voters' choices under all circumstances, even personally and contextually widely divergent ones, the *heterogeneity perspective* claims that different voters may rely on different calculi to help them make up their minds. The discipline's homogeneity premise was first challenged by Rivers (1988), but this move has found only limited resonance so far. It mainly inspired research into the moderating role of political sophistication, which often found voters who had little political understanding relying on less complex decision calculi than voters who were more knowledgeable and interested (Blumenstiel 2016). Cross-national multilevel research adds a fundamentally new angle to this line of reasoning by investigating how institutional or other contextual circumstances affect the considerations on which voters rely when making up their minds. It has a profound potential to undermine the field's widespread reliance on the "one-equation-fits-all" assumption by demonstrating how differing contextual conditions give rise in systematic ways to variations in voters' calculi of choice.

3 Turnout

The most influential theories used to explain turnout are the "calculus of voting" (Downs 1957; Riker and Ordeshook 1968) and the "civic voluntarism" model (Verba et al. 1995). According to these views, voters can be induced to go to the polling stations by instrumental motives concerning the benefits expected from future governments' policies and by expressive motives that relate to the act of voting itself. The calculus of voting in particular also draws attention to the fact that voting is costly, for instance because obtaining and processing relevant information requires time, effort, and cognitive capacity. Correspondingly, factors such as partisanship, political interest and knowledge, internal and external efficacy, as well as voters' sense of civic duty and political support, but also distrust and cynicism, are important predictors of turnout at the individual level. These factors mediate the role of differences in voters' endowment with socioeconomic resources to some extent. Correspondingly, a person is more likely to vote if she or he is sufficiently motivated, but also sufficiently competent to make up her or his mind and reach a meaningful choice, as well as to cope with the practical requirements and challenges of actually casting her or his vote.

Citizens' active engagement and political equality are two core elements of democracy (Dahl 1998). From a normative point of view, it is therefore desirable that turnout be high and egalitarian. Against this background, contextual studies can be seen as an attempt to understand which institutional as well as social conditions are conducive or detrimental to achieving these goals. Even the most basic electoral statistics unequivocally show that aggregate levels of turnout differ hugely across countries. Macro studies indicate that this variation is closely related to institutional as well as socioeconomic differences between countries (Cancela and Geys 2016; Stockemer 2017). These contextual conditions may theoretically influence turnout in two ways. First, they may determine whether the act of voting can be executed rather conveniently and thus at little cost, or whether it is difficult and thus more

costly, with regard to both aspects of voting—deciding which party or candidate to support, and registering this choice in the formally required way (Gallego 2015; Wass and Blais 2017). Second, contextual characteristics may also have an impact on the motivational components of voters' turnout calculus, thus increasing or decreasing the intensity of their involvement in an election. Cross-nationally comparative multilevel research seeks to discern, by applying more appropriate methods than aggregate studies do, whether and in what ways this is the case. They clearly move beyond what aggregate studies could deliver by demonstrating how contextual conditions interact with voters' attributes in order to increase or decrease their likelihood to vote.

3.1 Effects of Institutional Contexts

Regulations that govern how elections are to be conducted are of obvious importance for voters' decisions to vote or abstain. To expect compulsory voting to boost turnout seems almost tautological. However, studies not only demonstrate that individual citizens' likelihood to vote is indeed higher in countries where they are legally required to do so, but also that the specific arrangement of compulsory voting makes a difference (Singh 2015). When voting is not only mandatory, but not voting is indeed subject to sanctions, turnout is further increased. Moreover, it is then also less dependent on factors such as income, political savvy, efficacy, or partisanship. In a similar vein, according to a study using EES data, compulsory voting mutes the impact of political interest on turnout (Söderlund et al. 2011). Research by Gallego (2010) and by Córdova and Rangel (2017) demonstrates that compulsory voting, especially when subject to sanctions, helps close the gender and education gaps in turnout by particularly strongly increasing the likelihood that women and less well-educated persons will vote. All in all, the findings suggest that when voting is enforced, thus rendering abstention costly, the voting population corresponds more closely to the electorate at large (Singh 2015).

A key attribute of electoral systems is their proportionality, that is the extent to which the mode by which they transform votes into parliamentary seats mirrors or distorts the vote shares obtained by the parties. Karp and Banducci (2008) confirm findings from macro studies that proportional representation (PR) systems are more conducive to turnout than plurality or majority systems are. Their findings suggest that this effect is mediated through voters' motivations: PR systems appear to exert a positive influence on partisanship and efficacy. Moreover, it is assumed, although not shown, that PR incentivizes parties to campaign, which in turn may mobilize voters. More recent studies offer more convincing tests of both attitudinal mediation and moderation. Singh (2011) for instance looks at how the disproportionality of electoral systems curbs the effect of benefits generated from left–right proximity on turnout. He finds that this association only emerges when voting takes place under PR. Fisher et al. (2008) observe that under PR, political knowledge is less relevant for turnout than under a plurality system. According to Blais et al. (2014), PR systems enhance turnout via the feeling of being represented by a party. Chen (2011) also claims that the effect of electoral systems works through voters' motivations, although in a complex way mediated through party systems. Since PR stimulates

centrifugal political competition among a larger number of parties, voters are more highly motivated to use their votes not only to improve the electoral prospects of parties to which they feel close, but also to fend off the threat posed by parties that are very distant and thus seen as disagreeable.

Other studies concentrate more directly on features of *party systems*, that is the supply of alternatives between which voters can choose. To some extent, party systems are the outcomes of electoral systems, but they nonetheless unfold their own effects. Proportional electoral systems presumably give rise to larger party systems (Duverger 1959). Remarkably, while PR systems appear to boost turnout, larger numbers of parties have been found to depress turnout, presumably because having a wider range of choices makes voting more difficult and thus costly (e. g., Gallego 2010). Having said that, when the polarization of ideological competition is also taken into account, the picture becomes more complex. Brockington (2009) shows that turnout is increased if the electorally viable parties are spread more broadly across the left–right spectrum. He argues that if left–right polarization is more intense, and the various parties' positions therefore more distinct, more voters can find an ideologically proximate party that corresponds closely to their interests. This in turn is presumed to strengthen their motivation to vote. Polarization also has contingent effects. It appears to primarily increase the turnout motivation of more sophisticated voters (Moral 2017). Furthermore, it has been found to increase turnout among those who believe that elections make a difference to policy outcomes, but to depress it among those that see relatively little meaning in elections (Kittilson and Anderson 2011).

Focusing on differences in education, Gallego (2010, 2015) is more broadly interested in how the institutionally induced difficulty of voting affects the consequences of status inequality for turnout. She reconfirms not only that the turnout gap shrinks when the effective number of parties is smaller, but also demonstrates that similar effects occur when simpler *types of ballots* are used (such as closed lists instead of procedures that allow preferential voting), when *registration* is automatic instead of self-initiated and voluntary, when *governments* are not fragmented, i. e., formed by single parties and not by coalitions, and when *media systems* entail a strong public service component and thus offer a richer supply of political information.

3.2 Effects of Socioeconomic Contexts

Some studies were also interested in the effects of social inequality at the societal level. Combining Gini coefficients obtained from the Luxemburg Income Study (LIS) with survey data from different sources (besides the CSES also the WVS and pooled national election studies), several studies unanimously demonstrate that *income inequality* depresses turnout. Their findings are, however, mixed with regard to whether this especially affects persons towards the lower end of the income scale (Jensen and Jespersen 2017; Solt 2008) or impacts on lower- and higher-income individuals to an equal degree (Beramendi and Anderson 2008; Gallego 2015). Some studies try to assess whether mobilization through traditional working-class organizations such as trade unions and leftist parties may balance out the participatory disadvantages faced by citizens on a lower income. Gallego (2015)

finds no indication that such targeted mobilization takes place, whereas Anderson and Beramendi (2012) suggest that the particular vulnerability of citizens on a lower income to aggregate income inequality might be attenuated by special mobilization efforts undertaken by parties of the left, but only when these parties have an incentive to engage in such activities. This, in turn, is only the case when they face competition from other left-leaning parties.

4 Vote Choice

As an object of study, vote choice is more challenging than turnout because researchers need to deal with the problem that their dependent variable concerns objects that are by definition nation specific. The candidates or parties that compete for citizens' votes are singular. They exist only in one country, but not in others. Comparative studies therefore need to move up to a more abstract level of analysis where parties' or candidates' "proper names" no longer occur (see also Spies and Franzmann 2019 in this special issue). One way to achieve this is to sort parties into generic categories that are meaningful across all the countries under study (Sartori 1970). One example is the distinction between incumbent and opposition parties, typically used by studies of retrospective performance voting (see Sect. 4.1). Sorting parties into those that formally participate in governments and those that do not poses no major problems of cross-national equivalence. Assigning parties to party families is another possible strategy, thus far mainly utilized by analyses of voting for right-wing extremist parties (e. g., Arzheimer 2009). Another strategy employed to achieve a more general perspective is stacking the data according to the logic of discrete-choice modeling, that is by moving from the level of respondents to the level of parties within respondents. This also eliminates the idiosyncrasies of individual parties from the analysis, but it adds considerable complexity because the number of analytical levels is increased (e. g., van der Brug et al. 2007).

Cross-national multilevel research into electoral choices has thus far focused primarily on the mechanism leading to these decisions. Most studies are not interested in explaining the outcome variable itself, but in how context conditions moderate the "calculus" through which voters reach their decisions. Exceptions are studies of partisanship (Singh and Thornton 2013; Lupu 2015), right-wing voting (e. g., Arzheimer 2009), or electoral volatility (Dejaeghere and Dassonneville 2017). The bulk of the research aims at disentangling whether and how the importance of certain predictors of vote choice varies depending on contextual conditions (Klingemann and Weßels 2009). From a methodological point of view, the main objects of these studies are random slopes in models of vote choice and cross-level interactions that might explain variations of effect sizes across contexts.[2] Of the considerable variety of factors contained in the "funnel of causality" that is commonly accepted as a heuristic overarching model for understanding vote choice (Campbell et al. 1960,

---

[2] Some studies transform this random slope problem into a random intercept problem by using as dependent variables not vote choice itself, but marker variables that indicate how it relates to the predictor of interest (see, e. g., Singh 2010).

pp. 24–32; Miller and Shanks 1996, pp. 190–193), only a small subset has been more thoroughly investigated so far. How contextual attributes condition the effects of other factors, such as social structure (Magalhães 2014) or candidates (Klingemann and Weßels 2009; Curtice and Lisi 2015), has been explored only sporadically. Others, such as most notably partisanship, have been completely neglected; whether and how the electoral role of this crucial predictor of voting varies according to system attributes has not yet been studied in a proper multilevel setup.

Two areas stand out as particularly productive in the study of contextual effects on how people vote: performance voting and left–right voting according to the spatial logic (see Spies and Franzmann 2019 in this issue for a discussion from the point of view of theories of party competition). As with research into turnout, this literature also has a normative connotation—it is concerned about whether and under which conditions elections can fulfill their function as instruments for holding governments to account for their actions. Electoral accountability means that "[e]lections make politicians pay attention to what the public wants and reflect public demands in their policies and performance" (Franklin et al. 2014, p. 399). This presupposes as a necessary, but not sufficient, condition that voters' choices should reflect their policy preferences and performance assessments (Ashworth 2012). Whether and to what degree this is the case depends on contextual circumstances, often in interaction with voters' personal attributes—that is the main outcome of this research. The same theme is raised in a more comprehensive way by research into "correct" voting, which also appears highly sensitive to context.

4.1 Performance Voting

An important approach to explaining electoral choices conceives of voters as sanctioning agencies (Healy and Malhotra 2013). It interprets vote choices as constituting referenda on the achievements of incumbent governments, assuming that when at the polls, voters look back at the incumbents' performance and decide according to whether they like or dislike what they see. Voting thus functions according to a basic reward–punishment mechanism. Voters hold governments liable for the conditions experienced under their leadership, rewarding them if they were good or punishing them (by "throwing the rascals out") if they were bad—a view that has important normative implications because it embodies the idea that come election time, voters retrospectively hold governments to account for their accomplishments in office.

4.1.1 Economic Voting

By far the most thoroughly investigated manifestation of performance-based electoral choice is economic voting. Traditionally being "the branch of voting studies most deeply engaged in cross-national research" (van der Brug et al. 2007, p. 16), it offered particularly fertile ground for the recent surge of interest in cross-national multilevel research. Since economic conditions are a macro phenomenon that varies only between countries or within countries across elections but is a constant at each election within a country, the study of economic voting gravitated quite naturally towards comparative approaches using aggregate data. These macro studies have

generated some, albeit seemingly somewhat shaky, evidence for the hypothesized basic reward–punishment mechanism. Micro-level studies based on election surveys also provided some support for this idea, even though it was also not very robust (Duch and Stevenson 2008, pp. 24–27) and came at the cost of a major compromise: to be amenable to micro-level research, the real economy had to be replaced by voters' economic perceptions. This raises problems of validity since it is unclear to which aspects of the economy respondents refer when they answer these questions, as well as problems of endogeneity because their answers might be biased by political preferences (Evans and Andersen 2006). All in all, this research indicates that there might be something to the hypothesized reward–punishment mechanism, although with considerable variation across countries. Against this background, multilevel studies combining cross-national survey data with macro indicators have the potential to break new ground in a variety of ways. In the same way as conventional survey studies, they make it possible to model vote choice at the micro level where it occurs. At the same time, they offer more valid insights by investigating the predictive power of objective economic conditions as well as their mediation through economic perceptions. In addition, they are ideally suited to model the conditioning impact of institutional settings on economic voting.

The relevance of institutional contexts for economic voting was first highlighted by Powell and Whitten (1993). Their macro analysis strongly suggests that the reward–punishment mechanism is not universal, but strongly contingent on countries' institutional makeup. The key concept here is "*clarity of responsibility*," and the basic intuition is that voters can mete out rewards or punishments only if it is sufficiently clear who is actually to blame for the conditions that they experience. Generally speaking, this is easier if responsibility for policy-making is concentrated instead of fragmented or divided. If institutional circumstances blur responsibility, it is hard for voters to identify the proper target for their satisfaction or dissatisfaction, so that they cannot decide who to reward or punish (Anderson 2007a). "Horizontal" responsibility is, for instance, less clear if a government is not made up of just one party, but of a coalition of several parties, if parties are not very cohesive and disciplined, or if legislative power is not concentrated in one powerful parliamentary chamber, but distributed more evenly in a bicameral system. Likewise, regional decentralization in systems of multilevel governance dissipates "vertical" responsibility. Obviously, integrating micro- and macro-level data in a multilevel design is the appropriate strategy to test such hypotheses.

In a comprehensive study based on pooled EES data collected in fifteen countries in the course of three European elections, van der Brug et al. (2007) evaluate the basic reward–punishment hypothesis as well as a range of supplementary hypotheses on conditioning factors. Combining three levels of analysis (voters, parties, countries), they investigate how changes in three indicators of macroeconomic performance (unemployment; inflation; gross domestic product, GDP) influence individual voters' party preferences and, mediated through these, how they impact on electoral choices. The study confirms that objective economic conditions do indeed affect governing parties' electoral prospects. Its test of the clarity hypothesis yields quite complex results. Using Powell and Whitten's (1993) index of horizontal clarity of responsibility, it is found "that economic voting as traditionally formulated

does not exist in low-clarity countries" (van der Brug et al. 2007, p. 173). Where governmental responsibility is clear, governments are held to account for how the economy developed under their leadership (although in quite complex patterns that are jointly conditioned by the governing parties' ideological profiles and sizes as well as dimensions of the economy). Under conditions of low clarity, by contrast, voters appear to think forward rather than backward. In a situation where it is hard to attribute past experiences to a specific party, they seem not to care so much about who governed than about what can be expected from the competing parties' future policies, with a specific focus on the large parties, given that they are presumably more influential.

An even more ambitious study by Duch and Stevenson (2008) starts from the premise that voters are instrumentally rational and offers a more sophisticated explanation of economic voting than the conventional, purely backward-oriented reward–punishment model. According to the proposed "competence theory" of economic voting, voters use retrospective information about macroeconomic developments not in order to sanction governing parties for their past performance, but to assess their competence to perform well in the future, out of a desire to choose the most capable economic manager for the next government. Micro-level analyses of data from 163 surveys conducted in 18 Western democracies again confirm that the magnitude of the economic vote differs widely between countries, as well as within them. The authors' theory predicts this variation to be a result of features of economic and political contexts that affect the amount of control that elected decision-makers can exert over the economy, relative to the amount of control exerted by agencies outside the realm of a country's electoral politics. Voters are assumed to be aware that in a globalized world, economic developments are not fully under the control of national governments, and to therefore discount that part of economic development which is not attributable to their governments' actions. When choosing a party, voters are thus expected to respond to the economy only to the extent to which its development can be attributed to the government, and not to developments beyond its control. In line with these expectations, it is demonstrated that the intensity of economic voting is decreased in open economies whose GDP heavily depends on international trade. To counter adverse effects of international trade dependency, open economies often adopt institutional and policy strategies that lead to an expansion of the state sector. These policies lead to a relative increase in the number of non-elected agencies, and thus can also be expected to decrease the strength of economic voting. This is shown using various indicators of an expansive state sector (public expenditure quota, indices of corporatism, and the extent of government economic regulation). Economic *globalization* thus diminishes the role of the economy for governments' electoral accountability not only directly, but also indirectly, mediated through governments' domestic responses.

Looking at effects of institutional features of context, the study offers further support for the idea that the manner in which responsibility for policy-making at the time of an election is *distributed across parties* matters for economic voting. However, in line with the proposed competence theory, the theoretical explanation offered for this phenomenon again differs from the conventional understanding. According to the authors, it is not the difficulty to attribute responsibility under

constellations of dispersed authority that depresses the amount of economic voting, but an adjustment of the degree to which voters base their party choices on economic considerations to the relative power exerted by the chosen parties, which may be small or large, depending on institutional conditions and the parties' competitive standing. Accordingly, the strength of economic voting is responsive in a number of ways to the amount of concentration or diffusion of governmental power within a system. Confirming findings from other studies, the importance of the economy for vote choices overall is found to increase if governmental responsibility is more highly concentrated. More specifically, in presidential systems, economic voting is stronger when the president's party controls parliament than it is under a divided government. In parliamentary systems, it makes a difference whether a government consists of a single party or of several coalition parties, and whether or not it controls the majority of seats. In addition, economic voting for a party increases with its share of cabinet seats, and it is comparatively stronger for the party of the head of government, as well as for the party of the minister of finance. Moreover, the economy's influence on the vote is greatest when the leading party of the incumbent government is a competitive—but not certain—aspirant to the leadership of the next government. Both certain re-election and certain loss of office depress economic voting. This suggests that economic voters indeed, as hypothesized, not only look backward, but also forward, at least when the election appears to be contested.

Subsequent research has added important details to these findings. Using the same dataset and applying an advanced multilevel technique of mediation analysis, Becher and Donnelly (2013) demonstrate that the influence of the real economy on the vote is indeed mediated in its entirety (GDP change), or at least to a substantial extent (unemployment change), by economic perceptions, something that has appeared doubtful in the light of previous criticism of these measures as being ambiguous and potentially endogenous to the vote. Kayser and Peress (2012) provide direct evidence of the "*benchmarking*" mechanism of economic assessments. For GDP change, they show that economic voting responds to the positive or negative deviation of a country's economic development from the global trajectory of the economy. Additional analyses by the same authors suggest that the news media are indeed the source of the quite sophisticated information needed for such a calculus of choice. Also claiming that voters infer a government's future capacity to deliver desirable economic outcomes from retrospective evaluations, Marinova (2016, pp. 83–101) shows that *party system instability* is another important moderator of economic voting. When the parties that compete at subsequent elections undergo abrupt organizational changes, fusing, splitting, emerging, or disappearing between one election and the next, voters cannot use past experience as a heuristic to predict the future. As a consequence, the reward–punishment mechanism is undermined.

4.1.2 Other Forms of Performance Voting

While the bulk of comparative multilevel research on performance-based voting is concerned with the economy, some studies also apply broader or completely different perspectives. Using EES data, Hellwig (2015, pp. 76–95) finds that globalization diminishes the intensity of economic voting, but in a compensatory way, at

the same time strengthens the electoral impact of other dimensions of governments' performance, such as healthcare and immigration. Giger (2011) is more specifically interested in *social policy*. Exploring the electoral consequences of welfare retrenchment, she finds welfare cutbacks leading to more negative ratings of governmental performance on social policy, and these ratings in turn decrease the likelihood of voting for an incumbent party. Since welfare policies are constants in elections, this association can only be detected by multilevel modeling. However, while this kind of performance voting appears not to be similarly pronounced in all countries, its cross-country variation is not a function of the various governments' clarity of responsibility.

Other studies of the role of the clarity of responsibility for performance voting resort to a more general perspective. Rather than focusing on voters' reactions to government activities in defined areas of public policy, they are interested in *general performance voting*. Fisher and Hobolt (2010), for instance, show that global performance judgments on incumbent governments affect votes for government parties less pronouncedly for coalition governments than they do for single-party governments. Among coalition governments, furthermore, the strength of performance voting declines as the number of member parties increases. Hellwig (2011) studies the same association in conjunction with a broader set of contextual features, and obtains partly similar results. He finds performance voting to be strengthened if administrative responsibility is concentrated on a smaller number of parties. Moreover, performance judgments also affect electoral choices more strongly in highly polarized party systems where the electoral alternatives supplied by the party system can be more easily distinguished in terms of their policy positions.

4.2 Left–Right Voting

The proximity model of vote choice originates from Downs' (1957) seminal theory of policy-oriented instrumental voting. According to this view, voters choose prospectively and use their votes rationally, in order to maximize their benefit from the work done by the next government. In other words, "voters use parties as instruments to attain their preferred policy" (Kedar 2009, p. 94). By voting for a particular party or candidate, they seek to contribute to installing a government that in turn can be expected to deliver policies that correspond as closely as possible to their policy preferences. To study this mechanism, comparative research commonly refers to the left–right dimension, conceived as a "super issue" that encapsulates a broad range of more specific issues and policy concerns (Mair 2007, p. 212). This makes it possible to circumnavigate the potentially intractable problem of how to establish the equivalence of specific policy issues across countries (as well as across elections within the same country). That the left–right dimension is meaningful in most of the world's democracies and that the left–right scale can therefore be used as a measurement tool in cross-national research has been demonstrated repeatedly. Most studies conceive of left–right voting as choices based on simple proximity, the expectation being that voters should pick the party whose left–right position is closest to their own ideal point. A range of methods are available to establish parties' left–right positions, including perceptions as registered by surveys of voters, political elites,

or experts, and content analyses of party or media sources (see Spies and Franzmann 2019 in this issue for a detailed discussion).

Many comparative studies of left–right voting start from the normative premise that proximity voting on left–right terms is a desirable mode of choosing at elections, since it has a clear relationship to political accountability, encapsulated in the normative model of responsible party government (see, e.g., Weßels and Schmitt 2008, 2014). According to this notion, a functioning system of political representation requires parties to offer meaningful choices in terms of distinct policy profiles. They are elected on the basis of these offers' fit to their voters' policy preferences, and then expected to translate these programs into actual policies (Franklin et al. 2014). Comparative research is interested in how institutional contexts affect voters' ability to turn their left–right positions into choices that best suit their policy preferences. This concerns not only the process of comparing voters' own ideal points with the positions of the parties, although this is a *sine qua non* for left–right-based voting. Under certain institutionally defined circumstances, it may indeed appear more rational in terms of expected policy outcomes to choose not the closest party, but a more distant one. Examples are the notions of "strategic" and "compensatory" voting.

4.2.1 Information Effects on Proximity Voting

Since the mechanism of electoral choice implied in a spatial model of left–right voting requires a substantial amount of information and processing skills on the part of voters, it is considered to be cognitively rather demanding. Many studies are therefore concerned with the question of whether and how it is made easier or harder by contextual conditions. Particular interest attaches to features of *party systems*, most notably their *complexity and cognitive manageability* (Kroh 2009; Singh 2010), drawing on the general assumption that a highly structured supply with distinctly profiled electoral alternatives makes proximity voting more feasible.

Weßels and Schmitt (2008), for instance, hypothesize that voters can only rely on distances between their own policy preferences and parties' offers when the party system is sufficiently differentiated to offer real choices in terms of a broad range of distinct offers. Confirming this reasoning, their study shows that left–right proximity is more closely associated with vote choices in systems where the effective number of parliamentary parties is larger (see also Singh 2010; contradictory findings are presented by Kroh 2009), whose party systems cover a wider range of the left–right spectrum, and where parties are sufficiently distant from one another to indicate clear policy profiles. A follow-up study that measures the clarity of parties' policy profiles more directly by means of the association between the content of parties' election platforms and parties' perceived left–right positions confirms that left–right proximity voting is promoted by the clarity of parties' policy positions (Weßels and Schmitt 2014). Kroh (2009) arrives at the same conclusion using a measure building on the clarity of parties' left–right positions as perceived by voters.

A similar story is told by analyses which explore the impact of party system polarization on the strength of left–right voting (Dalton 2011; Kroh 2009; Lachat 2008). Using EES data, Lachat (2008) furthermore shows that voters with lower and

higher motivation and cognitive skills do not profit from this variation in context to the same extent. More polarized party systems increase the likelihood of resorting to this type of decision-making especially among more sophisticated voters. Kroh (2009) likewise finds party polarization to enhance left–right voting. Another way the party system can become cognitively more manageable for voters is the ordering of parties in standing clusters as it happens in systems with frequent coalition governments; accordingly, the effective number of parties in government has a positive effect on left–right voting (Kroh 2009). Singh (2010) shows that the dimensionality underlying voters' evaluations of the competing parties is another layer of party-system complexity that affects the likelihood of left–right voting. The more dimensions structure party competition, the more proximity voting is impeded.

Gingrich (2014) offers yet another take on the translation of left–right positions into party choices. She interprets the left–right dimension rather specifically as an indicator of redistributional and social policy preferences, and shows that voters' propensity to permit these preferences to guide their electoral choices is a function of the *informational structure of the welfare state*. According to her reasoning, direct transfers are more informative about what the welfare state does because voters can experience them immediately and easily attribute them to their source, whereas indirect benefits such as tax breaks are much more opaque. Consequently, the study shows that the effect of left–right proximity on vote choice is stronger the more pronounced direct transfers are within a country's welfare mix, presumably because it is easier for voters to connect their policy preferences to the parties that compete for their votes when the welfare state is more visible.

4.2.2 "Strategic" and "Compensatory" Voting

The relevance of proximity voting is also a function of institutional arrangements that affect *how votes are translated into seats, and seats into governmental power*. As a consequence, simple proximity voting, i.e., "sincerely" voting for the closest party (Gschwend 2009, p. 290), is not always the instrumentally most rational choice for voters that seek to maximize electoral utility. Depending on circumstances, it may indeed be more rational to choose a more distant party. Voters acting this way may be motivated by a desire to weaken the least-liked party's electoral prospects or more simply by a desire to obtain representation in parliament.

The notion of "*strategic*" voting is directly concerned with effects of electoral systems. It rests on the assumption that voters do not want to "waste their votes" by choosing parties or candidates that appear unviable at the election, and therefore shift to less preferred alternatives if these have better chances of electoral success. When choosing strategically, a voter takes two ingredients into account: a personal ranking of parties according to their preferableness on policy terms, and expectations about the outcome of the election. Blais and Gschwend (2011) analyze strategic defection from voters' most preferred parties at parliamentary elections. Context is found to matter in one very specific, though highly plausible way. The degree to which electoral systems distort the translation of the distribution of votes into parliamentary seat shares in favor of large parties interacts with the size of parties that would be the objects of sincere votes if voters had not chosen strategically: The stronger the

disproportionality induced by the electoral system, the greater the likelihood that voters desert smaller parties. Wagner (2013) rephrases Duverger's (1959) famous claim that voters tend to small parties under majoritarian electoral systems into a testable macro-micro theory. He shows that ideological voting is moderated by strategic considerations based on the relationship between a preferred party's size and the effective threshold for gaining entry into parliament that is jointly defined by two components of electoral systems: district magnitudes and legal thresholds. Votes for smaller parties are typically sincerely ideological, whereas votes for larger parties contain a mix of sincere and strategic motives, the latter resulting from voters that defect from small parties to avoid the danger of wasting their votes. This pattern is found to be most pronounced in unitary political systems, whereas the institutional complexity induced by the multilevel politics of federal systems appears to weaken the impact of strategic considerations.

Whereas strategic voting is merely seat-maximizing (Wagner 2013, p. 93), *compensatory* voting is policy-maximizing, and takes the entire distance between votes and policy outcomes into account. According to Kedar (2009, pp. 65–101), voters are conscious that institutional arrangements mediate between parties' policy programs and the actual policies to be expected from them when they get the chance to govern. While majoritarian systems concentrate power, and therefore allow for a direct translation of voters' choices into governmental power, the connection is diluted in systems whose institutions disperse the power to determine policy outcomes among many actors, so that no single actor can be expected to push through its policy program in an unfettered way. Policy outcomes will instead result from bargaining processes, and therefore take the form of compromises between several parties. This means that eventual policies will typically deviate from the parties' promises. According to Kedar's theory, voters may respond to this weakening of the link between party platforms and actual policies by means of compensatory voting. They choose parties whose positions deviate from those of the most proximate parties and are typically more extreme than their own positions. The rationale behind such a choice is that these more extreme parties may serve as effective counterweights to prevent the ultimate policy outcome from being deflected too far from the respective voter's ideal point. The study relies on four indicators to measure the extent to which power in parliaments is shared or concentrated. It demonstrates that the degree to which voters choose according to the compensatory logic instead of simple proximity is greater the more a system is characterized by coalition instead of single-party governments, the higher the effective number of parties in parliament, the larger the magnitude of its electoral districts, and the less the power to set the agenda of parliamentary debates is vested in the government.

4.3 "Correct" Voting

In a more encompassing way, the concept of "correct" voting echoes the normative concerns implied in the literature on performance and left–right voting reviewed above. Choosing "correctly" means choosing consistently with one's own interests, broadly conceived. A correct decision is the one that a voter would take if she or he had comprehensive information about the election and the competing parties

(Lau and Redlawsk 1997). Lau et al. (2014) investigate how system characteristics affect whether or not voters cast correct votes. To measure whether voters make, in this sense, the right choices, they predict which of the competing parties they should elect, given their evaluations of the parties' economic and general performance, closeness on the left–right dimension, as well as partisanship, and compare the result of this prediction with the same voters' actual choices. Votes are considered to be correct if the chosen and predicted parties are identical. Otherwise, they are classified as incorrect, indicating that the chosen party does not conform to the respective voters' own attitudes. Partly paralleling research on performance and left–right voting, the study shows that correct choices are more likely in systems whose institutional conditions simplify the choice situation and decrease information costs. Specifically, the likelihood that citizens will vote correctly is higher if their choices relate to parties and not to candidates, so that less information is necessary to make a decision, if the supply of electoral alternatives (as indicated by the effective number of parties) is smaller and the alternatives themselves are ideologically more distinct (as measured by party-system polarization), and if governmental responsibility is clear (as indicated by single-party governments in parliamentary systems and unified governments in presidential systems). Moreover, if the media environment offers abundant information, correct voting also becomes more likely, especially for more efficacious voters, as indicated by a cross-level interaction. At least in part, these findings parallel results of research on turnout, suggesting that some system attributes that increase voters' likelihood to go to the polls also improve their prospects to pick the parties that best represent their preferences and interests.

5 Conclusion

As a form of political action, casting votes is distinct. It is not only the most important, but also the most highly institutionalized form of political behavior. Inherently, it is therefore a multilevel phenomenon that cannot be fully understood at the level of individual voters alone. A multilevel perspective that takes the micro level of voters and the macro level of the political systems to which they belong into account simultaneously is the appropriate way to fully comprehend citizens' decisions about whether to vote at elections, and if so, for who. Recent years have seen a rapid increase in the number of studies that take this premise seriously. They investigate turnout and vote choice as activities of "embedded" citizens that respond in their behaviors to institutional but also socioeconomic particularities of the contexts in which they reside. Surveying this emerging field after the first decade reveals a rich landscape of research. It is highly differentiated, but when viewed more closely, some thematic clusters emerge where research activity has been particularly intense. They relate to two questions, both of which have a normative tinge: (i) Which conditions promote turnout that is high and more egalitarian, thus giving citizens an equal say in politics? And (ii) Which conditions promote electoral choices that are in line with voters' own interests, thus strengthening the role of elections as instruments for holding governments to account?

Without too grossly oversimplifying a voluminous body of research, one can, for instance, record that aggregate social inequality has been established to be detrimental to individual voters' turnout. An effective institutional precaution to boost turnout is, on the other hand, compulsory voting, especially if it is subject to sanctions. That most people vote when they are forced to seems hardly surprising. But the implications of this effect are not trivial. When voting is mandatory, the voting population corresponds more closely to the electorate at large, whereas with voluntary participation, voters' motivations become critically important. By strengthening relevant motivations, proportional electoral systems seem to facilitate higher and more equal turnout. The extent to which institutions render the act of voting easier or more difficult is another major theme of the extant body of research. First off, this concerns seemingly mundane features of electoral logistics, such as registration laws or ballot formats that determine how much effort is required of citizens if they wish to cast their votes. But it also concerns more fundamental attributes of party systems, electoral systems, and systems of government. By simplifying or complicating the choice situation, they decrease or increase the cognitive challenge voters need to meet in order to arrive at meaningful choices. Extant research suggests that the supply structure offered by party systems is particularly important. Choosing appears easier when party systems are organizationally stable and offer a wide range of internally cohesive alternatives with distinct ideological profiles within a conflict structure of low dimensionality. The amount and quality of the information provided by media systems also makes a difference. Under these conditions, not only turnout, but also choices in line with voters' interests are facilitated.

Research on vote choice has concentrated especially on two forms of policy-based decision-making—performance voting and left–right voting. This research has found clear-cut alternatives in the competition for governmental power also to decrease the cognitive burden of choosing. Single-party governments are ideal in this regard, but in multi-party systems, established coalition patterns may serve as functional equivalents. Nonetheless, coalition governments also have their downsides, especially if they include large numbers of parties. In the same way as other manifestations of systemic power fragmentation, such as divided government in presidential systems, bicameralism, federalism, and powerful non-elected agencies such as supreme courts or central banks, they blur responsibility for policy outcomes. Such conditions render performance-based vote choices difficult, if not impossible, because it is hard to recognize who to reward or blame. If voters cannot easily locate the right target, they lack the necessary guidance for translating their experiences into electoral choices. Likewise, left–right voting is more challenging in systems with fragmented power arrangements because voters cannot simply cast straightforward votes, but need to engage in complicated calculations of strategic and compensatory voting. Moreover, party systems that are more clearly structured in terms of characteristics such as the number and distinctiveness of competing parties and the polarization between them, improve voters' ability to transform their left–right preferences into choices for the most proximate party (see also Spies and Franzmann 2019 in this issue).

Already during its first decade, comparative multilevel research into electoral behavior has thus collected ample evidence that the institutional makeup of political systems has far-reaching consequences for normative core requirements of repre-

sentative democracy. Whether citizens turn out or not, and how they choose when they go to the polls, is very much influenced by the institutional architecture of political systems. If responsibility for policy outcomes is blurred, as is systematically the case in systems were power is not concentrated but diffused, the accountability mechanism of retrospective voting is undermined. If party systems display little structuration, so that voters have a hard time to connect their ideological preferences to electoral alternatives, the accountability mechanism of responsible party government runs dry.

These findings mark clear achievements of the new approach to studying electoral behavior. They indicate a strong potential to inform initiatives for electoral engineering (Norris 2004) and institutional design more generally (Goodin 1996). However, the nature of these achievements varies. Not all findings are substantively new; some primarily contribute towards consolidating the field on the basis of better data and more appropriate methods. All in all, findings of multilevel approaches may appear less surprising with regard to turnout. It can be argued that the contribution of comparative multilevel research in this field thus far consists primarily of validating existing knowledge of institutional preconditions of higher and lower turnout by placing it on firmer ground through modeling techniques that avoid the pitfalls of the ecological and individualistic fallacies. More obviously new are its findings about the effects of contextual features (such as voter registration requirements, government fragmentation, or public service broadcasting; see Gallego 2015) on the equality of political participation. Only comparative research that relies on multilevel approaches can unequivocally demonstrate how attributes of contexts help open or close turnout gaps between social groups.

A similar validating function for extant research into vote choice can be attributed to studies of how the clarity of responsibility affects performance voting. In terms of their potential to break substantively new ground, two developments stand out as particularly innovative in research on vote choice. One emerges from the ability to connect individual voting behavior to factors that, at single elections, are constants at the macro level, such as the state of the economy or governments' policies. Comparative multilevel research not only allows transcendence of the reliance on perceptual proxies for such phenomena when analyzing voting behavior at the micro level of individuals, but also permits exploration of how subjective impressions of this kind mediate between objective circumstances and electoral choices. A prime example of what can be learned from such an approach is Becher and Donnelly's (2013) study of the joint direct and indirect effects of macroeconomic conditions and perceptions of the economy on vote choices.

In a much more general vein, these studies also have a far-reaching potential to force the entire field of electoral studies into reconsidering the hitherto dominant outlook on its object. As findings accumulate that show how institutions moderate the calculi by which voters arrive at their choices, the homogeneity assumption (and the related "essentialist" bias in thinking about vote choice) that for decades has been a hallmark of electoral research appears increasingly questionable. Findings of multilevel studies that show how certain predictors of vote choice, routinely treated as universally relevant by extant single-country research, are important only in some contexts, but irrelevant in others, are grist to the mill of the fundamental

challenge to psephological thinking that has been mounted by proponents of the heterogeneity perspective on vote choice. First proposed in a seminal article by Rivers (1988), this view questions the premise of traditional electoral studies that all voters choose according to the same formula. It posits instead that different groups of voters rely on divergent modes of decision-making. Empirical support for this view comes thus far mainly from studies that show how voters' political knowledge moderates the effects of standard predictors, such as issue proximity, on electoral choices (Blumenstiel 2016). The comparative research reviewed above suggests that not only personal attributes but also institutional circumstances are a source of heterogeneity in decision-making. It demonstrates that certain contextual settings are conducive to some styles of electoral decision-making and detrimental to others—a finding with potentially massive implications for the ability of elections to function as instruments of democratic accountability. Not only whether, but also how you choose at elections depends crucially on where you live—that is the most powerful message of cross-national multilevel research on voting behavior.

The emerging field of cross-nationally comparative research on electoral behavior using multilevel approaches has thus already demonstrated its potential in several ways to improve scholarly understanding of the democratic role of elections. Nonetheless, there is room for further development. Important potentials of cross-national multilevel research have thus far been insufficiently exploited. To begin with, many aspects of voting behavior that for decades were important topics of single-country studies have thus far attracted only little or no attention among comparativists. Candidate and issue voting are obvious examples. Moreover, it seems surprising that research on vote choice has, as yet, not shown a more pronounced interest in the question of why certain parties are supported or deserted at elections. The concept of party families (Mair and Mudde 1998) offers pathways to deal with the problem that the objects between which voters choose exist only in one country, but applications of this strategy have thus far concentrated on voter support for right-wing extremist parties (e. g., Arzheimer 2009). Similar studies of other party families would be of obvious value. A closer reading of the literature also gives the impression that, all in all, rather little effort has so far been invested to come closer to the theoretical mechanisms behind electoral behavior. Some standard measures of political systems' institutional attributes have become quite popular in multilevel studies, but the theoretical rationales proposed to connect them to individual behavior vary substantially. Research thus appears to some extent to be data driven. More energy should therefore be invested to design inquiries in such a way that they allow deeper empirical insights into the theoretical mechanisms that connect the macro level of political systems and the micro level of voters' decision-making. An important task ahead is to move from merely assuming mediating mechanisms to also demonstrating them, thus ultimately allowing comparative assessments of competing theories on the system dependence of individual electoral behavior.

References

Almond, Gabriel A., and Sidney Verba. 1963. *The civic culture. Political attitudes and democracy in five nations.* Princeton/NJ: Princeton University Press.

Anderson, Christopher J. 2007a. The end of economic voting? Contingency dilemmas and the limits of democratic. *Annual Review of Political Science* 10:271–296.

Anderson, Christopher J. 2007b. The interaction of structures and voter behavior. In *The Oxford handbook of political behavior,* eds. Russell J. Dalton and Hans-Dieter Klingemann, 588–609. Oxford: Oxford University Press.

Anderson, Christopher J. 2009. Nested citizens: Macropolitics and microbehavior in comparative politics. In *Comparative politics. Rationality, culture, and structure,* eds. Mark Irving Lichbach and Alan S. Zuckerman, 314–332. Cambridge: Cambridge University Press.

Anderson, Christopher J., and Pablo Beramendi. 2012. Left parties, poor voters, and electoral participation in advanced industrial societies. *Comparative Political Studies* 45:714–746.

Anderson, Christopher J., and Russell J. Dalton. 2011. Nested voters: Citizen choices embedded in political contexts. In *Citizens, context, and choice: How context shapes citizens' electoral choices,* eds. Russell J. Dalton and Christopher J. Anderson, 241–256. Oxford: Oxford University Press

Arzheimer, Kai. 2009. Contextual factors and the extreme right vote in western Europe, 1980–2002. *American Journal of Political Science* 53:259–275.

Ashworth, Scott. 2012. Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science* 15:183–201.

Barnes, Samuel H., Max Kaase, Klaus R. Allerbeck, Barbara G. Farah, Felix J. Heunks, Ronald Inglehart, Kent M. Jennings, Hans D. Klingemann, Alan Marsh and Leopold Rosenmayr. 1979. *Political action. Mass participation in five western democracies.* Beverly Hills: Sage Publications.

Becher, Michael, and Michael Donnelly. 2013. Economic performance, individual evaluations, and the vote: Investigating the causal mechanism. *The Journal of Politics* 75:968–979.

Beramendi, Pablo, and Christopher J. Anderson. 2008. Income, inequality, and electoral participation. In *Democracy, inequality, and representation in comparative perspective*, eds., 278–311. New York: Russell Sage Foundation.

Blais, André, and Thomas Gschwend. 2011. Strategic defection across elections, parties, and voters. In *Citizens, context, and choice: How context shapes citizens' electoral choices,* eds. Russell J. Dalton and Christopher J. Anderson, 176–193. Oxford: Oxford University Press.

Blais, André, Shane P Singh and Delia Dumitrescu. 2014. Political institutions, perceptions of representation, and the turnout decision. In *Elections and democracy: Representation and accountability,* ed. Jacques Thomassen, 99–112. Oxford: Oxford University Press.

Blumenstiel, Jan Eric. 2016. *Wie sich Wähler beim Entscheiden unterscheiden. Wählerheterogeneität bei den Bundestagswahlen 1998 bis 2009* (How voters differ when making up their minds. Voter heterogeneity in the Bundestag elections 1998 to 2009). Baden-Baden: Nomos.

Brockington, David. 2009. It's about the benefits. Choice environments, ideological proximity and individual participation in 28 democracies. *Party Politics* 15:435–454.

Van der Brug, Wouter, Cees Van der Eijk and Marc N. Franklin. 2007. *The economy and the vote. Economic conditions and elections in fifteen countries.* Cambridge: Cambridge University Press.

Campbell, Angus, Philip E. Converse, Warren E. Miller and Donald E. Stokes. 1960. *The American voter.* New York: Wiley.

Cancela, João, and Benny Geys. 2016. Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral Studies* 42:264–275.

Chen, Tse-hsin. 2011. Uncovering the micro-foundations of turnout and electoral systems. *Electoral Studies* 30:295–308.

Coleman, James S. 1990. *Foundations of social theory.* Cambridge/Mass.: Harvard UP.

Curtice, John, and Marco Lisi. 2015. The impact of leaders in parliamentary and presidential regimes. In *Personality politics? The role of leader evaluations in democratic elections,* eds. Marina Costa Lobo and John Curtice, 63–86. Oxford: Oxford University Press.

Córdova, Abby, and Gabriela Rangel. 2017. Addressing the gender gap. *Comparative Political Studies* 50:264–290.

Dahl, Robert. 1998. *On democracy*. New Haven: Yale University Press.

Dalton, Russell J. 2008. The quantity and the quality of party systems. *Comparative Political Studies* 41:899–920.

Dalton, Russell J. 2011. Left-right orientations, context, and voting choices. In *Citizens, context, and choice: How context shapes citizens' electoral choices,* eds. Russell J. Dalton and Christopher J. Anderson, 103–125. Oxford: Oxford University Press.

Dalton, Russell J. 2014. *Citizen politics. Public opinion and political parties in advanced industrial democracies*. Los Angeles: Sage.

Dalton, Russell J., and Christopher J. Anderson. 2011b. Citizens, context, and choice. In *Citizens, context, and choice: How context shapes citizens' electoral choices,* eds. Russell J. Dalton and Christopher J. Anderson, 3–30. Oxford: Oxford University Press.

Dejaeghere, Yves, and Ruth Dassonneville. 2017. A comparative investigation into the effects of party-system variables on party switching using individual-level data. *Party Politics* 23:110–123.

Downs, Anthony 1957. *An economic theory of democracy.* New York: Harper & Brothers.

Duch, Raymond M., and Randolph T. Stevenson. 2008. *The economic vote. How political and economic institutions condition election results.* Cambridge: Cambridge University Press.

Duverger, Maurice. 1959. *Political parties: Their organization and activity in the modern state.* London: Methuen.

Esser, Hartmut. 1999. *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln* (Sociology. Special foundations. Vol. I: Situational logic and actions). Frankfurt a. M.: Campus.

Eulau, Heinz. 1986. *Politics, self, and society: A theme and variations.* Cambridge/London: Harvard University Press.

Evans, Geoffrey, and Robert Andersen. 2006. The political conditioning of economic perceptions. *Journal of Politics* 68:194–207.

Fisher, Stephen D., and Sara B. Hobolt. 2010. Coalition government and electoral accountability. *Electoral Studies* 29:358–369.

Fisher, Stephen D., Laurence Lessard-Phillips, Sara B. Hobolt, and John Curtice. 2008. Disengaging voters: Do plurality systems discourage the less knowledgeable from voting? *Electoral Studies* 27:89–104.

Franklin, Mark, Stuart Soroka, and Christopher Wlezien. 2014. Elections. In *Oxford handbook of public accountability*. eds. Mark Bovens, Robert Goodin and Thomas Schillemans, 389–404. Oxford: Oxford University Press.

Franzese, Robert J. 2005. Empirical strategies for various manifestations of multilevel data. *Political Analysis* 13:430–446.

Friedrichs, Jürgen, and Alexandra Nonnenmacher. 2014. Die Analyse sozialer Kontexte (The analysis of social contexts). In *Soziale Kontexte und soziale Mechanismen. Kölner Zeitschrift für Soziologie und Sozialpsychologie* (Social contexts and social mechanisms. Cologne Journal for Sociology and Social Psychology) special issue 66:1–16.

Gallagher, Michael. 1991. Proportionality, disproportionality and electoral systems. *Electoral Studies* 10:33–51.

Gallego, Aina. 2010. Understanding unequal turnout: Education and voting in comparative perspective. *Electoral Studies* 29:239–248.

Gallego, Aina. 2015. *Unequal political participation worldwide.* Cambridge: University Press.

Giger, Natalie. 2011. *The risk of social policy?* New York: Routledge.

Gingrich, Jane. 2014. Visibility, values, and voters: The informational role of the welfare state. *Journal of Politics* 76:565–580.

Goodin, Robert. Ed. 1996. *The theory of institutional design*. Cambridge: Cambridge University Press.

Goodman, Leo A. 1953. Ecological regressions and behavior of individuals. *American Sociological Review* 18:663–664.

Gschwend, Thomas. 2009. District magnitude and the comparative study of strategic voting. In *The comparative study of electoral systems,* ed. Hans-Dieter Klingemann, 289–310. Oxford: Oxford University Press.

Healy, Andrew, and Neil Malhotra. 2013. Retrospective voting reconsidered. *Annual Review of Political Science* 16:285–306.

Heath, Anthony, Stephen Fisher and Shawna Smith. 2005. The globalization of public opinion research. *Annual Review of Political Science* 8:297–333.

Hellwig, Timothy. 2011. Context, political information, and performance voting. In *Citizens, context, and choice: How context shapes citizens' electoral choices,* eds. Russell J. Dalton and Christopher J. Anderson, 149–175. Oxford: Oxford University Press.

Hellwig, Timothy. 2015. *Globalization and mass politics. Retaining the room to maneuver*. Cambridge: Cambridge University Press.

Huckfeldt, Robert, and John Sprague. 1995. *Citizens, politics, and social communication. Information and influence in an election campaign.* Cambridge: Cambridge University Press.

Jensen, Carsten, and Bjarke Bøgeskov Jespersen. 2017. To have or not to have: Effects of economic inequality on turnout in European democracies. *Electoral Studies* 45:24–28.

Jusko, Karen Long, and W. Phillips Shively. 2005. Applying a two-step strategy to the analysis of cross-national public opinion data. *Political Analysis* 13:327–344.

Kaase, Max. 2015. Data bases and statistical systems: Political science (general). In *International encyclopedia of the social & behavioral sciences,* ed. James R. Wright, 830–835. Amsterdam: Elsevier.

Karp, Jeffrey A., and Susan A. Banducci. 2008. Political efficacy and participation in twenty-seven democracies: How electoral systems shape political behaviour. *British Journal of Political Science* 38:311–334.

Kayser, Mark Andreas, and Michael Peress. 2012. Benchmarking across borders: Electoral accountability and the necessity of comparison. *American Political Science Review* 106:661–684.

Kedar, Orit. 2009. *Voting for policy, not parties. How voters compensate for power sharing.* Cambridge: Cambridge University Press.

Kittilson, Miki Caul, and Christopher J. Anderson. 2011. Electoral supply and voter turnout. In *Citizens, context, and choice: How context shapes citizens' electoral choices,* eds. Russell J. Dalton and Christopher J. Anderson, 33–54. Oxford: Oxford University Press

Klingemann, Hans-Dieter. Ed. 2009. *The comparative study of electoral systems.* Oxford: Oxford University Press.

Klingemann, Hans-Dieter, and Bernhard Weßels. 2009. How voters cope with the complexity of their political environment: Differentiation of political supply, effectiveness of electoral institutions, and the calculus of voting. In *The comparative study of electoral systems,* ed. Hans-Dieter Klingemann, 237–265. Oxford: Oxford University Press.

Kroh, Martin. 2009. The ease of ideological voting: Voter sophistication and party system complexity. In *The comparative study of electoral systems,* ed. Hans-Dieter Klingemann, 220–236. Oxford: Oxford University Press.

Laakso, Markku, and Rein Taagepera. 1979. Effective number of parties: A measure with application to west Europe. *Comparative Political Studies* 12:3–27.

Lachat, Romain. 2008. The impact of party polarization on ideological voting. *Electoral Studies* 27:687–698.

Lau, Richard R., and David P. Redlawsk. 1997. Voting correctly. *American Political Science Review* 91:585–598.

Lau, Richard R., Parina Patel, Dalia F. Fahmy and Robert R. Kaufman. 2014. Correct voting across thirty-three democracies: A preliminary analysis. *British Journal of Political Science* 44:239–259.

Lewis-Beck, Michael S., and Mary Stegmaier. 2000. Economic determinants of electoral outcomes. *Annual Review of Political Science* 3:183–219.

Lijphart, Arend. 1999. *Patterns of democracy: Government forms and performance in thirty-six countries.* New Haven: Yale University Press.

Lubbers, Marcel, and Take Sipma. 2017. Multi-level modelling of voting behaviour. In *The SAGE handbook of electoral behaviour,* eds. Kai Arzheimer, Jocelyn Evans and Michael S. Lewis-Beck, 934–951. London: SAGE.

Lupu, Noam. 2015. Party polarization and mass partisanship: A comparative perspective. *Political Behavior* 37:331–356.

Magalhães, Pedro. 2014. Political institutions and the social anchoring of the vote. In *Elections and democracy: Representation and accountability,* ed. Jacques Thomassen, 79–98. Oxford: Oxford University Press.

Mair, Peter. 2007. Left-right orientations. In *The Oxford handbook of political behavior,* eds. Russell J. Dalton and Hans-Dieter Klingemann, 206–222. Oxford: Oxford University Press.

Mair, Peter, and Cas Mudde. 1998. The party family and its study. *Annual Review of Political Science* 1: 211–229.

Marinova, Dani. 2016. *Coping with Complexity. How voters adapt to unstable parties.* Colchester: ECPR Press.

McClurg, Scott, Casey Klofstad and Anand Sokhey. 2017. Discussion networks. In *The Oxford handbook of political networks,* eds. Jennifer Nicoll Victor, Alexander H. Montgomery and Mark Lubell, Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190228217.013.21.

Miller, Warren Edward, and J. Merrill Shanks. 1996. *The new American voter.* Cambridge, MA: Harvard University Press

Moral, Mert. 2017. The bipolar voter: On the effects of actual and perceived party polarization on voter turnout in European multiparty democracies. *Political Behavior*. https://doi.org/10.1007/s11109-016-9386-0.

Norris, Pippa. 2004. *Electoral engineering: Voting rules and political behavior*. Cambridge: Cambridge University Press.

Norris, Pippa. 2009. The globalization of comparative public opinion research. In *Handbook of comparative politics,* eds. Neil Robinson and Todd Landman, 522–540. London: Sage.

Powell, G. Bingham, and Guy D. Whitten. 1993. A Cross-national analysis of economic voting: Taking account of the political context. *American Journal of Political Science* 2:391–414.

Przeworski, Adam, and Henry Teune. 1970. *Logic of comparative social inquiry*. New York: Wiley.

Riker, William H., and Peter C. Ordeshook. 1968. A theory of the calculus of voting. *American Political Science Review* 62:25–42.

Rivers, Douglas. 1988. Heterogeneity in models of electoral choice. *American Journal of Political Science* 32:737–757.

Sartori, Giovanni. 1970. Concept misformation in comparative politics. *The American Political Science Review* 64:1033–1053.

Scheuch, Erwin K. 1969. Social context and individual behavior. In *Quantitative ecological analysis in the social sciences,* eds. Mattei Dogan and Stein Rokkan, 133–155. Cambridge: MIT Press.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *International comparative social research*, Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenbauer and Heiner Meulemann. https://doi.org/10.1007/s11577-019-00607-9.

Schmitt-Beck, Rüdiger. 2015. Data bases and statistical systems: Political behavior and elections. In *The international encyclopedia of social and behavioral sciences,* ed. James R. Wright, 824–829. Amsterdam: Elsevier.

Schoen, Harald, Sigrid Roßteutscher, Rüdiger Schmitt-Beck, Bernhard Weßels and Christof Wolf. 2017. Voters and voting in context: An introduction. In *Voters and voting in context,* eds. Harald Schoen, Sigrid Roßteutscher, Rüdiger Schmitt-Beck, Bernhard Weßels and Christof Wolf, 1–12. Oxford: Oxford University Press.

Singh, Shane P. 2010. Contextual influences on the decision calculus: A cross-national examination of proximity voting. *Electoral Studies* 29:425–434.

Singh, Shane P. 2011. Contradictory calculi. Differences in individuals' turnout decisions across electoral systems. *Political Research Quarterly* 64:646–655.

Singh, Shane P. 2015. Compulsory voting and the turnout decision calculus. *Political Studies* 63:548–568.

Singh, Shane P., and Judd Thornton. 2013. Compulsory voting and the dynamics of partisan identification. *European Journal of Political Research* 52:188–211.

Söderlund, Peter, Hanna Wass and André Blais. 2011. The impact of motivational and contextual factors on turnout in first- and second-order elections. *Electoral Studies* 30:689–699.

Solt, Frederick. 2008. Economic inequality and democratic political engagement. *American Journal of Political Science* 52:48–60.

Spies, Dennis Christopher, and Simon T. Franzmann. 2019. Party competition and vote choice. In *International comparative social research*, Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenbauer and Heiner Meulemann. https://doi.org/10.1007/s11577-019-00611-z.

Steenbergen, Marco R., and Bradford S. Jones. 2002. Modeling multilevel data structures. *American Journal of Political Science* 46:218–237.

Stockemer, Daniel. 2017. What affects voter turnout? A review article/meta-analysis of aggregate research. *Government and Opposition* 52:698–722.

Thomassen, Jacques. Ed. 2005. *The European voter. A comparative study of modern democracies.* Oxford: Oxford University Press.

Verba, Sidney, and Norman H. Nie. 1972. *Participation in America : Political democracy and social equality.* New York: Harper Row.

Verba, Sidney, Kay Lehman Schlozman and Henry E Brady. 1995. *Voice and equality: Civic voluntarism in American politics.* Cambridge, London: Harvard University Press.

Wagner, Aiko. 2013. *Die Mikrofundierung von Duvergers Gesetz. Strategisches Wahlverhalten als Wirkungsweise politischer Institutionen.* (The microfounding of Duverger's Law. Strategic voting behavior as a modus operandi of political institutions) Baden-Baden: Nomos.

Wass, Hanna, and André Blais. 2017. Turnout. In *The SAGE handbook of electoral behaviour,* eds. Kai Arzheimer, Jocelyn Evans and Michael S. Lewis-Beck, 459–487. London: SAGE.

Weßels, Bernhard, and Hermann Schmitt. 2008. Meaningful choices, political supply, and institutional effectiveness. *Electoral Studies* 27:19–30.

Weßels, Bernhard, and Hermann Schmitt. 2014. Meaningful choices: Does parties' supply matter? In *Elections and democracy: representation and accountability,* ed. Jacques Thomassen, 38–59. Oxford: Oxford University Press.

Zuckerman, Alan S. Ed. 2005. *The social logic of politics: Personal networks as contexts for political behavior.* Philadelphia: Temple University Press.

Rüdiger Schmitt-Beck 1956, Dr. phil, Professor of Political Science—Political Sociology at the University of Mannheim. Areas of Research: Electoral Behavior, Political Communication. Publications: Parties and voters at the 2013 German federal election. New York 2018 (ed. with R. Rohrschneider); Voters and voting in context. Multiple contexts and the heterogeneous German electorate. Oxford 2017 (ed. with H. Schoen, S. Roßteutscher, B. Weßels and C. Wolf).

Kölner Zeitschrift für Soziologie und Sozialpsychologie

Multilevel Models for the Analysis of Comparative Survey Data: Common Problems and Some Solutions

Alexander W. Schmidt-Catran · Malcolm Fairbrother · Hans-Jürgen Andreß

Abstract This paper provides an overview over the application of mixed models (multilevel models) to comparative survey data where the context units of interest are countries. Such analyses have gained much popularity in the last two decades but they also come with a variety of challenges, some of which are discussed here. A focus lies on the small-*N* problem, influential cases (outliers) and the issue of omitted variables at the country level. Summarizing the methodological literature, the paper provides recommendations for applied researchers when possible or otherwise points to the more detailed literature. Some solutions for the small-*N* problem and omitted variable bias are discussed in detail, recommending the pooling of multiple survey waves to increase statistical power and to allow for the estimation of within-country effects, thereby controlling for unobserved heterogeneity. All issues are illustrated using an empirical example with data from the European Social Survey. The online appendix provides detailed syntax to adopt the presented procedures to researchers' own data.

Online Appendix: http://www.schmidt-catran.de/mixedmodels.html

A. W. Schmidt-Catran (✉)
Institut für Soziologie, Lehrstuhl für Soziologie mit dem Schwerpunkt Methoden der quantitativen empirischen Sozialforschung, Goethe-Universität Frankfurt
Theodor-W.-Adorno-Platz 6, Campus Westend, 60323 Frankfurt am Main, Germany
E-Mail: alex@alexanderwschmidt.de

M. Fairbrother
Department of Sociology, Umeå University
Norra Beteendevetarhuset, Umeå universitet, 901 87 Umeå, Sweden
E-Mail: malcolm.fairbrother@umu.se

H.-J. Andreß
Fakultät für Wirtschafts- und Sozialwissenschaften, Institut für Soziologie und Sozialpsychologie, Lehrstuhl für empirische Sozial- und Wirtschaftsforschung, Universität zu Köln
Albertus-Magnus-Platz, 50923 Cologne, Germany
E-Mail: hja@wiso.uni-koeln.de

🖄 Springer

Mehrebenenmodelle zur Analyse von vergleichenden Umfragedaten: Häufige Probleme und ausgewählte Lösungsansätze

Zusammenfassung Die vorliegende Arbeit bietet einen Überblick über die Anwendung von Mehrebenenmodellen auf international vergleichende Umfragedaten. Mehrebenenanalysen, in denen die relevanten Kontexteinheiten Länder sind, haben in den letzten 2 Jahrzehnten eine weite Verbreitung gefunden, sind allerdings aus statistischer Perspektive in einigen Aspekten problematisch. Dieser Artikel zielt auf einige der Probleme ab, die bei der Anwendung von Mehrebenenanalysen auf internationale Umfragedaten auftreten. Ein Fokus liegt dabei auf dem small-*N*-Problem, einflussreichen Fällen („Ausreißern") und dem Problem unbeobachteter Heterogenität auf der Länderebene. Dieser Beitrag bietet eine Zusammenfassung der methodischen Literatur zu Mehrebenenmodellen und versucht, in Forschung Tätigen möglichst konkrete Empfehlungen zu geben oder – wo dies nicht möglich ist – auf die tiefergehende Literatur zu verweisen. Lösungsansätze für das small-*N*-Problem und das Problem unbeobachteter Heterogenität werden im Detail diskutiert. Aus dieser Diskussion ergibt sich die Empfehlung, vorhandene Wellen international vergleichender Umfragedaten zu poolen. Zur Illustration verwendet dieser Artikel ein empirisches Beispiel auf Basis der Daten des European Social Survey. Der Online-Anhang enthält zu diesen Beispielen eine detaillierte Syntax, die sich leicht für andere Daten und Forschungsfragen anpassen lässt.

1 Introduction

Multilevel models, also known as random effects, hierarchical, or mixed models, are regression models for the analysis of hierarchical data. Such models can be applied to a wide variety of data structures, but applications to two types of data are particularly common in the social sciences: (1) panel data, where measurement occasions are nested in persons or some other unit of analysis (e.g. firms, nations); and (2) datasets where the primary units of analysis (e.g. survey respondents, employees, students) are nested in higher-level social groups (e.g. nations, companies, schools). This paper focuses on the latter type, and particularly on the decisions confronting researchers analyzing comparative survey data, though it also considers insights developed in the tradition of panel data analysis.

Due to the vast increase in the availability of comparative surveys during the last two decades, the expansion of computational power, and improvements to statistical software, multilevel models have become a commonly used tool of social science. To illustrate the point, Fig. 1 shows the share of multilevel analyses out of all articles appearing in the *European Sociological Review* (ESR) from 2000 to 2016;
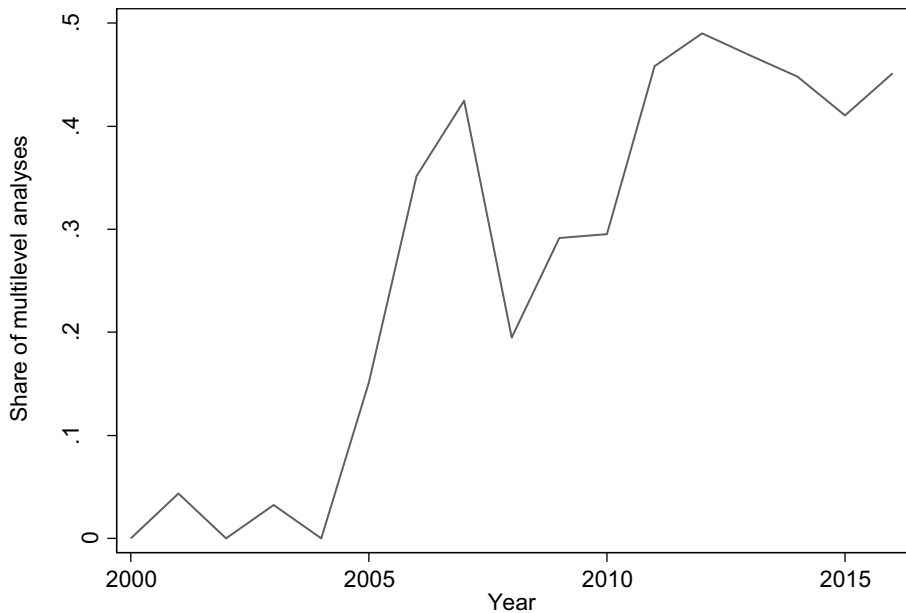
Fig. 1 Share of multilevel analyses from all publications in *European Sociological Review* (ESR). Notes: Based on a keyword search for the term "multilevel" in the search engine of ESR (on October 10, 2017) and the total number of articles published between 2000 and 2016

the proportion has reached almost 50%. Specifically with respect to comparative survey datasets (i.e., surveys conducted in multiple countries simultaneously, such as the European Social Survey or World Values Surveys), multilevel models are a popular analytical tool because they help identify how individual outcomes like attitudes and behaviors vary according to social context. All the social sciences take an interest in how people's economic, social, political, or institutional circumstances shape their lives.

In the face of the dramatically expanding popularity of multilevel modelling, and the creative application of such models to new kinds of data and research questions, methodologists have started to point out problems and challenges in specific analyses and common research practices (e.g., Bryan and Jenkins 2016; Heisig and Schaeffer 2018; Schmidt-Catran and Fairbrother 2016; Te Grotenhuis et al. 2015). Drawing on this literature, this paper discusses some issues particularly relevant for analyses of comparative survey data: statistical inference with nonrandom samples; the problem of having only a small number of higher-level units; and issues of omitted variable bias. These issues are not unique to analyses using multilevel models, but are rather general problems for all kinds of regression techniques, and therefore where appropriate we bring in insights from more general literature.

Throughout the discussion, to provide a concrete illustration of the general points we make, the paper uses a running example inspired by a recent study by Te Grotenhuis et al. (2015). Investigating the relationship between social security and religious involvement, Te Grotenhuis et al. (2015) demonstrate, in their words, "the danger of

testing hypotheses cross-nationally." Substantively, their study tests whether state-provided social security, along with general increases in economic wealth, can substitute for some of the benefits to individuals that come from religion. For a detailed theoretical treatment of this hypothesis, we refer readers to the paper by Te Grotenhuis et al. (2015) and the literature cited therein. Methodologically, Te Grotenhuis et al. (2015) used Eurobarometer data, but we employ data from the European Social Survey (ESS; 2016), like a prior study on the same subject by Immerzeel and Tubergen (2013). All analyses in this paper can be replicated using the Stata data set and do-file provided in the online appendix.[1]

We will focus on linear multilevel models for continuous dependent variables. We begin with a very brief introduction to these models and their assumptions. For ease of presentation, we will from now on always refer to the example of individuals (at level 1) nested in countries (at level 2).

1.1 A Very Brief Introduction into Multilevel Models

A multilevel model for continuous dependent variables is a generalization of the linear regression model, which includes a separate error component at each of its levels and may be written as

$$y_{ji} = \beta_0 + \beta_1 x_{1ji} + ... + \beta_k x_{kji} + \gamma_1 z_{1j} + ... + \gamma_l z_{lj} + u_j + e_{ji},$$

where the index i indicates individuals and j indexes countries. From left to right, y_{ji} is an individual-level outcome (e. g. church attendance), and the model includes 1 to k individual-level variables x (e. g. age, education), with corresponding coefficients β, and 1 to l country-level variables z (e. g. social spending, GDP/capita where GDP is gross domestic product), with the coefficients γ. These coefficients are conventionally also referred to as fixed effects. In addition, the model also includes random effects (or error terms) at the individual (e_{ji}) and the country level (u_j), both of which are assumed to be normally distributed with a mean of zero and a constant variance and to be uncorrelated with each other and with the observed variables. Where the purpose of the analysis is to identify a causal relationship, the latter assumption is called the *exogeneity* assumption and is crucial for the estimation of unbiased fixed effects. The variances of the error terms are estimated, with the term u_j capturing the country-level disturbances from the overall intercept β_0. Each individual element of u_j is called a random intercept.

In fitting multilevel models, it is common for researchers to calculate the intraclass correlation (ICC): the share of the total unexplained variance attributable to the higher level. The formula for this is $\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$, where σ_u^2 and σ_e^2 are the variances of the individual- and country-level random effects, respectively (Hox 2010, p. 15). In an empty model—a model that includes no observed independent variables—the ICC indicates what proportion of the overall variance is at the country level, a figure equivalent to the average correlation of observations within countries. If it were zero, the observations would not violate the assumption of independence,

---

[1] The online appendix is available at www.schmidt-catran.de/mixedmodels.html.

there would be no intercountry differences to explain, and a multilevel model would not be necessary.

Considering our research example, we can examine the degree to which religious involvement varies across countries, ahead of explaining that variation by social security and other variables. We follow Grotenhuis et al. (2015) in operationalizing religious involvement as church attendance, and in their treatment of this variable as interval-scaled (such that a linear model can be estimated). Using the ESS wave from 2014, we find $\rho = 0.335/(0.335 + 2.030) = 0.142$. Thus, 14.2% of the total variance in church attendance is attributable to the country level (Table 4 in the appendix describes the sample used for this analysis).

The model above can be extended and made more flexible, allowing not only for the intercept β_0 to vary cross-nationally, but also for any individual-level variable's effect to vary between countries. Such a model is often called a random intercept and random slope model:

$$y_{ji} = \beta_0 + \beta_1 x_{1ji} + ... + \beta_k x_{kji} + \gamma_1 z_{1j} + ... + \gamma_l z_{lj} +$$
$$u_{0j} + u_{1j} x_{1ji} + ... + u_{kj} x_{kji} + e_{ji}$$

The random effects u_{1j} to u_{kj} are country-level variances that capture the deviation of country-specific slopes from the average effects across all countries (β_1 to β_k).[2] Thereby the model explicitly allows for heteroscedasticity due to effect heterogeneity in individual-level variables. The random effects at the country level—random intercepts and slopes—are assumed to have a multivariate normal distribution and be independent of the idiosyncratic error term e_{ji}.

The covariances between random intercept and slopes, however, are not or rather should not be assumed to be zero (Hox 2010, p. 13). This means we generally estimate a variance-covariance matrix for the random effects (intercepts and slopes) of the form

$$\Sigma_u = \begin{pmatrix} \sigma_{u0}^2 & \cdots & \sigma_{u0}\sigma_{uk} \\ \vdots & \ddots & \vdots \\ \sigma_{u0}\sigma_{uk} & \cdots & \sigma_{uk}^2 \end{pmatrix},$$

where the diagonals of this matrix describe the variances of random effects and the off-diagonals include the covariances between each pair of random effects. The number of unique entries in this symmetric matrix, together with the number of country-level variables in the fixed part of the model, constitutes the total number of parameters estimated from country-level information. For example, a model including two country-level variables (e.g. social spending and GDP/capita) and three random slopes of individual-level variables (e.g. gender, age, education) will estimate 12 country-level parameters in total: two country-level fixed effects, four random effect variances (intercept plus three slopes) and six covariances between

---

[2] Note that the country-level random effects now have an additional subscript $(0,1, ... ,k)$, indicating to which fixed effect the random effect belongs.

Table 1 Random intercept models of church attendance, European Social Survey (ESS) 2014

Variable	M0		M1		M2	
Individual-level variables						
Urban vs. rural	–	–	0.0669	***	0.0669	***
Education (in years)	–	–	–0.0156	***	–0.0156	***
Subjective income	–	–	–0.0120	–	–0.0123	–
Male (ref = female)	–	–	–0.2077	***	–0.2076	***
Age	–	–	0.0091	***	0.0091	***
Country-level variables						
Social spending (% of GDP)	–	–	–	–	–0.0465	–
GDP/capita	–	–	–	–	0.0000	–
Average urban vs. rural	–	–	–	–	0.4394	–
Average education	–	–	–	–	–0.0996	–
Constant	1.4620	***	1.1409	***	2.7045	–
Variance components						
Country level	0.3348	***	0.3265	***	0.3005	**
Individual level	2.0231	***	1.9662	***	1.9662	***
Statistics						
N (Country)	20		20		20	
n (Individual)	37,028		37,028		37,028	

See text for explanation of M0–M2

*$p < 0.05$, **$p < 0.01$, ***$p < 0.01$ (two-sided tests). All models are estimated via Restricted Maximum Likelihood. Models based on ESS data 2014 (compare Table 4)

GDP gross domestic product

them.[3] We will return to this point when discussing the small-*N* problem; suffice to say here that it can be hard not to ask too much of the data while still accounting for an adequate number of fixed and random country-level effects.

Table 1 presents a first analysis of the example data, using a single wave of the ESS.[4] It shows the basic stepwise procedure usually applied with multilevel models. Model M0 is an empty model which is used to decompose the total variance into its individual- and country-level components. As already noted, 14.2% of the variance is at the country level. The next step, as is typical, adds the individual-level variables to the model (M1). Older people, people living in rural areas, women, and people with less education attend religious services more often. Subjective income does not have a significant effect.

By adding individual-level variables first, the analysis reveals how much of the country-level variance can be explained by individual-level differences: $1 - (0.3265/0.3348) \approx 0.025$. This is, 2.5% of the differences between countries can be explained by differences in the populations of the individuals living in those coun-

---

[3] A symmetric variance-covariance matrix of size m contains $m \cdot ((m + 1)/2)$ unique entries, m of which are variances and the rest being covariances.

[4] A detailed description of all involved variables, their descriptive statistics and correlations, can be found in the online appendix to this paper.

tries. This is often called a *compositional effect* and in this application only a small fraction of the between-country variance can be explained by differences in composition, which means there is substantial variance left that is due to country-level effects. If most of the variance between countries could be explained by compositional effects, we would have to conclude that any differences between countries are not related to contextual effects—only to characteristics of the individuals making up the populations of these countries.

The third step (M2) adds country-level effects, which after controlling for compositional effects can be interpreted as contextual effects. These reduce the unexplained country-level variance from Model M1 by about 8% ($1 - (0.3005/0.3265) \approx 0.080$). Social spending (as % of GDP) has the hypothesized negative effect on church attendance, consistent with the results of Immerzeel and Tubergen (2013). However, in contrast to their analysis, the effect of social spending is not significant, which may not be a surprise given that we use 20 observations to estimate five parameters (four fixed and a random effect).

A fourth step could be to test for random slopes and a fifth one the inclusion of cross-level interaction effects, which might explain the variation in individual-level effects identified in step four. (For a detailed description of the stepwise procedure see Hox 2010, p. 54 ff.). Following Te Grotenhuis et al. (2015) and Immerzeel and Tubergen (2013) we are not interested in cross-level interactions and therefore stop here.[5]

This has clearly been a very brief introduction, but it should have served the purposes of introducing some notation and core ideas, and starting some analysis of the example dataset. For a detailed introduction to multilevel models, readers may wish to consult one of the classic introductory textbooks by Hox (2010) or Snijders and Bosker (2012). Rabe-Hesketh and Skrondal (2012) provide an easily accessible introduction into multilevel models using Stata. Gelman and Hill (2007) discuss multilevel models in both frequentist and Bayesian frameworks, using the software packages R and BUGS.

2 Challenges in Analyses of Comparative Survey Data

Multilevel analyses of comparative survey data are not without their complications. Measurement equivalence with respect to latent variables, for example, can be a limitation—as explained in the paper by Cieciuch et al. in this special issue. Setting aside problems of measurement, however, here we address a different set of issues.

First, the countries included in international surveys are never random samples, but are instead selected or self-selected in ways that make them, effectively, convenience samples (Ebbinghaus 2005). This raises questions about the justifiability of statistical inferences to a larger population of countries, and about the use of infer-

---

[5] But see Heisig et al. (2017) who argue for the inclusion of random slopes even if the research interest is not in cross-level interactions, i.e. in explaining differences in individual-level effects by country-level characteristics. Barr et al. (2013) and Bell et al. (2019) also demonstrate and discuss the importance of random slopes.

ential statistics generally (see Goerres et al. 2019).. Second, the number of countries included in such surveys is typically rather small. Most international surveys include about 30 countries (e. g. European Social Survey [ESS]; European Union Statistics on Income and Living Conditions), and only a few include more than about 50 (such as by combining samples from the World Values Surveys [WVS] and European Values Studies [EVS]). Many studies analyze an even more limited number of countries because right-hand-side national-level variables are often unavailable for some countries (Bryan and Jenkins 2016, p. 3). This increases both the selectivity of the sample (Ebbinghaus, 2005: p. 136) and the severity of the small-N problem. Third, in a model aiming at identifying a causal relationship, the small degrees of freedom at the country level limits the number of higher-level control variables that can be included (see Goldthorpe 1997, p. 5f.; Jaeger 2013). We discuss each of these issues in turn.

2.1 Nonrandom Country-level Sampling in International Surveys

From the point of view of some researchers, inferential statistics are only applicable to random samples, which leaves rather unclear the statistical status of analyses conducted on, in effect, convenience samples of countries. Some researchers conclude that inferential statistics are completely meaningless in these settings; others argue that the use of inferential statistics is justified even with these nonrandom samples (compare Ebbinghaus 2005 and Babones 2013, 107 ff.).

When observations on entire countries are the units of analysis, as in the analysis of pooled time-series cross-section data, the research community tends not to object to the use of inferential statistics. That is true even though the nonrandom sampling of countries prohibits the straightforward generalization of findings to a larger population of countries; instead "all inferences of interest are conditional on the observed units" (Beck 2001, p. 273).

While samples of countries in international surveys are clearly not random—and therefore country-level effects must be viewed as conditional on the specific sample of countries—at the very least individuals within countries are sampled at random.[6] Therefore individual-level results should be generalizable within countries. However, individual-level effects in multilevel models are not only identified by variation within countries, but also by between-country variation (see Bell et al. 2018; Andress et al. 2013, particularly p. 157 ff.). This also implies that inference to the populations within countries may be problematic. One way of addressing this problem is to group-mean center the individual-level variables, stripping them of any country-level variation (Hox 2010, p. 68 ff.; Bell et al. 2018; see Fairbrother 2016 for an applied example).[7] Enders and Tofighi (2007) suggest doing this if the interest is purely in individual-level relationships, though multilevel models are typically employed because of a specific interest in country-level effects or their interactions with individual-level variables. However, if the interest is really just in individual-

---

[6] The issue of nonrandom missing values, i. e. sample selection effects at the individual level, is left aside here.

[7] This is equivalent to the introduction of country-dummies, i. e. country fixed effects.

level effects, other modelling techniques may be better suited (Bryan and Jenkins 2016).

There is an informal working consensus in the literature that inferential statistics are also relevant at the country level, despite the fact that the countries included in international surveys are not selected at random from the population of all countries. The basic argument for this is that there are several other relevant sources of random variation, aside from sampling errors (e.g. measurement errors, omitted variables), which justify the usefulness of *p*-values for separating real effects from random noise.

What does this imply for the research example? The ESS data used here are obviously not a random sample of countries and certainly cannot be used to generalize results to the world population of countries in a statistical sense (see Table 4 for a sample description). The original data set from the ESS included 32 countries[8] and covered most EU member countries. So, one might think that models based on this data should allow to make statements about EU member countries. Due to missing data for social spending and/or GDP per capita, however, some countries were excluded from the analysis. If the missing observations were truly random, the data would allow for generalization to the population of EU member countries.[9] However, the excluded countries are Bulgaria, Cyprus, Croatia, Lithuania and the Ukraine, seemingly not a random set of countries.

2.2 The Small-*N* Problem

We coded articles with multilevel analyses in the *European Sociological Review* and found 103 such analyses using countries as contextual units. In those analyses, the average number of countries is 22.6 (Min = 9, Max = 78). Setting aside the issue of nonrandom sampling, then, what are the implications of using such small country-level samples in multilevel models of comparative survey data?

First, with higher-level *N*s in this range, the estimated coefficients of country-level variables will often be quite sensitive to single (outlying) countries (Wilkes et al. 2007; Van der Meer et al. 2010). Figure 2 tests this possibility for the example data. It presents the simple bivariate relationship between church attendance and social spending (as % of GDP) using the complete ESS data (rounds 1 to 7, compare Table 4), aggregating each variable to the country level. The set of grey lines describes the bivariate relationships when each country is excluded from the sample one at a time; the black line indicates the relationship in the full data. In terms of correlations the strength of the relationship in the full sample is –0.34. When leaving out each country once, it varies between –0.27 (leaving out Turkey) and –0.41 (leaving out Estonia), a substantive difference of about 52%.

One can take two perspectives on this. On the one hand, we can accept that any statistical inference is conditional on the sample and thus it is to be expected that different samples will provide results that deviate from each other by more than

---

[8] The data set has been obtained from the cumulative data wizard, which does exclude Albania, Kosovo and Latvia.

[9] Ignoring for now the fact that two EU members are not in the sample: Malta and Latvia.
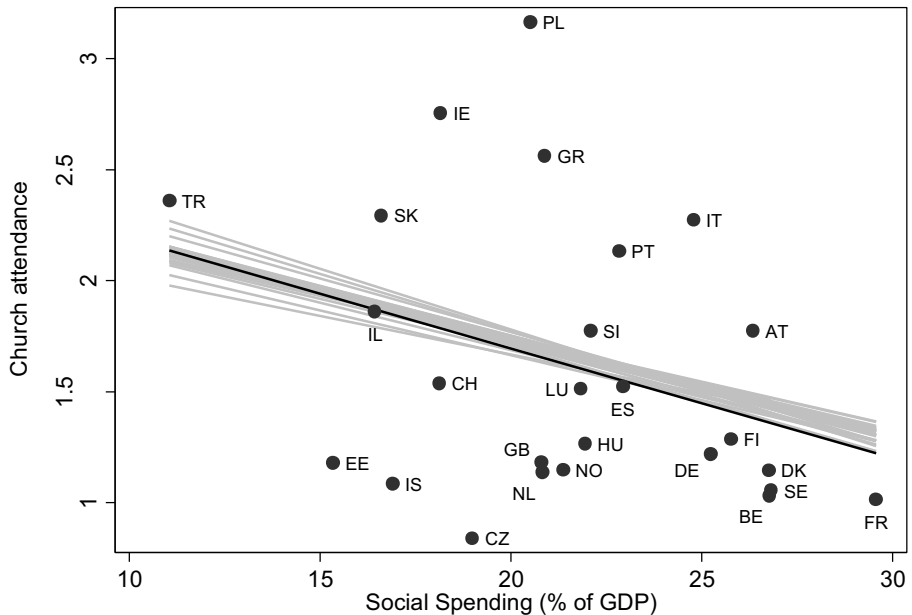
Fig. 2 Bivariate country-level relationships between social spending and church attendance. Notes: Based on ESS data 2002–2014 (compare Table 4). The *black line* represents the association in the full sample, while the *grey lines* represent the associations when leaving out each country one at a time. *AT* Austria, *BE* Belgium, *CH* Switzerland, *CZ* Czech Republic, *DE* Germany, *DK* Denmark, *EE* Estonia, *ES* Spain, *FI* Finland, *FR* France, *GB* Great Britain, *GR* Greece, *HU* Hungary, *IE* Ireland, *IL* Israel, *IS* Iceland, *IT* Italy, *LU* Luxembourg, *NL* Netherlands, *NO* Norway, *PL* Poland, *PT* Portugal, *SE* Sweden, *SI* Slovenia, *SK* Slovakia, *TR* Turkey

what could be explained by sampling error. On the other hand, the model parameters ought to describe the data in the best possible way. In some cases, outliers can have such strong influences that the regression line primarily describes the position of the outlier relative to the rest of the countries, rather than the relationship in the bulk of the data. Van der Meer et al. (2010) provide such an example where a strong positive relationship between church attendance and volunteering completely dissolves once outliers are considered (also see Hox 2010, p. 29).

Investigating outlying cases can be done graphically by means of scatter plots, as in Fig. 2. But scatter plots show only simple bivariate relationships of aggregated data and it may be hard to decide which countries are too influential.[10] An alternative are outlier statistics such as Cook's Distance (Cook 1977) or DFBETAs (Belsley et al. 1980, p. 13), which can also be applied to multilevel models (Snijders and Berkhof 2008, p. 157). Later, in Sect. 4, we demonstrate how to apply these outlier statistics to multilevel models. For now, we simply note that the bivariate cross-sectional relationship between aggregated church attendance and social spending is in line with our expectations: higher spending is associated with less religious

---

[10] See Bowers and Drake (2005) for more information on how to use exploratory data analysis and visualization when the number of level 2 units is small.

involvement. While the estimated relationship is dependent on the specific countries in the sample, ranging from –0.27 to –0.41, this influence may not be regarded as overly problematic since the complete range of values confirms our theory.

Second, while all available estimation techniques for multilevel models (e.g., Full Maximum Likelihood [FML], Restricted Maximum Likelihood [RML]) are consistent, meaning that they converge to the true parameters with increasing sample size, their behavior in small samples is sometimes problematic (Hox 2010, p. 40 ff.). This issue has motivated several methodological studies asking variations on "how many countries do you need for multilevel modelling?" (Stegmueller 2013; also see Maas and Hox 2005; Bell et al. 2014; Bryan and Jenkins 2016; Heisig et al. 2017; Elff et al. 2016). Such studies have also examined how different estimators behave under conditions of varying sample sizes, violations of the normality assumptions, and other data characteristics.

Both FML and RML, the most commonly applied estimators (Hox 2010, p. 40), provide unbiased point estimates of the fixed effects in *linear mixed models* but the variance components and their standard errors (SEs) are underestimated in small samples. Due to the uncertainty in the random part of the model, the SEs of the fixed effects are also biased downwards, resulting in unclear distributions of test statistics and the risk of performing anticonservative tests[11] (Bryan and Jenkins 2016, p. 7; also see Elff et al. 2016, p. 14 ff. for some solutions). The same biases are found with nonlinear multilevel models with the additional caveat that the unbiasedness of fixed effects coefficients cannot be clearly demonstrated for these models (Bryan and Jenkins 2016, p. 7 f.).

The small-sample bias appears to be much stronger with FML than with RML (Hox 2010, p. 41). In fact, RML was introduced to deal with the FML bias in variance component estimation (Patterson and Thompson 1971). Nevertheless, Maas and Hox (2005) find somewhat substantial biases of RML with small samples. Most studies, however, find very small or nonexisting biases with RML even if the country-level N is as small as 10 or 5 (Bryan and Jenkins 2016; Browne and Draper 2000; Elff et al. 2016). With FML, in contrast, the bias can be quite substantial with small samples at the country level (Elff et al. 2016; Browne and Draper 2000).

Should one always prefer RML over FML then? FML has one clear advantage vis-a-vis RML, which is that it allows the use of likelihood-ratio tests (LR tests) to compare nested models (Hox 2010, p. 41).[12] Such comparisons can be very useful in the process of model building and may also be helpful for testing hypotheses. Thus, there is a trade-off between RML and FML: If the bias of FML estimates is negligible, FML may be preferred over RML. Above an ICC of 0.142 was obtained from the example data on church attendance. This model was estimated with RML. Using FML, the ICC is estimated to be 0.136. As expected the FML estimates yield a smaller variance at the country level but the difference may be regarded as trivial.

---

[11] With anticonservative tests, the risk of falsely rejecting the null hypothesis of no effect increases. In other words, results look too significant.

[12] To be precise, RML does also allow to compare nested models but only if they differ in their random but not in the fixed part.

This is consistent with a recent simulation study by Elff et al. (2016, p. 13 ff.), who show that the bias of FML compared to RML is substantial with fewer than 15 countries but relatively unimportant with 20 or more. Nevertheless, we suggest that instead of applying simple rules of thumb, researchers should compare the results of both methods to decide whether the bias of FML can be ignored. Formulating a rule of thumb is difficult because the performance of any estimator is highly dependent on the specifics of the data and the complexity of the model fitted to them (Bryan and Jenkins 2016, p. 8).

In addition to FML and RML, there are several other estimators for multilevel models available: Generalized Least Squares (GLS), Generalized Estimation Equation (GEE), and Bayesian methods. GLS is asymptotically equivalent to FML but in practice often less efficient (Hox 2010, p. 42 f.). GEE and cluster robust SEs can be a remedy against too optimistic (underestimated) SEs but also involve the risk of obtaining overestimated SEs (Hox 2010, p. 262 f.), which are to be avoided given that the statistical power to estimate country-level effects is rather small anyway. With violated distributional assumptions, which can be a consequence of a small N at the country level, bootstrapping can reduce the bias in SEs but it is implemented only in a few statistical software packages, is computationally quite demanding, and is not *per se* useful with small samples (Hox 2010, p. 264 ff.). For now valid bootstrapping with multilevel models is implemented only in MLwiN.

Finally, there is the option to turn away from classical frequentists statistics and use Bayesian methods. Obviously, this paper does not offer the space to deal with Bayesian methods in any detail. Readers who are interested in Bayesian multilevel modelling may want to start with Jackman (2009), who gives a general introduction into Bayesian modelling and treats multilevel models in Chap. 7. Hox (2010) has a large section on Bayesian multilevel modelling (p. 271 ff.); Gelman and Hill (2007) and Draper (2008) may also be good starting points.

In a nutshell, frequentists view the population *parameter as* an unknown but *fixed* quantity, which they estimate from data. The uncertainty in the estimate results from the sampling distribution, i. e. the distribution of the parameter in an indefinite number of samples. Bayesians view the *data as fixed* and the parameter of interest as an unknown quantity that must be described by probabilistic statements and can always be updated by data. This leads Bayesians to formulate a prior distribution, which reflects the belief, or rather (un)certainty, about the parameter before seeing the data. The data then is used to update the prior distribution by conditioning it on the observed data, resulting in the so-called posterior distribution. This posterior distribution, the result of the analysis, characterizes the researcher's new beliefs about the parameter, in light of the prior distribution and the likelihood of the data.

With large Ns and uninformative priors—priors that do not favor any specific parameter region—Bayesian estimates are identical to ML estimates. There is some controversy about the question of whether a Bayesian approach deals better with the small-N problem than frequentist analysis does. Stegmueller (2013) claims that Bayesian methods have an inherent advantage over frequentists methods when it comes to the analysis of hierarchical data with few clusters. Elff et al. (2016) disagree. In our reading of the literature, the unbiasedness of Bayesian methods with small Ns is more straightforward than it is for the frequentist approach, within which

special adjustments and estimation methods are needed for small samples (compare Elff et al. 2016). On the other hand, some literature suggests that seemingly uninformative priors can result in biased Bayesian estimates when the sample size is small (Gelman 2006; Van Erp et al. 2017). In sum, there does not seem to be a general advantage of Bayesian methods over frequentist approaches.

It is a different game, of course, if a researcher has useful prior information on parameters, in which case the Bayesian approach can be recommended. But we have yet to see a convincing implementation of a model using informative priors in the context of comparative survey data. It is telling that out of the (just) six Bayesian multilevel analyses published in ESR since 2000[13] none used (true) informative priors—one analyses (Sutton 2012) implemented so-called skeptical priors, which drag coefficients slightly towards zero to create conservative tests.

2.3 Omitted Variable Bias

To identify a causal effect of a variable x on y, any alternative explanation for an association between them must be ruled out. In experiments this is of course achieved by randomization. With observational data, it must be done by partialing out the effects of any variable that is a cause of both y and x. Technically, the omission of a variable which affects y and is related to x violates the exogeneity assumption and therefore results in biased coefficient estimates (Wooldridge 2013, p. 88 ff., also see 45 ff.). This very basic insight is no different for multilevel models (Kim and Frees 2006).

However, with multilevel models fitted to comparative survey data, the small-N problem makes the issue of omitted variables even more delicate: First, as we argued above, the limited degrees of freedom at the country level create a trade-off between the need to control for all necessary variables and respecting the limits of what the data can do (Heisig et al. 2017). Second, country-level characteristics of interest are often strongly correlated with each other and with necessary control variables (Babones 2013, p. 94 ff.).[14] Additionally, any attempts to control for an adequate number of country-level (fixed and random) effects are practically limited far below the theoretically absolute limit set by the country-level degrees of freedom because multilevel models tend to run into convergence problems if they include too many covariates at the country level (Heisig et al. 2017, p. 823 f). This combination of high multicollinearity coupled with few degrees of freedom will often result in inefficient estimates and thereby create the temptation to ignore important variables (Arceneaux and Huber 2007).

This has led to a questionable practice in applied research where many researchers make arguments like this: "If all country-level variables are included at the same

---

[13] Brännström (2008); Sutton (2012); Stadelmann-Steffen (2012); Stegmueller et al. (2012); Giger (2012); Mewes (2014).

[14] In the example data, the country-level variables are not too strongly related. The average (absolute) correlation across the four variables amounts to 0.31 (min = 0.19, max = 0.47), so collinearity is not a pressing issue. However, it is much stronger than the average (absolute) correlation across the individual-level variables which is 0.09 (min = 0.01, max = 0.25).

time, nothing is significant; so, I test and/or control each variable separately".[15] From a causal identification standpoint this strategy is problematic. This is not to say that researchers should include any (control) variable they can think of. In contrast, the model building strategies developed in the framework of directed acyclical graphs provide very good guidance on which variables need to be included in a model and which *not* (for an overview, see Elwert 2013). But to control only piecewise—one variable at a time— is certainly not a good strategy to identify causal effects.

Third, with countries it is arguably very difficult to operationalize all relevant factors (Babones 2013, Chap. 3). Thus, biased estimates due to omitted variables are quite likely outcomes in the analysis of comparative survey data—maybe even more so than with plain individual-level analyses, where the available degrees of freedom tend to be much higher, and measurement in many domains, specifically of latent variables, is arguably easier (Fontaine 2015). There are good reasons to be cautious before concluding that the model has no omitted variables, even if we can include all *available* variables without running into issues of nonconvergence or multicollinearity. After about a decade of elated investigations into country effects, social science researchers started to increasingly worry about such unobserved heterogeneity (for examples, see Fairbrother 2013, p. 911; Jaeger 2013, p. 156; Wulfgramm 2014, p. 263; Schmidt-Catran 2016, p. 124; Te Grotenhuis et al. 2015, p. 644; Finseraas 2012, p. 167).

3 Some Solutions and Caveats

With just a few countries in cross-sectional analyses, and few degrees of freedom at the country level, models may yield imprecise estimates of country-level effects. One way to get more variation at the country level, however, is to observe the same countries multiple times. And many international surveys have now been fielded on multiple occasions (e. g. ESS, ISSP, EVS, WVS), providing an opportunity to pool comparative survey data across time. The resulting data structure may be called comparative *longitudinal* survey data (Fairbrother 2014) and promises to not only increase statistical power but also to provide less biased estimates in the presence of unobserved country-heterogeneity. The former is a direct result of pooling across time, while the latter can be achieved by the identification of country-level effects via within-country variation, i. e. changes of country-level variables over time.

3.1 Comparative Longitudinal Survey Data

As Schmidt-Catran and Fairbrother (2016, p. 26) show in their literature review, many researchers have attempted to apply multilevel models to comparative longitudinal survey data. But they also demonstrate that there are right and wrong ways of analyzing such data, and previous studies have often used problematic specifications.

---

[15] An example of such a paper is Semyonov et al. (2006, p. 437): "Because of restrictions associated with the limited degrees of freedom at the country level, only three hierarchical linear model equations are estimated [...], with each equation including only one country-level variable."

Specifically, the introduction of a longitudinal dimension into the data creates an additional level in the hierarchical structure of the data, and this level must be accounted for to obtain unbiased SEs. In other words, incorrectly specifying the statistical model can lead to significance levels that are not actually supported by the data. Moreover, Schmidt-Catran and Fairbrother (2016, p. 30, 34) also demonstrate that a failure to model the correct random effects structure may not only yield overly optimistic SEs, but also biased coefficient estimates.

So, what is the correct hierarchical structure for a given analysis? This depends on two questions: First, at which levels are the variables measured and, second, at which levels is there variation in the data? Comparative longitudinal survey data can be viewed as having four levels: countries, survey waves (typically years, which will be used synonymously from here), combinations of countries and waves (here called country-years), and individuals. Thus, there are potentially three levels above the individuals (years, countries and country-years). At each of these levels there may be variation, meaning the observations within these clusters can be dependent. For example, individuals within the same countries are more similar than individuals from different countries; but they may even be more similar if they are observed in the same year. Alternatively, individuals observed in the same year may be more similar than individuals observed in different years, even if they are observed in different countries. Such variation needs to be accounted for by random or fixed effects. The latter can be done via the introduction of dummy variables for the clusters.

Including such dummies, however, takes up all the degrees of freedom at that level, which means no variables can be included at this level.[16] Thus, for each variable of interest, there needs to be a corresponding level in the random part of the model. This leaves only levels as candidates for cluster-dummies at which no variables of interest are measured. The final question then is the following: At what level is a variable measured? In the simple two-level model from above, with cross-sectional data, this question is easy to answer. Individual-level variables (e.g. age, gender) are measured at the individual level and country characteristics (e.g. social spending, GDP/capita) are measured at the country level.

When a longitudinal dimension comes into play, this question becomes more complicated. By definition, a cluster-level variable must be constant within clusters. Thus, a country-level variable that changes over time, like social spending, is not a country-level variable. For this reason, Schmidt-Catran and Fairbrother (2016) argue that comparative longitudinal survey data are—in most cases—best analyzed with the following model:

$$y_{jti} = \beta_0 + \sum_{t=1}^{T} \delta_t D_t + \beta_1 x_{1jti} + \dots + \beta_k x_{kjti} + \gamma_1 z_{1jt} + \dots + \gamma_l z_{ljt}$$
$$+ u_j + u_{jt} + e_{jti}$$

---

[16] Technically, there is perfect collinearity between country-level variables and country-dummies.

This is a hierarchical three-level model with individuals (i) nested in country-years (jt) nested in countries (j). The term u_j captures (unexplained) variance between countries and u_{jt} accounts for the (unexplained) variance within countries over time. The potential variance at the year level is not modelled via random effects but with year-dummies ($\sum_{t=1}^{T} \delta_t D_t$). This model allows for the inclusion of time-constant country-level variables (e.g. legal tradition) and of time-varying country-level variables (e.g. social spending); note that the z-variables now have the indices jt because they can (but need not) vary within countries over time. If researchers have a genuine interest in year-level variables (e.g. number of global terror attacks), this model does not work and the model of choice would be a four-level model with individuals nested in country-years, which are cross-classified in countries and years (for more details, see Schmidt-Catran and Fairbrother 2016).

Let us see how our research example plays out with this model. While the models in Table 1 have been fitted to the 2014 wave of the ESS only, the models presented in Table 2 are based on all available ESS data (compare Table 4). Model M3 uses the specification presented above—a three-level model with individuals nested in country-years nested in countries. Model M4 is identical in the fixed part but is a two-level model with individuals nested in countries, i.e. it omits the country-year level. This is a common mistake (compare Schmidt-Catran and Fairbrother 2016, p. 26), as many researchers assume that variables which capture country characteristics are just country-level variables, and do not need a country-year level random effect. As explained above, this is not true if these variables vary over time, as they do in the research example.

Using the pooled data approach and the correct random effects structure (M3), we now find a significant negative effect of social spending, in line with our hypothesis and the result of Immerzeel and Tubergen (2013). Note that the effect of social spending is much weaker than in Table 1 (–0.0137 vs. –0.0465), but it is nonetheless statistically significant. Model M4 demonstrates how a failure to include country-years as a separate level will provide anticonservative SEs. While the point estimates in M3 and M4 are very similar to each other, the z-statistics are much higher in the incorrectly specified model M4 (|z| = 6.32 as compared to |z| = 2.2). The latter model erroneously treats social spending as an individual-level variable, since it cannot be a country-level variable—because it is not constant within countries.[17]

Model M5 in Table 2 is a two-level model but its random effects structure matches the fixed effects. That is, all country-level variables in the fixed part of the third model have been entered as means, across all years; so they are constant within each country. Consequently, this model should yield correct SEs but it does not benefit from the increase in statistical power. In fact, we gain statistical power at the individual level, where we now have many more observations than in the models in Table 1, but not at the country level.[18] Statistical power at the individual level, however, is typically not scarce with comparative longitudinal survey data and this

---

[17] Note that this is an oversimplification. Technically, the level at which a variable is measured is not one specific level but it depends on how the variance components of a variable distribute over the levels.

[18] Except for the fact that we now include six additional countries which have been in the ESS at some point but not in the 2014 wave used in Table 1.

Table 2 Random intercept models of church attendance, European Social Survey (ESS) 2002–2014

Variable	M3 b/|z|		M4 b/|z|		M5 b/|z|	
Individual-level variables						
Urban vs. rural	0.094	***	0.0939	***	0.0937	***
	40.86	–	40.8	–	40.78	–
Education (in years)	–0.0131	***	–0.0131	***	–0.0134	***
	17.7	–	17.74	–	18.17	–
Subjective income	–0.0272	***	–0.0292	***	–0.0309	***
	7.72	–	8.3	–	8.79	–
Male (ref = female)	–0.2493	***	–0.2499	***	–0.2501	***
	46.5	–	46.56	–	46.6	–
Age	0.0116	***	0.0117	***	0.0117	***
	77.08	–	77.35	–	77.37	–
Country-level variables						
Social spending (% of GDP)	–0.0137	*	–0.0152	***	–0.0385	–
	2.2	–	6.32	–	1.33	–
GDP/capita	–0.0000	–	–0.0000	–	–0.0000	–
	1.01	–	1.83	–	1.17	–
Average urban vs. rural	–0.0542	–	–0.0407	–	0.2738	–
	0.65	–	1.27	–	0.66	–
Average education	–0.0277	–	–0.0198	*	–0.1098	–
	1.3	–	2.32	–	1.37	–
Year FEs						
2004	0.0013	–	0.0009	–	–0.0041	–
	0.05	–	0.09	–	0.41	–
2006	–0.0187	–	–0.0268	–	–0.0473	***
	0.49	–	1.78	–	4.47	–
2008	–0.0158	–	–0.0269	–	–0.0649	***
	0.33	–	1.38	–	6.41	–
2010	–0.0278	–	–0.0388	–	–0.1144	***
	0.52	–	1.8	–	11.22	–
2012	–0.0366	–	–0.048	–	–0.1269	***
	0.6	–	1.9	–	12.43	–
2014	–0.0109	–	–0.0271	–	–0.1171	***
	0.15	–	0.92	–	11.27	–
Constant	2.0519	***	1.9339	***	3.0919	*
	5.15	–	9.91	–	2.55	–
Variance components						
Country	0.339	***	0.3442	***	0.3267	***
Country-year	0.0063	***	–	–	–	–
Individual	1.968	***	1.9729	***	1.9732	***

Table 2 (Continued)

Variable	M3	M4	M5
	b/|z|	b/|z|	b/|z|
Statistics			
N (country)	26	26	26
N (country-years)	149	–	–
n (individuals)	277,505	277,505	277,505

See text for explanation of M3–M5

*p<0.05, **p<0.01, ***p<0.01 (two-sided tests). All models are estimated via Restricted Maximum Likelihood. Models based on ESS data 2002–2014 (compare Table 4). Country-level variables in Models M3 and M4 are original variables (yearly record), while in M5 they are averaged across years for each country. Note that Model M4 is incorrectly specified for demonstrational purposes

*GDP* gross domestic product, *FE* Fixed Effects

is not a recommendation to estimate such models. In fact, the model is presented to motivate the next section. A comparison of the effect of social spending in Models M3 and M5 reveals that it is much larger in the latter, where it is close to the estimate from Table 1 (M2).

3.2 Within-country Estimation of Country-level Effects

The reason for this difference in the effect size is that M2 in Table 1 and M5 in Table 2 are purely cross-sectional estimates; they are the multivariate equivalents of the relationship from the scatter plot in Fig. 2. The estimates from Model M3 in Table 2, which allows country-level variables to vary over time, are identified by two sources of variation: between- and within-country variation and the resulting coefficient is a weighted average of the relationships (Bell et al. 2018; Bell and Jones 2015). Using a variant of Mundlak's (1978) formulation, Fairbrother (2014) demonstrates how comparative longitudinal survey data can be modelled to decompose the total effect into its within- and between-country components. Using the notation from above (but excluding, for ease of presentation, all country-level variables but one), the model can be written like this:

$$
y_{jti} = \beta_0 + \sum_{t=1}^{T} \delta_t D_t + \beta_1 x_{1jti} + \dots + \beta_k x_{kjti} + \gamma^{\text{BE}} \overline{z}_j + \gamma^{\text{WE}} (z_{jt} - \overline{z}_j)
$$
$$
+ u_j + u_{jt} + e_{jti}
$$

The variable $\overline{z}_j$ is the country-level mean of z_{jt} across years; it exhibits only between-country variation, and accordingly γ^{BE} is the between-country effect. The term $(z_{jt} - \overline{z}_j)$ describes the variation of z around the country-specific mean and captures within-country variation; its country-specific mean is zero. The correlation between $(z_{jt} - \overline{z}_j)$ and u_j must be zero. This may sound like a technical detail but it is of utmost importance: The coefficient γ^{WE} provides the within-country effect of z and it cannot suffer from omitted variable bias due to any *time-constant* country-level characteristic because any such unobserved variable would be part of u_j. Thus, the within-effect has an advantage over the between-effect, and the nondecomposed
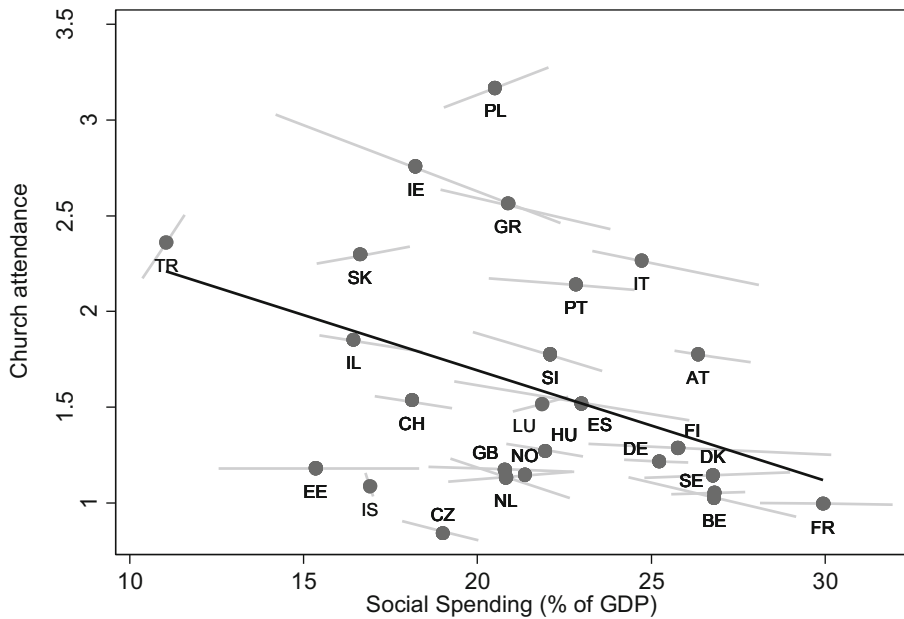
Fig. 3 Bivariate relationships of church attendance and social spending within and between countries. Notes: Based on ESS data 2002–2014 (compare Table 4). The *black line* represents the between-country association, while the *grey lines* represent the associations within the single countries. Compare Fig. 2 for the definition of country codes.

total effect, in terms of the necessary assumptions for unbiasedness (Fairbrother 2014).[19]

In less technical words, the standard interpretation of regression estimates is that "*y* increases by β units if *x* increases by one unit". This interpretation clearly implies the notion of change over time. We expect that for any given unit we will observe a change in *y* because of a change in *x*. For such a statement to be validly drawn from between-country differences, we assume that the countries in our sample differ only in their observed variables but not in any unobserved (correlated) characteristic. As Gelman (2005, p. 461) puts it "it is a big leap to interpret differences between countries as a potential effect of a change within a country" (Fairbrother 2014, p. 3). It may be a better test to directly investigate change over time within countries.

Figure 3 presents bivariate relationships of social spending and church attendance between countries, as in Fig. 2, but also within each country. The black line represents the between-country association and the grey lines show the within-country relationships. The graph reveals that there are indeed negative relationships between social spending and church attendance in many countries (e. g. Ireland, Italy, Spain,

---

[19] The idea to identify an effect solely by within-unit variation and thereby to control for any time-constant unobserved variables originates from the analyses of panel data. Readers who want to get a detailed understanding of this may want to read this literature: Allison (2009); Andress et al. (2013, Chap. 4); Bell and Jones (2015).

Table 3 Decomposing country-level effects into within and between components

Variable	M6 b/p		M7 b/p	
Individual-level variables				
Urban vs. rural	0.0940	***	0.0827	***
	0.0000	–	0.0000	–
Education (in years)	–0.0131	***	–0.0131	***
	0.0000	–	0.0000	–
Subjective income	–0.0273	***	–0.0130	***
	0.0000	–	0.0006	–
Male (ref = female)	–0.2494	***	–0.2334	***
	0.0000	–	0.0000	–
Age	0.0116	***	0.0108	***
	0.0000	–	0.0000	–
Country-level variables				
Social Spending (% of GDP) [BE]	–0.0382	–	–0.0139	–
	0.1888	–	0.5394	–
Social Spending (% of GDP) [WE]	–0.0112	–	–0.0081	–
	0.0826	–	0.2338	–
GDP/capita [BE]	–0.0000	–	–0.0000	–
	0.2512	–	0.6884	–
GDP/capita [WE]	–0.0000	–	–0.0000	–
	0.8811	–	0.9078	–
Average urban vs. rural [BE]	0.2796	–	–0.1784	–
	0.5031	–	0.5785	–
Average urban vs. rural [WE]	–0.0485	–	–0.0509	–
	0.5726	–	0.5111	–
Average education [BE]	–0.1109	–	–0.1532	*
	0.1668	–	0.0085	–
Average education [WE]	–0.0139	–	–0.0021	–
	0.5379	–	0.9231	–
Year FEs				
2004	–0.0056	–	–0.0226	–
	0.8401	–	0.3807	–
2006	–0.0428	–	–0.0414	–
	0.2819	–	0.2641	–
2008	–0.0537	–	–0.0514	–
	0.2999	–	0.2794	–
2010	–0.0725	–	–0.0759	–
	0.2109	–	0.1514	–
2012	–0.0899	–	–0.0919	–
	0.1802	–	0.1348	–
2014	–0.0746	–	–0.0815	–
	0.3419	–	0.2566	–

Table 3 (Continued)

| Variable	M6 b/p		M7 b/p	
Constant	3.0431	*	3.8998	***
	0.0124	–	0.0000	–
Variance Components				
Country	0.3270	***	0.1663	***
Country-year	0.0062	***	0.0044	***
Individual	1.9680	***	1.9903	***
Statistics				
N (country)	26		23	
N (country-years)	149		128	
n (individuals)	277,505		239,881	

See text for explanation of M6 and M7

$*p < 0.05$, $**p < 0.01$, $***p < 0.01$ (two-sided tests). All models are estimated via Restricted Maximum Likelihood. Models based on ESS data 2002–2014 (compare Table 4). Model M7 excludes Ireland, Poland and Denmark

GDP gross domestic product, *FE* Fixed Effects, *WE* Within effect, *BE* Between effect

Portugal, Greece, Slovenia) but there are also countries with a positive association (e. g. Norway, Slovakia, Turkey, Luxembourg, Poland) and countries with no apparent relationship (e. g. Estonia, France, Sweden, Great Britain). This casts doubt about the unbiasedness of the cross-sectional analyses presented in Table 1 (M2) and Table 2 (M3). Te Grotenhuis et al. (2015, p. 650) show a similar graph and find a very similar picture. In the example by Te Grotenhuis et al. (2015), it is obvious from the graphical inspection that the vast majority of countries does not show a negative relationship, while in our example one may find an—on average—negative relationship among the countries.

Decomposing country-level effects into their within and between components yields the results presented in Model M6 (Table 3). Within and between-country effects are not identical for any of the four country-level variables, indicating that Model M3 (Table 2), which presented a weighted average of within and between effects, was misleading (see Fairbrother 2014 for a detailed discussion). In all instances, the between effect is much larger than the within effect, indicating that cross-sectional models will often provide overestimated effects due to omitted variable bias. Regarding the effect of social spending, the between effect is –0.038, resembling the effect estimated in the purely cross-sectional Model M2, while the within effect is only –0.011. This coefficient is not significant at the 5%-level but it is close. If the hypothesis is tested one-sided, for which there is a good reason to do because the hypothesis is directed, one could conclude that there is a negative effect of social spending on church attendance; albeit much smaller than a cross-sectional model suggests. So, for now one may conclude that the results of the cross-sectional analyses by Immerzeel and Tubergen (2013), who also tested one-sided, can be replicated by a within-country estimator (but also see the further discussion in the next section).
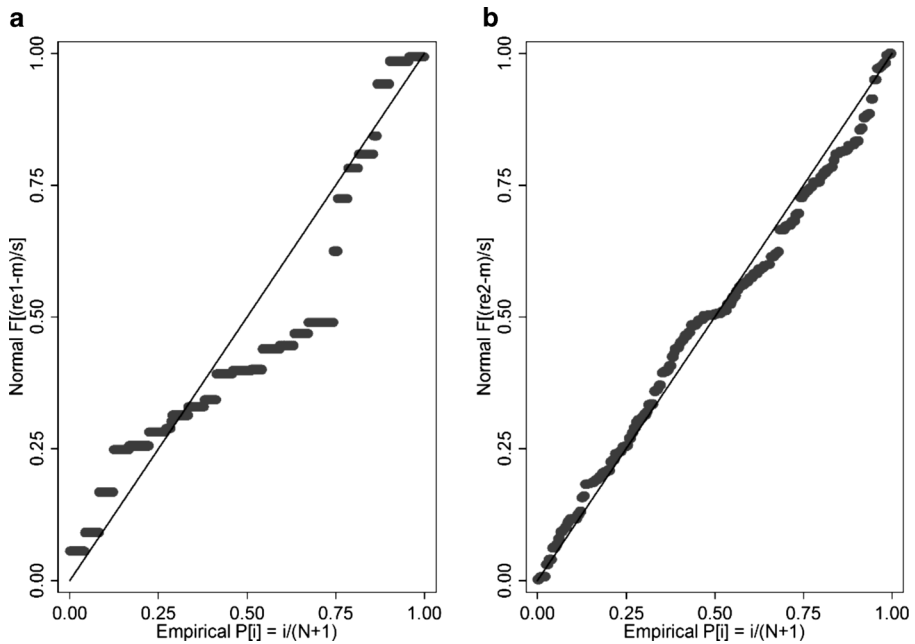
Fig. 4 *P–P* plot of (**a**) country- and (**b**) country-year-level residuals from Model M6 (Table 3). Notes: Based on ESS data 2002–2014 (compare Table 4)

To summarize, by pooling multiple waves of comparative survey data the statistical power can be increased and the option to test hypotheses via within-specifications emerges. From a causal identification standpoint this should be superior to between country estimates which are more prone to omitted variable bias. This is obviously not possible if the variables of interest do not change over time. Similarly, using this technique becomes less useful if the variables of interest change only marginally. In that case, the available variance to identify the effect is small and the estimates will be imprecise. Clearly, the general statistical power and the feasibility of the within-country estimator increases with the number of available survey waves.

4 Diagnostics

Before concluding this article, we briefly discuss diagnostics for multilevel models—specifically, diagnostics for influential cases. There are many statistical tests and graphical inspections that can be used to check for violations of some of the assumptions implicit to the model. Hox (2010, p. 23 ff) provides a very good overview of such tests, and Snijders and Berkhof (2008) discuss many issues in greater and more technical detail. This paper does not have space to discuss regression diagnostics in detail, but that should not be taken as a sign they are not important and

useful. It is valuable to investigate regression diagnostics, particularly through graphical inspection of the residuals at each level. For example, Fig. 4 shows so-called *P–P* plots of the residuals from Model M6 at the country and at the country-year level. These plots allow for a basic visual test of the normality assumption: Perfectly normally distributed residuals form a straight diagonal line. As expected, residuals at the country level (u_j)—with just 26 observations—are not normally distributed, while residuals at the country-year level (u_{jt}), with 139 observations, are close to a normal distribution. In principle, violations of the normality assumption can result in biased SEs and require some caution with respect to statistical inference, though simulations reported by Bell et al. (2019) suggest that such biases are in practice quite modest.

Cook's Distance (Cook's D, for short) is a measure describing the influence of single observations on all estimated coefficients (Cook 1977). In the context of multilevel models, it can be applied to the random and the fixed part separately (Snijders and Berkhof 2008, p. 157 ff.):

$$C_j^F = \frac{1}{r} \left(\widehat{\beta} - \widehat{\beta}_{(-j)} \right)' \widehat{S}_{F(-j)}^{-1} \left(\widehat{\beta} - \widehat{\beta}_{(-j)} \right), \text{ for the fixed part and}$$

$$C_j^R = \frac{1}{p} \left(\widehat{\eta} - \widehat{\eta}_{(-j)} \right)' \widehat{S}_{R(-j)}^{-1} \left(\widehat{\eta} - \widehat{\eta}_{(-j)} \right), \text{ for the random part;}$$

where $\widehat{\beta}$ and $\widehat{\eta}$ are *vectors* of parameter estimates from the fixed and the random part, respectively, and $\widehat{\beta}_{(-j)}$ and $\widehat{\eta}_{(-j)}$ are the same estimates when country j is left out from the sample. Finally, $\widehat{S}_{F(-j)}$ and $\widehat{S}_{R(-j)}$ are the estimated covariance matrices of the fixed and random part and r and p are the numbers of parameters estimated in the fixed and random part. Cook's D can be interpreted as the standardized average squared difference in parameter estimates with and without country j (Van der Meer et al. 2010, p. 175). The total Cook's D measure for the model equals the weighted average of Cook's D for the random and the fixed part:

$$C_j = \frac{1}{r + p} \left(r C_j^F + p C_j^R \right).$$

Since hypotheses are most often about the fixed part of the model, researchers may want to examine the single components rather than the total measure. And with very few countries it is entirely possible that every country will appear to be influential.

That of course depends on the definition of "too influential". Belsley et al. (1980, p. 28) propose the cut-off value $4/n$ for Cook's D. Table 5 (appendix) presents Cook's D for the fixed part of Model M6 and 19 out of 26 countries are deemed too influential if we follow the proposal of Belsley et al. (1980).[20] Obviously, this is not very helpful because the exclusion of 19 countries is not an option. Nevertheless, the Cook's D measure indicates which of the countries has the strongest influence

---

[20] For two-level models, Stata users can use the mlt ado-package to calculate Cook's D and DFBETAs (Möhring and Schmidt 2013). In the online appendix we provide a syntax for three-level models which is very general and can be easily adapted to researchers' own applications.

on the sum of all (fixed) parameter estimates. In the example data the by far most influential countries are Ireland (Cook's D = 5.12) and Israel (Cook's D = 4.84, where the next highest-ranked country has a value of about 2.6). Looking at Fig. 3, one may wonder why Israel (IL) appears as influential, given its position in the scatter plot; recall though that Cook's D is based on the *sum of all parameter estimates* from a multivariate model, not on bivariate relationships.

Nevertheless, to decide how robust the conclusion about the social spending effect is, Cook's D may not be the best measure. After all, Israel does not appear to be a suspicious case in Fig. 3, neither regarding the between- nor the within-country relationship. DFBETA is a measure that describes the influence of a single unit on a selected coefficient (Belsley et al. 1980, p. 13) and can be applied in the context of multilevel models as well (Van der Meer et al. 2010, p. 175):

$$\text{DFBETA}_{zj} = \frac{\widehat{\beta}_z - \widehat{\beta}_{z(-j)}}{\text{SE}\left(\widehat{\beta}_{z(-j)}\right)},$$

where $\widehat{\beta}_z - \widehat{\beta}_{z(-j)}$ is the difference between the effects of variable z with and without country j.[21] This difference is divided by the SE of the effect in the model without country j. DFBETAs can be understood as the standardized difference between the coefficients with and without unit j. For DFBETAs, Belsley et al. (1980, p. 28) propose the cut-off value $2/\sqrt{n}$. Table 5 in the appendix presents DFBETAs for the within effect of social spending and identifies three influential cases: Denmark, Ireland and Poland, with Ireland having a strong negative impact on the estimates (DFBETA = −1.85) and Denmark and Poland having positive influences (0.54 and 0.76, respectively). Again, not all of these countries seem suspicious from inspecting Fig. 3. Given the country-specific relationships presented in Fig. 3, it is no wonder that Ireland has a strong negative impact and that Poland has a positive effect on the estimates but Denmark seems to be a rather inconspicuous case. This is precisely the reason why graphical inspections of bivariate relationships alone are not sufficient to identify influential cases (Van der Meer 2010, p. 175).

The blind exclusion of countries, because they exceed some cut-off value in an outlier statistic, is not a useful strategy; but paying attention to these cases certainly is. One may argue that these outliers are valuable candidates for case studies and/or hint at the need for better theories. Since the space in this paper is limited, we simply present estimates without these three influential cases (Ireland, Poland and Denmark) in Model M7 (Table 3). Focusing only on the effect of interest, Model M7 yields a smaller within effect of social spending than Model M6, and the *p*-value for the coefficient on social spending has increased substantially.

---

[21] DFBETAs can of course also be calculated for individual-level variables (x) but in the context of multilevel modeling its application to country-level variables (z) is typically of interest.

5 Conclusions

We will not attempt to settle the debate between Immerzeel and Tubergen's (2013) argument that social spending has a negative effect on church attendance and the objection by Te Grotenhuis et al. (2015) that this result does not stand up when tested longitudinally. Instead, the purpose of this exercise has been to demonstrate how sensitive results from multilevel models with comparative survey data can be to various decisions taken during the research process, and to suggest useful ways of thinking about that sensitivity.

This paper has addressed a selection of issues, but there are others it has ignored. The research example in this paper used a linear multilevel model, and while all issues are also relevant for nonlinear models, the nonlinear case presents some additional challenges (see Bryan and Jenkins 2016). We also did not address in great detail the estimation of cross-level interactions and random slopes, both of which are important topics (see Bell et al. 2019; Elff et al. 2016; Giesselmann and Schmidt-Catran in press). Finally, we also did not address any issues of model building. For the example analysis, we simply took the model from Te Grotenhuis et al. (2015). Particularly where degrees of freedom are limited, researchers need to choose what variables to include very carefully, on both theoretical and empirical grounds.

Comparative survey data are characterized by a small number of higher-level units (countries) which are not random samples. This presents researchers with several challenges, including questions about whether inferential statistics are useful at all, what the appropriate estimation method is, and whether estimates are sensitive to single countries. There is also a risk of omitted variable bias, or the inability to include a full complement of variables and/or random slopes. While inference about country-level effects must be viewed as conditional on the observed sample, inferential statistics are, from our view, still useful in the context of multilevel models fitted to comparative survey data. With small samples at the country level, researchers would do well to test the robustness of their findings to the choice of different estimation methods. While statistical power at the country level is typically scarce, the contrary is true for the individual level, where observation numbers are typically very large. Particularly at this level, researchers should always consider the practical size of the effects in addition to their levels of significance.

The issue of omitted variables can—to some extent—be addressed by employing within-country estimators, though this requires observing sufficient change over time in the variable of interest and to have a decent number of waves that can be pooled. Thus, not every research question can be tested with these methods. Obviously, such an estimator does only control for time-constant omitted variables but it can still suffer from omitted variables if these too vary over time.

Appendix

Table 4 Sample sizes of example data—European Social Survey (ESS) rounds 1 to 7

Country	Year							Total
	2002	2004	2006	2008	2010	2012	2014	
AT	2138	2154	2288	–	–	–	1778	8358
BE	1725	1742	1781	1747	1662	1863	1759	12,279
CH	2002	2111	1787	1785	1473	1477	1508	12,143
CZ	1224	2525	–	1934	2279	1737	1943	11,642
DE	2849	2723	2803	2688	2986	2920	3001	19,970
DK	1461	1451	1443	1584	1557	1621	1487	10,604
EE	–	1970	1471	1596	1784	2358	2016	11,195
ES	1532	1589	1743	2481	1834	1838	1849	12,866
FI	1972	1997	1880	2181	1852	2169	2064	14,115
FR	–	–	1965	2039	1714	1952	1895	9565
GB	2007	1862	2342	2300	2352	2222	2221	15,306
GR	2515	2388	–	2039	2667	–	–	9609
HU	1672	1471	1481	1511	1548	1937	1645	11,265
IE	1930	2195	1586	1743	2514	2582	2316	14,866
IL	2289	–	–	2264	2020	2353	2460	11,386
IS	–	551	–	–	–	720	–	1271
IT	1160	1484	–	–	–	856	–	3500
LU	1403	1567	–	–	–	–	–	2970
NL	2309	1847	1870	1748	1784	1825	1881	13,264
NO	2027	1754	1744	1541	1540	1614	1432	11,652
PL	2071	1695	1687	1595	1703	1840	1563	12,154
PT	1456	1989	2072	2237	2003	2066	1239	13,062
SE	1979	1924	1903	1811	1488	1812	1761	12,678
SI	1487	1369	1433	1234	1352	1233	1210	9318
SK	–	1373	1649	1725	1790	1784	–	8321
TR	–	1805	–	2341	–	–	–	4146
Total	39,208	43,536	34,928	42,124	39,902	40,779	37,028	277,505

Source: ESS data 2002–2014, obtained from the cumulative data wizard on 20th September 2017

Note: For definition of the country codes compare Fig. 2.

Table 5 Cook's D of fixed part and DFBETAs of within-effect of social spending from Model M6

Country	Cook's D	DFBETAs
AT	0.2983*	0.0025
BE	0.4002*	–0.1199
CH	0.0952	–0.0782
CZ	0.0786	0.0474
DE	0.5876*	–0.2020
DK	0.2882*	0.5436*
EE	0.6936*	0.1051
ES	1.7909*	–0.0312
FI	0.6411*	0.2576
FR	0.3415*	0.1015
GB	1.2517*	0.0654
GR	0.6241*	–0.0859
HU	0.1458	0.1163
IE	5.1165*	–1.8460*
IL	4.8394*	–0.3309
IS	0.0062	–0.0238
IT	0.0525	0.0137
LU	0.0316	0.1129
NL	0.1516	–0.0785
NO	0.2899*	0.0436
PL	0.4649*	0.7625*
PT	2.6162*	0.0936
SE	0.4284*	–0.2783
SI	0.6803*	0.1920
SK	0.5251*	0.0055
TR	2.1563*	0.2088

* cut-off value (Cook's D = 0.1538, DFBETAs = 0.3922) exceeded
Note: For definition of the country codes compare Fig. 2.

References

Allison, Paul D. 2009. *Fixed effects regression models*. Thousand Oaks: SAGE.

Andress, Hans-Jürgen, Katrin Golsch and Alexander W. Schmidt. 2013. *Applied panel data analysis for economic and social surveys*. Springer: Berlin, Heidelberg.

Arceneaux, Kevin, and Gregory A. Huber. 2007. What to do (and not do) with multicollinearity in state politics research. *State Politics & Policy Quarterly* 7:81–101.

Babones, Salvatore J. 2013. *Methods for quantitative macro-comparative research*. London: Sage.

Barr, Dale J., Roger Levy, Christoph Scheepers and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Beck, Nathaniel. 2001. Time-series–cross-section data: What have we learned in the past few years? *Annual Review of Political Science* 4:271–293.

Bell, Andrew J., and Kelvyn Jones. 2015. Explaining fixed effects: Random effects modeling of time-series cross-sectional and data panel data. *Political Science Research and Methods* 3:133–153.

Bell, Andrew J., Kelvyn Jones and Malcolm Fairbrother. 2018. Understanding and misunderstanding Group mean centering: A commentary on Kelley et al.'s dangerous practice. *Quality & Quantity* 52:2031–2046.

Bell, Andrew, Malcolm Fairbrother and Kelvyn Jones. 2019. Fixed and random effects models: Making an informed choice. *Quality & Quantity* 53:1051–1074.

Bell, Bethany A., Grant B. Morgan, Jason A. Schoeneberger, Jeffrey D. Kromrey and John M. Ferron. 2014. How low can you go? *Methodology* 10:1–11.

Belsley, David A., Edwin Kuh and Roy E. Welsch. 1980. *Regression diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley: New York.

Bowers, Jake, and Katherine W. Drake. 2005. EDA for HLM: Visualization when probabilistic inference fails. *Political Analysis* 13:301–326.

Brännström, Lars. 2008. Making their mark: The effects of neighbourhood and upper secondary school on educational achievement. *European Sociological Review* 24:463–478.

Browne, William J., and David Draper. 2000. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics* 15:391–420.

Bryan, Mark L., and Stephen P. Jenkins. 2016. Multilevel modelling of country effects: A cautionary tale. *European Sociological Review* 32:3–22.

Cook, R. Dennis. 1977. Detection of influential observation in linear regression. *Technometrics* 19:15–18.

Draper, David. 2008. Bayesian multilevel analysis and MCMC. In *Handbook of multilevel analysis*, eds. Jan De Leeuw and Erik Meijer, 77–139. Springer: New York.

Ebbinghaus, Bernhard. 2005. When less is more: selection problems in large-N and small-N cross-national comparisons. *International Sociology* 20:133–152.

Elff, Martin, Jan P. Heisig, Merlin Schaeffer and Susumu Shikano. 2016. No need to turn Bayesian in multilevel analysis with few clusters: How frequentist methods provide unbiased estimates and accurate inference. *Open Science Framework* (osf.io/fkn3u).

Elwert, Felix. 2013. Graphical causal models. In *Handbook of causal analysis for social research,* ed. Morgan, Stephen L., 245–273. Springer Netherlands: Dordrecht.

Enders, Craig K., and Davood Tofighi. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods* 12:121–138.

European Social Survey Cumulative File, ESS 1–7 (2016). Data file edition 1.0. NSD – Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Fairbrother, Malcolm. 2013. Rich people, poor people, and environmental concern: Evidence across nations and time. *European Sociological Review* 29:910–922.

Fairbrother, Malcolm. 2014. Two multilevel modeling techniques for analyzing comparative longitudinal survey datasets. *Political Science Research and Methods* 2:119–140.

Fairbrother, Malcolm. 2016. Trust and public support for environmental protection in diverse national contexts. *Sociological Science* 3: 359–382.

Finseraas, H. 2012. Poverty, ethnic minorities among the poor, and preferences for redistribution in European regions. *Journal of European Social Policy* 22:164–180.

Fontaine, Johnny R.J. 2015. Traditional and multilevel approaches in cross-cultural research: An integration of methodological frameworks. In *Multilevel analysis of individuals and cultures,* eds. Van de Vijver, Fons J.R., Dianne A. Van Hemert and Ype H. Poortinga, 65–93. New York: Psychology Press.

Gelman, Andrew. 2005. Two-stage regression and multilevel modeling: A commentary. *Political Analysis* 13:459–61.

Gelman, Andrew. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:72–91.

Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Giesselmann, Marco and Alexander W. Schmidt-Catran. 2018. Getting the within estimator of cross-level interactions in multilevel models with pooled cross-sections: Why country dummies (sometimes) do not do the job. *Sociological Methodology*. https://doi.org/10.1177/0081175018809150

Giger, Nathalie. 2012. Is social policy retrenchment unpopular? How welfare reforms affect government popularity. *European Sociological Review* 28:691–700.

Goerres, Achim, Markus B. Siewert, and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In Cross-national comparative research – analytical strategies, results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Goldthorpe, John H. 1997. Current issues in comparative macrosociology: A debate on methodological issues. *Comparative Social Research* 16:1–26.

Te Grotenhuis, Manfred, Marijn Scholte, Nan Dirk de Graaf and Ben Pelzer. 2015. The between and within effects of social security on church attendance in Europe 1980–1998: The danger of testing hypotheses cross-nationally. *European Sociological Review* 31:643–654.

Heisig, Jan P., Merlin Schaeffer and Johannes Giesecke. 2017. The costs of simplicity: Why multilevel models may benefit from accounting for cross-cluster differences in the effects of controls. *American Sociological Review* 82:796–827.

Heisig, Jan Paul, and Merlin Schaeffer. 2019. Why You Should Always Include a Random Slope for the Lower-Level Variable Involved in a Cross-Level Interaction. *European Sociological Review* 35:258–279.

Hox, Joop J. 2010. *Multilevel analysis: Techniques and applications, 2nd Edition*. New York: Routledge.

Immerzeel, Tim, and Frank Van Tubergen. 2013. Religion as reassurance? Testing the insecurity theory in 26 European countries. *European Sociological Review* 29:359–372.

Jackman, Simon D. 2009. Bayesian analysis for the social sciences. New York: John Wiley.

Jaeger, Mads M. 2013. The effect of macroeconomic and social conditions on the demand for redistribution: A pseudo panel approach. *Journal of European Social Policy* 23:149–163.

Kim, Jee-Seon, and Edward W. Frees. 2006. Omitted variables in multilevel models. *Psychometrika* 71:659–690.

Maas, Cora J. M., and Joop J. Hox. 2005. Sufficient sample sizes for multilevel modeling. *methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 1:86–92.

Van der Meer, Tom, Manfred Te Grotenhuis and Ben Pelzer. 2010. Influential cases in multilevel modeling: A methodological comment. *American Sociological Review* 75:173–178.

Mewes, Jan. 2014. Gen (d) eralized trust: women, work, and trust in strangers. *European Sociological Review* 30:373–386.

Möhring, Katja, and Alexander W. Schmidt. 2013. MLT: Stata module to provide multilevel tools (Statistical Software Components S457577). Boston, MA: Boston College Department of Economics.

Mundlak, Yair. 1978. Pooling of time-series and cross-section data. *Econometrica* 46:69–85.

Patterson, H. Desmond, and Robin Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. *Multilevel and longitudinal modeling using stata, 3rd Edition*. College Station, TX: Stata Press.

Schmidt-Catran, Alexander W. 2016. Economic inequality and demand for redistribution: Cross-sectional and longitudinal evidence. *Socio-Economic Review* 14:119–140.

Schmidt-Catran, Alexander W., and Malcom Fairbrother. 2016. The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review* 32:23–38.

Semyonov, Moshe, Rebeca Raijman and Anastasia Gorodzeisky. 2006. The rise of anti-foreigner sentiment in European societies, 1988–2000. *American Sociological Review* 71:426–449.

Snijders, Tom A.B., and Johannes Berkhof. 2008. Diagnostic checks for multilevel models. In *Handbook of multilevel analysis*, eds. Jan De Leeuw and Erik Meijer, 457–514. New York: Springer.

Snijders, Tom A.B., and Roel J. Bosker. 2012. *Multilevel analysis: An introduction to basic and advanced multilevel modelling, 2nd Edition*. London: Sage.

Stadelmann-Steffen, Isabelle. 2012. Education policy and educational inequality—evidence from the Swiss laboratory. *European Sociological Review* 28:379–393.

Stegmueller, Daniel. 2013. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* 57:748–761.

Stegmueller, Daniel, Peer Scheepers, Sigrid Roßteutscher and Eelke de Jong. 2012. Support for redistribution in western Europe: Assessing the role of religion. *European Sociological Review* 28:482–497.

Sutton, John R. 2012. Imprisonment and opportunity structures: A Bayesian hierarchical analysis. *European Sociological Review* 28:12–27.

Van Erp, Sara, Joris Mulder and Daniel L. Oberski. 2017. Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods* 23:363–388.

Wilkes, Rima, Neil Guppy and Lily Farris. 2007. Right-wing parties and anti-foreigner sentiment in Europe. *American Sociological Review* 72:831–840.

Wooldridge, Jeffrey M. 2013. *Introductory econometrics: A modern approach*. Mason, OH: South-West Cengage Learning.

Wulfgramm, M. 2014. Life satisfaction effects of unemployment in Europe: The moderating influence of labour market policy. *Journal of European Social Policy* 24:258–272.

Alexander W. Schmidt-Catran 1983, Dr. rer.pol., Professor for sociology with a focus on quantitative methods, Goethe-University Frankfurt. Areas of research: welfare states, migration, public opinion, statistical methods. Publications: The random effects in multilevel models: Getting them wrong and getting them right. European Sociological Review 32, 2016 (together with M. Fairbrother); Applied panel data analysis for economic and social surveys. Berlin, Heidelberg 2013 (together with H.-J. Andreß and K. Golsch); Immigration and welfare support in Germany. American Sociological Review 81, 2016 (together with D. Spies).

Malcolm Fairbrother 1975, PhD, Professor of Sociology, Umeå University, and researcher, Institute for Futures Studies, Stockholm. Areas of research: political sociology, environment/nature-society relations, globalization, trust, social science methodology. Publications: When will people pay to pollute? Environmental taxes, political trust, and experimental evidence from Britain. British Journal of Political Science 46, 2017; Economists, capitalists, and the making of globalization: North American free trade in comparative-historical perspective. American Journal of Sociology 119, 2014; Two multilevel modeling techniques for analyzing comparative longitudinal survey datasets. Political Science Research and Methods 47, 2014.

Hans-Jürgen Andreß 1952, Dr. phil., Professor for Empirical Social and Economic Research. Areas of research: social research, statistics and multivariate methods, social inequality, labor market research, social and family policy. Publications: Is material deprivation decreasing in Germany? A trend analysis using PASS data from 2006 to 2013. Journal for Labour Market Research 52, 2018; The economic consequences of divorce in Germany: What has changed since the turn of the millennium? Comparative Population Studies 40, 2015 (with M. Bröckel); Applied panel data analysis for economic and social surveys. Berlin 2013 (together with K. Golsch).

Kölner Zeitschrift
für Soziologie und
Sozialpsychologie

Check for updates

ANHANDLUNGEN

Immigrants' Labor Market Outcomes: Contributions from Multilevel Studies

Romana Careja

Abstract The article reviews 18 studies published between 2000 and 2017 which use multilevel techniques and focus on immigrants' labor market outcomes. It is important to understand why immigrants appear to be at particular risk of experiencing poor economic outcomes, be it participation in the labor market, labor market attainment or income. The multilevel analyses reveal that structural conditions in both origin and destination countries are significantly shaping immigrants' labor market outcomes, while other factors (such as social capital) have less clear effects. The review also shows that multilevel studies, although they reveal substantial insights, are in want of a better theoretical explanation of the contextual effects, especially when the contexts of interest are at the country level. It concludes that, by incorporating interaction effects and relying on longitudinal data, future multilevel studies have the potential to better account for immigrants' economic outcomes.

Keywords Immigrants' labor market outcomes · Multilevel analysis · Subnational units as macrolevel contexts · Structuralist · Institutionalist · Psychological and ideological explanations

Erfolg von Zuwanderern auf dem Arbeitsmarkt: Erkenntnisse aus Mehrebenenstudien

Zusammenfassung In dem Artikel werden 18 zwischen 2000 und 2017 veröffentlichte Studien untersucht, in denen Mehrebenenanalysen verwendet werden, um den Arbeitsmarkterfolg von Zuwanderern zu ermitteln. Es ist wichtig zu verstehen, warum Einwanderer einem besonderen Risiko auf dem Arbeitsmarkt ausgesetzt sind,

R. Careja (✉)
Department of Political Science and Public Administration, University of Southern Denmark
Campusvej 55, 5230 Odense M, Denmark
E-Mail: rca@sam.sdu.dk

🙋 Springer

eher benachteiligt zu sein; sei es bei der Erwerbsbeteiligung, beim beruflichen Vorankommen oder beim Einkommen. Die Mehrebenenanalysen zeigen, dass die strukturellen Bedingungen in den Herkunfts- und Zielländern den Arbeitsmarkterfolg der Einwanderer maßgeblich beeinflussen, während andere Faktoren (wie etwa ihr Sozialkapital) weniger klare Auswirkungen haben. Die Übersicht zeigt auch, dass für Mehrebenenanalysen, obwohl sie wesentliche Einsichten offenbaren, bei der theoretischen Erklärung von Kontexteffekten noch Nachholbedarf besteht, insbesondere wenn die interessenden Kontexteffekte auf Länderebene angesiedelt sind. Der Artikel kommt zu dem Schluss, dass zukünftige Mehrebenenanalysen das Potenzial haben, durch die Einbeziehung von Interaktionseffekten und die Verwendung von Längsschnittdaten die ökonomischen Erfolge von Einwanderern besser belegen zu können.

Schlüsselwörter Arbeitsmarkterfolg von Zuwanderern · Mehrebenenanalyse · Subnationale Einheiten als Makrokontexte · Strukturalistisch · Institutionalistisch · Psychologische und ideologische Erklärungen

1 Introduction

Immigrants' incorporation into the labor market is unanimously seen as an essential step in the integration process, ultimately leading to immigrants becoming productive and self-reliant members of the host societies. Notwithstanding this uplifting view, many studies find significant gaps between Western Europeans and immigrants on indicators such as labor market participation and attainment, welfare dependency, or poverty risks, and argue that they are gauging the limited integration of immigrants. Other studies argue that such bird's-eye views based on averages are not entirely correct, as there is greater diversity among immigrants: For example, the children of immigrants may encounter similar problems when entering the labor market as their parents did, but there are differences when it comes to attainment: Children of immigrants from other European countries outperform their parents, but remain somewhat behind the majority groups, while other groups face disadvantages in many contexts (for example, Turks in many European countries) (see the detailed review by Heath et al. 2008).

The sustained interest in immigrants' labor market outcomes is visible not only in the number of studies, but also in the large palette of methods: from qualitative and qualitative comparative analyses (QCA) studies (Agudelo-Suarez et al. 2009; Hooijer and Picot 2015) and panel/longitudinal studies (Büchel and Frick 2005; Dustmann and van Soest 2002; Kanas et al. 2012; Bratsberg et al. 2014), to various regression-based models (Kalter and Granato 2002; Adsera and Chiswick 2007; Munoz-Comet 2016; Kesler 2015), and event-history models (Kalter and Kogan 2006).

Many of these studies have emphasized the role of individual factors such as education, family background, accumulated experiences, years in host countries, contacts with natives, age, gender, and ethnicity (Adsera and Chiswick 2007; Evans 1989; Sanders and Nee 1996; Perreira et al. 2007; Flap et al. 2000; Waldinger et al.

2006), but contextual factors have come into the limelight more recently. Theoretical arguments stating that individual outcomes depend on the context in which individuals live are not new (Blalock 1984; Huber 1991). Socialization theories, for example, emphasize the part played by family values and conditions, the roles of peers in the neighborhood, or the impact of the work environment. The context can also have an effect through structures: As neighborhood effects theories show (see for a discussion Ainsworth 2002; Sampson et al. 2002), living in a poor versus a richer neighborhood can shape the way in which one dresses and speaks, which may have consequences on how a person presents himself or herself or comes over in a job interview situation. Similarly, institutional theories argue that different institutional settings open or close opportunities for individuals, with far-reaching implications for a person's success on the labor market (or lack thereof). In this context, multilevel modeling techniques emerged as an effective instrument to gauge the complex effects of context and individual factors on individual economic outcomes such as risk and duration of unemployment, employment attainment and status, or experience of self-employment.

Multilevel analysis has several advantages over the other techniques which can account for the effect of contextual factors. It allows the researcher to simultaneously examine the effects of context and individual-level variables, while accounting for the non-independence of observation within units. It also accounts for both cross-individual and cross-context variation, as well as for the contribution of individual-level and context-level variables to this variation (see Schmidt-Catran et al. 2019). In addition to its technical advantages, the basic multilevel model is flexible, thereby enabling researchers to test many theories using a variety of conceptualizations and measurements of outcomes of interest. It can accommodate different conceptualizations of "levels": Individuals can be seen as nested in groups, in neighborhoods, regions or countries, but they can also be considered to be nested in time units. The multilevel model can include two, three or more levels, and can be applied to different outcomes (binary, continuous, multiple-category or survival analysis) (Snijders and Bosker 2012; Hox 2010).

This review proceeds as follows. It begins with a brief summary of the findings of previous studies concerning immigrants' labor market outcomes, followed by an overview of the studies which rely on multilevel modeling to examine these outcomes. The following sections examine these studies in greater detail, focusing on explanatory frameworks (Sect. 4) and findings (Sect. 5). The review concludes with a discussion of the use of multilevel modeling in the reviewed studies, and of the contribution it brings, as well as its potential.

2 Labor Market Outcomes of Immigrants in Europe

Unlike the U.S. and Canada, Western European countries did not become a destination for immigration on a large scale until after the Second World War, when the guest-worker programs were enacted to help governments and employers deal with the massive labor force shortages. Workers from Yugoslavia, Morocco, Turkey and former colonies arrived in the 1950s and 1960s. The Oil Crisis in the early

1970s led to a freeze in labor recruitment from abroad, but foreigners arriving via family reunification and asylum constituted a pool of low-skilled labor throughout the 1970s and 1980s. The 1990s once more witnessed a reversal, as Western European governments came to see economic migration as beneficial, especially if it was channeled toward those economic sectors that need it the most.

During the 1990s, evidence started to accumulate that gaps between immigrants and natives existed along important dimensions such as educational attainment (Vermeulen and Perlman 2000; Grönqvist 2006; Van Zanten 1997), poverty risks (Blume et al. 2007; Galloway and Aaberge 2005), access to, and attainment in, the labor market (Berthoud 2000; Fieldhouse 1999; Herzog-Punzenberger 2003; Kogan 2006; Simon 2003; Wolbers and Driessen 1996). In spite of the fact that the concern with these lasting gaps has spilled over from the country level to the EU level, where it spurred efforts to discourage discrimination (by adopting the Racial Equality Directive (2000/43/EC) and the Employment Equality Directive (2000/78/EC)) and to systematically monitor immigrants' economic outcomes, Eurostat data indicate that, 16 years after the EU Equality Directives were put in place, there remain visible differences on average between natives and immigrants on the most important labor market indicators. Among immigrants, those coming from non-EU countries have the worst outcomes. For example, they have higher rates of long-term unemployment (i.e. lasting longer than 12 months), and the highest rates of youth unemployment. These two indicators are of particular concern because they indicate the long-term risk that these groups face, due to the cumulative nature of the effects of unemployment, which span across many areas, from deskilling, to loss of social rights, to psychological and health effects (Brand 2015).

3 Data Sources and Context Units

3.1 Descriptives of Studies Included in the Review

Following the 1999 publication of Snijders and Bosker's comprehensive textbook on multilevel modeling, a multitude of textbooks have been published (Hox 2010; Gelman and Hill 2007; Raudenbush and Bryk 2002; Skrondal and Rabe-Hesketh 2010; Goldstein 2011; Demidenko 2004; de Leeuw and Meijer 2008), which have deepened knowledge and broadened the field, providing insights for further modeling. As a consequence, the multilevel technique became an essential tool in the social scientist's toolbox during the 2000s, and came to be applied to an ever-increasing research domain. The year 2000 was therefore selected as the starting point in the literature search for this review. In order to retrieve articles focusing on immigrants' labor market outcomes, the search included the terms "multilevel analysis/model", "immigrants", "economic outcomes", "employment", "unemployment", "poverty", "welfare dependency", and "wages", in various combinations. As multilevel models are sometimes referred to as "random effects models" and "hierarchical models", these terms have also been used. The search spanned economic, political science, and sociological publications.

The literature on immigrants' economic outcomes is vast (and rapidly expanding) (see Heath, Rothon and Kilpi's 2004 review of the economic outcomes of the second generation, or Midtbøen's 2015 review of ethnic penalties on labor markets, in addition to the abovementioned publications), with a remarkable number of country studies and small-N comparisons. Studies where the country names are replaced by variables are growing steadily in number, this being a direct consequence of the increasing availability of comparable country-level data, but many of them could not be included in this review as they use other analytical strategies (see however Sect. 6 for a brief commentary on their findings). The literature search produced a comparatively small number of studies which explicitly use multilevel modeling techniques (see Table 1).

Table 1 summarizes the literature identified for this study along two dimensions: type of context units used, and the labor market outcome of interest. As it can be seen, in seven studies the outcome of interest is immigrants' job/labor market status, in six studies it is immigrants' unemployment, and in five studies it is the take-up of self-employment by immigrants. As to the macrolevel considered, eight studies follow the cross-country approach, and nine focus on subnational units within one country. While in both cases the researchers can evaluate the impact of contextual characteristics, the latter strategy has certain advantages: It enables the researcher to keep factors such as institutions, laws, or country-specific skills that immigrants are expected to acquire under control, while zooming in on social, structural, and economic characteristics of *immediate* contexts expected to affect individuals' economic outcomes such as neighborhood, the locality or the region where they live (for an elaboration of the role of different context levels, see Kloosterman and Rath 2001). As to design, all the studies reviewed here rely on one-year cross-sections or pool several survey waves (see column "Method & design" in Table 2).

3.2 Data and Units of Analysis

The multilevel modeling technique builds on the recognition that context units/ groups are not independent from each other (in fact, a basic assumption is that they are "extracted" from a larger population) and it is able to account for the non-independence of observation within context units. The correct estimation of the effects requires that the units at both levels can be identified. If it can be argued that individuals are clearly identifiable (biological bodies with sets of psychosocioeconomic characteristics), other units pose greater challenges. At the macro level, individuals can be nested in units with fixed borders, such as countries or subnational administrative units, but they can also be nested in less clearly defined units.

As Tables 1 and 2 (column "secondary level units") show, several studies focus on subnational units such as neighborhoods and enclaves as the relevant contexts shaping immigrants' labor market outcomes. The fact that these units do not have clear borders presents researchers with opportunities and challenges: Although researchers can define the context unit that makes most sense for the intended research, finding the appropriate size may cause problems. A major concern is the heterogeneity of neighborhood/enclave characteristics, specifically the relation of the within- to the between-variation: When the neighborhood/enclave size is large, variation within

Table 1 Summary of multilevel studies focusing on immigrants' labor market outcomes (author's own concept)

| | | Labor Market Outcomes | | | | |
		Self-employment (5[a])	Labor force participation (2[a])	Unemployment (5[a])	Job/occupation status (7[a])	Wages/income (2[a])
One country (9[a])	Neighborhoods	Wang (2009) (*US*) Zolin et al. (2015) (*Australia*)	–	–	Johnston et al. (2010) (*New England and Wales*)	–
	Localities	Ohlsson et al. (2012) (*Sweden*) Andersson et al. (2013) (*Sweden*)	–	–	Kogan and Kalter (2006) (*Germany*)	–
	Regions	–	–	Scheller (2017) (*Germany*)	Scheller (2017) (*Germany*) De Jong and Steinmetz (2004) (*US*)	Kim (2018) (*Korea*)
Several countries (8[a])	Western countries	Van Tubergen (2005)	Van Tubergen et al. (2004) Kislev (2017)	Van Tubergen et al. (2004) Fleischmann and Dronkers (2010) Kogan (2006, 2007) Kislev (2017)	Pichler (2011) Kogan (2007)	Kislev (2017)
	Western and non-Western countries	–	–	–	Spörlein and Van Tubergen (2014)	–

[a]In brackets—number of studies in the respective category

Table 2 Description of multilevel studies focusing on immigrants' labor market outcomes (author's own concept)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Single-country studies								
Ohlsson et al. (2012) Sweden	Micro: Register data Sweden 2007; Macro: Statistics Sweden	Multilevel (ML) logistic; cross-classified; Cross-section	Mixed embeddedness	Self-employment	81 local labor market regions; 44 countries of birth	ICC reported, varied for men (14%) & women (15%)	Direct: Local context and country of birth have limited impact	Individual factors explain most of the variation
Andersson et al. (2013) Sweden	Micro: Register-based LISA dataset Sweden 2007; Macro: Statistics Sweden	ML cross-classified; Cross-section	Mixed embeddedness	Self-employment	8 geographical areas of origin; 7 groups by year of arrival; 87 local labor markets	ICC reported, varies for men (18.2%) and women (17.8%)	Direct: Local labor markets; ethnic origin and time of arrival	Self-employment likelihood decreases with the % of non-Europeans; Limited effect of local labor market and ethnic origin
Wang (2009) US	Micro: Public Use Microdata Sample (PUMS) US 2000; Macro: various datasets	ML; Cross-section	Mixed embeddedness (not explicit)	Self-employment	241 metropolitan areas (MA)	Not reported	Direct: Various MA characteristics; residential segregation; local economic conditions	MA effects on propensity for self-employment differ by gender and ethnic group; Significant effects of macroeconomic conditions, business structure, ethnic composition, residential segregation

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Zolin et al. (2015) Australia	Micro: Own Survey of Chinese entrepreneurs 2010	ML, Cross-section	Social capital theory	Entrepreneur's success	Co-ethnic enclave (unclear number)	Not provided	Direct: Location of business; Cross-level: Location of business interacts with non-co-ethnic social capital	Non-co-ethnic social capital has a stronger positive relationship with business growth for businesses outside the enclave
Kim (2018) Korea	Micro: National Survey of Multicultural Families 2009; Macro: aggregated from survey	ML, Cross-section	Social capital	Monthly earnings (log)	46 regional clusters	ICC not provided, significant between-cluster variance mentioned	Direct: discrimination, social activities, bridging and bonding social capital; Cross-level: Interactions with weak ties	Finding a job through informal contacts (weak ties) is connected with low income, and the effect is enhanced if the person lives in a co-ethnic community characterized by strong co-ethnic social capital

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
De Jong and Steinmetz (2004) US	Micro: U.S. Bureau of the Census Current Population Survey (1995–1997); 1994 General Social Survey; 1990 Census	Multilevel, hierarchical, non-linear	Context of reception (structural and public attitudes)	Occupational attainment (lower skilled and higher skilled)	38 geographic units (metropolitan areas MAs)	ICC and variance not reported	Direct: Natives' receptivity attitudes; various social, economic and demographic of MAs	Favorable structural conditions and some receptivity attitudes—associated with occupational attainment of women and men in service and technical jobs
	Macro: Aggregated from individual data	Cross-section					Cross-level: Interaction between education and natives' receptivity attitudes	Receptivity attitudes—no direct impact on immigrants' attainment in managerial and professional positions; some structural characteristics associated with women's attainment in managerial and professional positions

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Kesler and Hout (2010) US	Micro: 5% Integrated Public Use Microdata Sample (IPUMS) 2000	ML (cross-classified)	Mixed embeddedness	Hourly wage (log)	490 ethnic communities	ICC not reported; variances reported in Annex	Direct: A community's level of and success with self-employment impacts immigrants' wages	In general, higher self-employment in one's ethnic community is associated with lower wages among immigrant employees
	Macro: aggregated from micro data	Cross-section					Cross-level: Community's experience with self-employment interacted with human capital and gender	However, when co-ethnic entrepreneurs are exceptionally successful financially, their presence has a positive effect on the wages of employees in the community. Immigrants with high human capital may be buffered against the negative effect of the entrepreneurial community

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Johnston et al. (2010) UK	Micro: Controlled Access Microdata Sample 2001 Macro: Deprivation and Isolation indexes constructed by authors	ML logistic Cross-section	Contexts of reception	School to work transition	Neighborhoods	ICC not reported	Direct: Spatial concentration in areas of deprivation and/or segregated areas tends to exacerbate immigrants' disadvantages in the labor market relative to educational attainment	All ethno-religious groups suffer what appear to be "ethnic penalties" in the labor market, but there are differences across ethnic groups. These penalties are also associated with place of residence: Those living in relatively deprived and/or ethnically segregated areas are more likely to be overqualified
Kogan and Kalter (2006) Austria	Micro: Labor Force Survey 1995, 1997, 1999, 2000 Polled cross-section	Random effects GLS (generalized least square)	Economic competition and queuing model (ethnic group resources)	Occupational status (male immigrants)	35 localities	ICC not reported; Variance within localities is reported	Direct: Size of immigrant ethnic group in locality Indirect: The ethnic group membership interacted with the size of the immigrant population in a community of residence	Relative size of ethnic group in locality direct positive effect on occupational status of immigrants. Interactions: The larger the proportion of Turks in localities, the better the occupational status of both Turks and Yugoslavs. Larger proportions of Yugoslavs has a negative effect on status

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Cross-country studies								
Van Tubergen et al. (2004)	Micro: the International File of Immigration Surveys 1980–2001	ML Logit	Human capital theory Discrimination theory	Labor market participation; employment	18 Western destination countries; 187 origin groups, 984 combinations of origins and destinations	ICC not reported; Variance components reported	Direct: 2 origin, 3 destination and 6 community effects	Significant composition and context effects
	Macro: various, ILO, OECD, WB	Pooled						
Van Tubergen (2005)	Micro: the International File of Immigration Surveys 1980–2001	ML (cross-classified)	Sociological rational-choice	Self-employment	17 Western destination countries; 180 origin groups; 850 settings	ICC reported: country of origin (10%), country of destination (6%); Variance components reported	Direct: 4 origin, 3 destination, and 9 setting effects	Odds of self-employment among immigrants increase if immigrants come from non-Christian origin countries, if natives experience unemployment, and if immigrant co-ethnic groups are small, highly-educated, and have a longer settlement history
	Macro: various, ILO, OECD, WB	Pooled						

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Kogan (2006)	Micro: European Labor Force Surveys 1992–2000 Macro: ILO; OECD	ML logistic Pooled	Contexts of reception (institutions)	Unemployment (only for third-country nationals, TCNs)	15 EU States	ICC reported: depending on model, between 41 and 55%; Variance components reported	Direct: structural and institutional characteristics (structure of economy, welfare regime, EPL) Cross-level interactions: (with being a TCN)	In countries with a larger share of unskilled/low-skilled jobs, TCNs are less disadvantaged compared to natives. Male TCNs are more likely to find jobs (than women) when the host country's labor market is more flexible, and when it has liberal welfare regimes
Kogan (2007)	Micro: European Labor Force Surveys 1992–2000 Macro: ILO; OECD	ML Pooled	Contexts of reception (institutions)	Unemployment; occupational status (male third-country nationals TNCs; EU immigrants)	14 EU countries	ICC reported: approx. 50%; Variance components reported	Direct: structural and institutional characteristics (structure of economy, welfare regime, EPL) Cross-level interactions: (with being a TCN)	The labor market outcomes of immigrants are influenced by the institutional characteristics of the receiving societies In countries with a larger share of unskilled/low-skilled jobs and flexible labor markets, particularly recent immigrants are less disadvantaged compared to natives. Scandinavian welfare regimes are a barrier to labor market entry

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Fleischmann and Dronkers (2010)	Micro: European Social Survey 2004 Macro: various sources: UNDP, CIA, Freedom House, etc.	ML cross-classified Cross-section	Not explicitly stated—follows van Tubergen et al. 2004	Unemployment status	13 Western European countries of destination; No. of countries of origin not reported	Not reported	Direct: 15 destination, 10 origin effects	Only mean unemployment in destination country, and coming from a EU-15 Member State, have a significant effect on the likelihood of being unemployed
Pichler (2011)	Micro: European Social Survey 2002–2008 Macro: OECD, WB, MIPEX	ML logistic regression Pooled	Contexts of reception (structural characteristics)	Occupational status	28 European countries	Random parts/ variation reported	Direct: GDP growth, welfare regime, EPL, integration policies (MIPEX)	Occupational attainment varies across countries. Policies unlikely to have an effect. Welfare regimes matter, with worse off odds in Southern European countries (women even worse than men). The positive effect of liberal welfare regimes (observed by Kogan (2006, 2007) not corroborated

Table 2 (Continued)

Study—author & Nr. of countries	Data sources & time covered	Method & design	Theory	Dependent variable	Secondary level units	Explanatory power of upper level (ICC)	Hypothesized effects	Main result(s)
Spörlein and van Tubergen (2014)	Micro: Integrated Public Use Microdata Series International 1987–2007; Macro: Various, ILO, OECD	ML linear regression, cross-classified; Pooled	Context and composition effects-Human capital theory; Discrimination theory	Occupational status	47 Western and non-Western destination countries; 201 origin groups, 1661 immigrant communities	ICC reported: varies between Western/non-Western and origin/destination/community Variance components reported	Direct: 2 destination, 2 origin and 6 communities	Occupational status depends on economic conditions in the countries of origin, geographical distance between host and origin countries, co-ethnic group size, religious and socio-economic distance between immigrant groups and natives
Kislev (2017)	Micro: ESS 2002–2012; Macro: Various, MIPEX, WB, OECD	ML cross-classified; Pooled	Reception context	Labor force participation, unemployment, household income	Country of origin (exact number not provided), 13 destination countries	ICC not reported; Variance components reported	Direct: Social acceptance of immigrants, pro-immigrant policies; inequality and human development	The context effects vary with the labor market outcome. Supporting social environment (positive attitudes of natives) positively affects the likelihood of finding a job. Immigrants' disadvantage in terms of household income and labor force participation depends on characteristics of both origin and host countries

ILO International Labor Organization, *OECD* Organisation for Economic Cooperation and Development, *WB* The World Bank, *ESS* European Social Survey, *MIPEX* Migration Integration Policy Index, *EPL* employment protection legislation

units may be high, but variation between units may be limited. The combination of large within variation and low between variation increases the difficulty encountered when identifying the context effects. If the neighborhoods/enclaves are defined to limit internal heterogeneity, the low within unit variation may create problems for the estimation of the random terms (O'Campo 2003).

Multilevel models' ability to take into account the effects of factors situated at different levels capture close-to-reality combinations of determinants of immigrants' labor market outcomes. However, this increased ability to capture the multilayered reality comes with its own price, as the models become more data demanding and require an increasingly complex estimation of their assumptions. Sample size and power calculations for multilevel hypothesis-testing are more demanding in terms of the number of contexts and the number of units per context than in the case of traditional regression methods. Due to large cross-national survey projects and various repositories of country-level data, as the column "Data sources and time" (Table 2) indicates, it has become possible to approach the minimum recommended number of 20 context-level units, albeit the lower limit of 20 is still hotly debated (Stegmueller 2013; Maas and Hox 2005; Meuleman and Billiet 2009; Hox et al. 2012; Bryan and Jenkins 2016). This problem is of course much smaller when subnational units are used as contexts, although in this latter case other problems might intervene—such as the availability of finely grained data (for example at the level of neighborhood/enclave, or some other lower-level administrative units).

As Table 2 (column "Data sources and time") illustrates, macrolevel data are retrieved from large repositories offered by the International Labor Organization (ILO), the Organisation for Economic Cooperation and Development (OECD), the World Bank (WB) and Eurostat, or national statistical institutes. Microlevel data are derived from various national surveys, from cross-national surveys such as the European Social Survey (ESS), or from datasets compiled from various domestic sources by the researchers themselves (van Tubergen et al. 2004; Spörlein and van Tubergen 2014; Van Tubergen 2005). The availability of high-quality data sources makes it possible not only to increase the number of macrolevel units, but also to combine them in novel ways, which enables original hypotheses to be tested (van Tubergen et al. 2004; Spörlein and van Tubergen 2014; Kogan and Kalter 2006). The column "Secondary level units" in Table 2 illustrates how researchers take advantage of the new data sources that are available in order to include diverse contexts: The subnational studies cover anywhere between 35 localities (Kogan and Kalter 2006) to 87 local labor markets (Andersson et al. 2013) to 490 ethnic communities (Kesler and Hout 2010). Depending on the modeling technique, some cross-national studies rely solely on destination countries (for example 28 countries in Pichler 2011), while others include destination, origin and combinations thereof (settings). The latter is an approach developed by van Tubergen and colleagues, which enables them to include a large number of context-level units: For example, van Tubergen et al. (2004) use 18 Western destination countries, 187 origin countries and 984 settings.

The issue of data availability partly affects the choice of context units. As country-level data is regularly collected and available free of charge, an ever-increasing number of countries can be included in analyses, as van Tubergen and colleagues have shown. Diverse situations occur at subnational level. For example, the availabil-

ity of administrative register data encourages scholars to opt for using administrative units (see Andersson et al. (2013) and Ohlsson et al. (2012) who use locality-level data in Sweden, or Scheller (2017) who uses planning regions in Germany). Such administrative data have the advantage that the researcher can include *all* existing subnational units. In contrast, where administrative data are absent or difficult to come by, researchers devise alternative strategies, which result in fewer than real-existing upper level units included in the analysis. For example, Kesler and Hout (2010) develop a set of criteria to identify the metropolitan areas and ethnic communities, while de Jong and Steinmetz's (2004) combination of several data sources provides them with unit-level data for 38 subnational geographical units. In these latter cases, an additional question emerges which is often encountered, but not readily acknowledged (Kesler and Hout 2010; De Jong and Steinmetz 2004; Kim 2018; Zolin et al. 2015): Do the units included in these analyses constitute nonbiased samples of all the potential subnational units? Given the fact that multilevel analysis assumes that the upper-level units are a random sample from a larger population of units, the reliability of the results is affected if this condition is not met. This problem, of course, does not occur if the upper level includes all the units from the population of subnational units (for example Andersson et al. 2013).

4 Theoretical Explanations

According to Parsons (2007), social scientists rely on explanations which, by and large, can be divided into four categories: institutional, structural, psychological, and ideational. *Institutional* explanations consider the individuals' position within institutions, namely man-made organizations and rules. *Structuralist* explanations focus on the role of structures such as geography, wealth distribution, or power differentials. Both institutions and structures are seen as exogenous to the phenomenon to be explained, and constrain individual actions and behaviors in specific ways. *Psychological* explanations focus on the role of cognition and affect as main determinants of individual outcomes. Similarly, *ideational* explanations also emphasize the role of cognitive and affective elements, but see them as "created by certain historical groups of people" (Parsons 2007, p. 12). Albeit attracting some criticism (see the "critical friendly" commentary by Daigneault and Béland (2015)), Parsons' typology in essence provides an elegant heuristic to map a field in which little effort had previously been made to systematize the theoretical approaches accounting for the effects of contexts on immigrants' labor market outcomes.

4.1 Institutionalist Explanations

Studies which focus on institutions generally distinguish between host- and home-country institutions. Most attention is devoted to understanding the role of the former, reflecting the general concern with the (economic) outcomes of immigrants in the host countries. However, as illustrated below, some authors argue that the home countries' institutions are also likely to have an impact on their emigrants' future. The *host-country* institutions most often expected to influence the economic out-

comes of immigrants are policies, as man-made regulatory frameworks which shape opportunities. Although a coherent theory is yet to be formulated, several studies elaborate on the likely effects of several policies, based on their liberal or restrictive features. For example, Pichler (2011) argues that policies that facilitate immigrants' settlement are likely to have positive effects, but their impact may be indirect, via other policies or processes. Other scholars pay attention to policies which enhance employment protection, institute constraining immigration regimes or provide too generous welfare benefits and expect to observe negative effects on immigrants' labor market outcomes, due to the fact that they directly affect the ability and incentives of immigrants to obtain jobs (Kogan 2006; Pichler 2011; Büchel and Frick 2005). Other host-country institutions which are likely to shape immigrant outcomes are the color of government or the type of welfare state (Kogan 2006; Pichler 2011; van Tubergen et al. 2004) because they are expected to influence the chances that immigrants are offered, as well as their incentives to work. The arguments concerning *home-country* institutions build on the idea that the institutional make-up of the home countries may influence the emigration decision of certain categories, which will experience differentiated economic success (van Tubergen et al. 2004; Fleischmann and Dronkers 2010; Spörlein and van Tubergen 2014). All these studies use countries as units of analysis at the upper level because they conceptualize institutions as country characteristics, and disregard possible differences at the subnational level that some institutions might exhibit, for example policy implementation may be different at regional/local levels.

4.2 Structuralist Explanations

Many scholars who take a structuralist perspective see individuals as being first and foremost embedded in local structures, which shape their opportunities. This approach has the advantage that it accounts for within-country diversity. For example, long-term complex processes have led to the formation of pockets of deprivation or segregation, and living in such areas may result in disadvantages (fewer opportunities, lower job quality) (Johnston et al. 2010). Similarly, local economies may vary with respect to their profile (agricultural production vs. highly competitive niche sectors), or vary with respect to their unemployment levels. Such differences not only have an effect on the type of immigrants settling in these localities (low vs. highly skilled), but also on their long-term employment prospects and career attainment (Ohlsson et al. 2012; Andersson et al. 2013; Wang 2009; Kesler and Hout 2010). Studies in this vein are located at subnational level, and use neighborhoods, localities, or regions as the upper level of analysis.

Other studies however focus on country-level structural conditions. For example, geographical distances or income inequality between home and host countries, size of immigrant communities or overall economic conditions, constitute structural characteristics which shape the migration flows and influence the life chances of immigrants in the host countries (van Tubergen et al. 2004; van Tubergen 2005; Fleischmann and Dronkers 2010; Spörlein and van Tubergen 2014; Kogan and Kalter 2006; Scheller 2017). Arguably, some of these structural conditions are not uniform within countries (income inequalities, economic conditions, immigrant communi-

ties), and the simplifying assumptions embedded in using countries as the upper level of analysis can be criticized. While this is a valid criticism, the choice is often based on pragmatic considerations spurred by the absence of detailed subnational data.

4.3 Psychological Explanations

By definition, psychological explanations are formulated at the individual level. In studies which explore the effect of contexts on individual economic outcomes, psychological factors are part of the mechanisms through which context characteristics are translated into individual outcomes such as unemployment, or entrepreneurial behavior. For example, the human capital (i.e. skills and cognitive capacities), the ability to make rational decisions or the awareness of discrimination risks, can be conceptualized as filters through which immigrants negotiate the positive or negative characteristics of their living environments (Van Tubergen et al. 2004, p. 707). A detailed discussion of these effects and the supporting mechanisms is presented in Sect. 4.5.

There are several studies which build on psychological explanations at contextual level. They are centered around "social capital". Although social capital is often seen as an individual characteristic, it can be conceptualized as a context characteristic if it is seen as a characteristic of networks. It must however be noted that social capital-based explanations cannot be easily placed into one single category. In fact, they are more correctly seen as being simultaneously structururalist and psychological explanations. Social capital has a structural component, as it depends on the structure of one's ties (i.e. the network per se), but it is more than that, as it depends on the quality of the ties, with key elements such as trust, affect, or cognitive connections. As in the studies taking this approach the authors emphasize the cognitive and affective links that are established between individuals in the network, I have decided to place social capital-based explanation in the psychological category.

As to the effects of social capital, the jury is still out, as theoretical expectations can be formulated in positive and negative directions. The presence of co-nationals in the host countries has long been proven to help immigrants by reducing the information and settling costs (see for an overview Gold 2005). The literature on economic outcomes postulates that co-ethnic networks have an effect well beyond the initial settling stage: They can provide not only information about employment opportunities, but also employment opportunities, and even demand services and become a potential market for entrepreneurial immigrants. Although bonding social capital, that is social capital emerging in homogeneous networks such as co-ethnic ones, has obvious advantages, some authors argue that there is a risk that co-ethnic networks provide low-quality jobs, thereby undermining individuals' development potential, and maintain that bridging social capital, that is social capital emerging in networks encompassing ethnically diverse groups, is more important (see for a discussion Kanas et al. 2009), as it offers access to more and more diversified resources which may have positive effects on immigrants' labor market outcomes. This line of argumentation sees immigrants embedded in *local* networks, and therefore the

studies pursuing this path often choose neighborhoods and enclaves as the upper-level units of analysis (Scheller 2017; Kogan and Kalter 2006; Zolin et al. 2015).

4.4 Ideational Explanations

Ideational explanations focus on the role that ideas espoused by groups of people have in shaping social actions. Public opinion has increasingly become the focus of a number of studies which explore the effects of ideas on individual outcomes such as policy preferences or policy responses. Several scholars have also adopted this approach in relation to immigrants' economic outcomes. The underlying argument is that public opinion creates an ideational sphere on its own, capable of producing effects at individual level because it influences the opportunities that immigrants are offered or denied. Public opinion can be treated as a country-level characteristic (see Kislev 2017), or as a subnational characteristic, if natives' attitudes toward immigrants are conceptualized as varying within countries, due for example to past patterns of localized interactions (or a lack thereof) (De Jong and Steinmetz 2004). These different conceptualizations result in the use of countries or regions and localities as macrolevel units in multilevel models.

4.5 Combinations of Institutionalist, Structuralist, and Psychological Explanations

Although studies often focus on one explanatory factor, and control for the effects of others, there are studies that reflect an endeavor to capture complex social phenomena by combining explanatory frameworks. One such approach is *mixed embeddedness,* which has been developed to explain immigrants' propensity to become entrepreneurs. Its starting point is a criticism of the literature that sees individual characteristics as *the* determinant of immigrant entrepreneurship (see for example Hammarstedt 2001; Constant and Zimmermann 2006; for a discussion see also Kloosterman and Rath 2001; Kloosterman 2010). In contrast, the position adopted by the proponents of mixed embeddedness reflects the view that the emergence of immigrants' entrepreneurship depends *simultaneously* on individual capital (financial and entrepreneurial spirit and knowledge), on the social capital of the neighborhood (which provides support and a market for the entrepreneur) *and* on the opportunity structure at the local level (local market situation and institutions) (Kloosterman and Rath 2001). Studies in this vein naturally zoom in on local contexts, neighborhoods, enclaves, or local labor markets (Ohlsson et al. 2012; Andersson et al. 2013; Wang 2009; Kesler and Hout 2010).

A different way of combining explanations is exemplified by the work of van Tubergen and colleagues (van Tubergen 2005; van Tubergen et al. 2004; Spörlein and van Tubergen 2014) as well as Fleischmann and Dronkers (2010), who argue that context characteristics affect a person's migration decisions, which are ultimately reflected in the labor market outcomes (for example, countries with labor market regimes which reward individual investment in skill acquisition are likely to attract immigrants who perceive high returns for their skill set). These scholars have theorized that immigrants make decisions under the simultaneous effects of host-countries' contexts (*destination* effects), of countries of origins (*origin* effects),

and under the influence of destination and origin characteristics, combined in what is designated as *community/setting* effects (for example skill transfer opportunities as shaped by laws, or religious and geographical distance). These three types of macrocharacteristics can be seen as directly influencing immigrants' labor market outcomes (i.e. they produce *contextual* effects), or they influence the choices that immigrants make, thereby resulting in an unequal distribution of immigrants across contexts (i.e. they produce *composition* effects).

In this overall framework, rational choice, human capital and discrimination theories (which arguably reflect psychological, i.e. cognitive and affective processes) are used to elaborate on how the composition and contextual effects come to be. For example, on the basis of rational choice, plausible mechanisms can be developed that explain how characteristics of host countries can be conducive to certain economic outcomes. Thus, van Tubergen (2005) argues that some immigrants opt to become entrepreneurs as a rational decision (self-interest) if they live in a context which prevents them from securing salaried employment and/or which opens a market niche for services that they can provide. Or, aware of their own skills, immigrants rationally choose the place of settlement in such a way as to maximize the returns on their skills (see this line of reasoning in Van Tubergen et al. 2004; reflecting economic arguments of migration as elaborated by Todaro and Maruszko 1987; Borjas 1990). However, when people flee from a country that has a bad political environment, the decision to migrate is not taken in order to maximize the return on skill investment, and thus refugee groups are more mixed in terms of their human capital. Van Tubergen and colleagues argue that human capital theory can be used to explain the unequal selection of human capital across different macrounits given the characteristics of those units, i.e. to elaborate on the *composition* effects exerted by contexts. In its turn, discrimination theory illuminates other mechanisms. For example, a person belonging to a "visible" minority group is more exposed to discrimination, and therefore more likely to have poor labor market outcomes (van Tubergen et al. 2004). Van Tubergen et al. (2004) argue that such mechanisms illustrate direct effects of a context (i.e. one where certain minorities are easily visible) on individual outcomes, i.e. they illustrate *contextual* effects. In these explanations, the focus is placed on the structural, economic, and political characteristics of origin and destination countries, and they are tested with models which use countries at the upper level. The well-known criticism that "country" may be too coarse a unit of analysis applies here too, but by defining *settings* as combinations of origin and destination country characteristics the researchers are able to produce more finely graded macrocharacteristics (van Tubergen et al. 2004; van Tubergen 2005; Spörlein and van Tubergen 2014).

5 Effects of Contexts: Findings

5.1 Overview of the Findings, by Economic Outcomes

As a cursory look at Table 2 indicates, data sources, time periods and context units, as well as the outcomes of interest, vary greatly, making it difficult to assess

to what extent the findings of these studies contradict or complement each other. Several broad statements concerning the role of contextual factors can nevertheless be formulated.

A first way to make sense of the findings is by looking at the studies which include multiple dependent variables. For example, van Tubergen et al. (2004) examined the likelihood of both employment and labor force participation, and found that both outcomes are negatively affected by political suppression in the country of origin, and by disparities between origin and destination countries, while being positively affected by left-wing governments, a relatively large community of co-nationals in destination countries, and a Christian origin. Surprisingly, the mean educational level of co-nationals and the geographical distance between origin and destination countries have negative effects on labor force activity, but exert positive effects on the likelihood of employment. The authors note the pattern, but do not provide an explanation. Kislev (2017) found that a supportive environment in the destination country significantly increases immigrants' chances to find a job, but not labor force participation. Immigration and integration policies do not have an effect on employment, but do affect immigrants' income, leading the author to conclude that policies by themselves can only remedy the challenges immigrants face to a limited extent. Kogan (2007) found that a large low-job sector and a liberal welfare regime diminish the risk of unemployment among third-country nationals (TCNs), i.e. immigrants coming from a non-EU country. The latter factor also increases the likelihood to improve their occupational status. Scheller's (2017) study in Germany found that the communities of co-nationals have no effect on the risk of unemployment, but have a differentiated effect on the job status of members of different ethnic groups: A larger group of co-nationals has no effect for Turks, increases the job status of Italians, and decreases that of Greeks. Regional economic conditions have no effect on either unemployment or on job status, with the exception of the regional unemployment rate, which increases immigrants' unemployment risk. All in all, these studies provide plausible evidence that the same individual and contextual factors do not have the same impact on different labor market outcomes. Unfortunately, due to the differences in data sources and operationalizations, their findings cannot be properly evaluated against each other.

Another way to look at the findings is to group studies which share the same outcome. While the individual-level effects are by and large in the expected directions, there is much greater diversity when it comes to the effects of contextual factors. With regard to *self-employment*, van Tubergen's (2005) cross-country study finds that immigrants take this option either pushed by the (perceived) discrimination on the labor market, or as the result of a lack of employment opportunities. Self-employment is also more likely to be taken up by individuals living in immigrant communities with relatively higher education levels and longer settlement periods. These findings were echoed by some studies conducted at subnational level, while being contradicted by others. For example, Ohlsson et al. (2012) and Andersson et al. (2013) concur that, compared to individual characteristics, the regional structural factors (such as local economic conditions) play only a minor role in explaining the propensity of self-employment among immigrants in Sweden, but Wang (2009) finds significant contextual effects (macro-economic structure, population structure)

in US metropolitan areas. Zolin et al. (2015) find that ethnic businesses' early survival depends on their embeddedness in co-ethnic communities, but that their long-term success is driven by being located outside co-ethnic communities.

With regard to *labor force participation* and the *likelihood of being (un)employed*, cross-country studies found that the likelihood of being employed is not affected by policies, such as a point-based immigration system (van Tubergen et al. 2004), employment, integration and education policies, or by employment protection legislation (Fleischman and Dronkers 2010; Kislev 2017), but is encouraged by factors which, overall, can be considered as describing favorable receiving contexts (structural factors such as the existence of a large low-skilled segment on the job market, high gross domestic product (GDP) per capita, institutional factors such as liberal welfare states (Fleischmann and Dronkers 2010; Kogan 2006, 2007)), or social environmental factors such as the presence of a well-educated community of co-ethnics or public opinion with positive views about immigrants (van Tubergen et al. 2004; Kislev 2017). Somewhat in contrast, Scheller's regional-level study finds that only the regional unemployment rate affects immigrants' likelihood to experience unemployment themselves, while other regional structural factors like presence of co-nationals, regional wealth or population density do not matter (Scheller 2017).

Occupational status is of particular interest because a great deal of literature documents immigrants' high risk of entrapment in low-status jobs, as well as the cross-generational handing down of low occupational status. By and large, it seems not to be affected by integration and immigration policies (Pichler 2011). Kogan (2006, 2007) finds a positive effect of liberal welfare states, but Pichler (2011) does not echo this finding. Some structural factors seem to have an effect (Spörlein and van Tubergen 2014; Pichler 2011), and, if looked at more in-depth as De Jong and Steinmetz (2004) document, their effect is unequal across different job statuses (management vs. low-skilled occupations) and gender. The relevance of structural factors is also documented at subnational level, where in particular the local unemployment rates and structure of the local economies (Kogan and Kalter 2006), and the relative size of co-ethnic groups have been found to have significant effects (albeit the effects may have different directions for different ethnic groups (Scheller 2017)). Last but not least, occupational status is significantly affected by the distance between natives and immigrants: The closer the socio-economic distance between the countries of origin and destination, and the more positive the attitudes adopted by natives, the higher is immigrants' occupational attainment (De Jong and Steinmetz 2004; Spörlein and van Tubergen 2014).

With regard to *income*, Kislev (2017) finds that immigrants' household income is positively affected by pro-immigrant factors in the destination country (education policies which contain prodiversity elements, and a positive public opinion vis-à-vis immigrants), as well as by good starting conditions (emigrating from a country with a high value on the Human Development Index). Surprisingly, he finds that labor market policies which target immigrants have a negative impact on income—and explains the effect as a consequence of the fact that labor market policies for immigrants often channel them toward low-wage jobs. Kim's (2018) community-level analysis finds that co-ethnic social capital affects women and men differently: Women immigrants earn lower incomes if they live in communities characterized

by high levels of bonding social capital (measured via frequency of social activities among co-ethnics). Moreover, living in such communities accentuates the income penalty experienced by women when they obtain employment via informal contacts.

The conclusion that could be drawn from this overview is that immigrants' labor market outcomes are influenced by a variety of factors, and that, considering each category of labor market outcomes separately, different context factors emerge as having significant effects. Do such findings in fact tell coherent stories about the role of contexts in shaping immigrants' labor market outcomes? To answer this question, I turn to consider the findings through the prism of the explanations and theoretical frameworks mentioned above.

5.2 Theoretical Explanations

In several articles at the end of the 1980s and the beginning of the 1990s, Portes and his collaborators argued that immigrants' integration, and indeed their success in the host countries, depended on what they called "contexts of reception" (Portes 1995; Portes and Böröcz 1989)—the totality of host county characteristics that are likely to shape the opportunities that immigrants may have: Institutions that enable or restrict access to labor market or skill recognition, economies that privilege certain skills over others, or even public opinion that is positively (or negatively) oriented toward immigrants. The authors posited that immigrants who encounter handicapped contexts of reception are more likely to have negative integration outcomes (including poor economic outcomes), while those encountering privileged contexts of reception are more likely to be successful.

By and large, the findings of the studies reviewed here tend to support this idea, but the results are by no means as clear cut as Portes and colleagues suggested. One of the reasons may lie in the strategy employed by these studies—namely to focus on certain explanatory factors, while controlling for others, thus not modeling the simultaneous action of macrolevel characteristics (which seems to be at the core of Portes and colleagues' argument).

For example, *mixed embeddedness* emphasizes the role of local structural conditions. However, several studies in this group reached different conclusions: Ohlsson et al. (2012) and Andersson et al. (2013) found that the ethnic context and local economic environment play only a minor role in explaining immigrant entrepreneurship in Sweden, and concluded that individual characteristics such as entrepreneurial spirit are more important explanatory factors. Andersson et al. (2013) also argued that the propensity toward self-employment is lower in areas inhabited by non-European immigrants, which he attributes to the lower purchasing power of the inhabitants. In contrast, US-based studies found strong and significant local context effects (Wang 2009; Kesler and Hout 2010).

Other structural characteristics also show effects in different directions. For example, van Tubergen et al. (2004) found, as expected, that high income inequality ratios between home and host countries is associated with lower odds of employment and of labor market participation of immigrants, but unexpectedly found that immigrants arriving from richer countries (relative to host countries) were more likely to be out of the labor market. Immigrants arriving from countries farther away have higher

odds of being employed, but within this group, women have higher odds to remain outside of the labor market.

Structural characteristics at neighborhood level also have divergent effects. For example, Johnston et al. (2010) find that the higher the neighborhood deprivation, the lower the immigrants' chances are of matching returns to education on the labor market, but the higher the neighborhood segregation, the higher the chances are of receiving returns that match the education level. The authors speculate that the latter finding might be linked to the effects of ethnic social capital—networks which provide individuals with employment in highly segregated areas.

Several institutional country-level factors such as policies indicating economic openness or closure vis-à-vis immigrants, institutional protections for workers that affect employers' inclination to hire immigrants, or immigration policies, have attracted sustained scholarly interest. Büchel and Frick (2005) find that more restrictive labor markets, as well as policies which limit economic opportunities, are associated with worse outcomes for immigrants. Kogan (2006) found no effects of employment protection legislation (EPL), but found fewer labor market disadvantages among immigrant men in countries with liberal welfare states, where mobility and flexibility in relation to the labor market are emphasized, unlike conservative and social-democratic welfare states, which are more rigid. Pichler (2011), however, has not found this effect of the welfare state.

De Jong and Steinmetz (2004) build explicitly on Portes and Böröcz's (1989) distinction between handicapped and advantaged contexts of reception, and conceptualize the latter "natives' receptivity", i.e. natives' positive orientations towards immigrants (public opinion). Their study finds that, after controlling for structural conditions such as local labor market characteristics, natives' attitudes have both a direct and an indirect effect (via education) on the occupational attainment of immigrant workers, but only for those in service and low-level occupations, and not for those in managerial and professional ones.

Studies which tested propositions based on *social capital* arguments have found significant effects on immigrants' economic outcomes. However, there is some disagreement as to their direction. For example, Kogan and Kalter (2006) tested the occupational queuing hypothesis, which recognizes that not all the groups provide the same level of social capital, and found that when several ethnic groups share the same geographical area, the job status of immigrants is influenced by ethnic hierarchies. Zolin et al. (2015) found that ethnic entrepreneurs profit in their first years from the presence of co-ethnic networks (bonding social capital), but their thriving and long-term success depends on their connection to non-co-ethnic communities (bridging social capital). In contrast, Scheller (2017) found no effect of social capital on immigrants' risk of unemployment, and discovered that job status was influenced by local social ties in the case of Greeks and Italians, but not in the case of Turks, while Kim (2018) documented that belonging to a tightly-knit co-ethnic group accentuates the ethnic penalty that immigrant women experience in terms of income.

The several studies which test contextual and composition hypotheses, capturing the effects of home and host countries, test the same theoretical model. However, they use different indicators and datasets, which may explain the differences in their

findings. Although significant effects of macrolevel characteristics are observed, many of them disappear when individual-level characteristics are controlled for. The most extreme example is the study carried out by Fleischmann and Dronkers (2010), which included 18 country-level factors, but only two (mean unemployment rate of natives and being an old EU Member State) remained significant after the introduction of individual characteristics. The studies by van Tubergen and colleagues show more significant macrolevel effects (both composition and contextual) after individual factors are controlled for (roughly 7 out of 10); however the direction of some effects changes across studies. For example, political suppression in the country of origin (which is associated with non-economic migration, and also with trauma which makes people less employable) has the predicted negative effect on employment and labor market activity (van Tubergen et al. 2004), but has non-significant effects on self-employment and occupational status (van Tubergen 2005; Spörlein and van Tubergen 2014). Relative group size (which is a setting effect) has a negative effect on self-employment (van Tubergen 2005) and occupational status (Spörlein and van Tubergen 2014), and a positive effect on labor force activity (van Tubergen et al. 2004).

Summing up, none of the four types of explanations receives incontestable support across the board. Structural factors, in particular related to the labor market, seem to explain better self-employment, participation and attainment, while pro-immigrant attitudes among natives seem to explain better occupational attainment, income, or labor force participation. Somewhat surprisingly, institutional explanations, in particular those including host-country policies, have limited explanatory power when it comes to labor market outcomes. One could argue that this is not entirely surprising, given the fact that most of these factors are measured at country level, i.e. far removed from individuals. However, even when the macro characteristics are measured in individuals' immediate environment, such as structural conditions at local level or the social networks, they do not generate unambiguous results.

Even though they provide valuable insights, these findings seem to suggest that macrolevel effects are less stable or less universal than one might have expected. This state of affairs may arguably be affected by the extent to which these studies tackle some of the issues that are specific to the multilevel explanatory frameworks and modeling techniques. Several of these issues are discussed below.

6 Discussion and Conclusion

The first question that one might ask is how much of the total variation can be attributed to the macrolevel? It is difficult to assess because, as Table 2 shows, only a minority of studies report intra-class correlation coefficients' (ICC) values.[1] Van Tubergen (2005) reports values between 6–10%, while Ohlsson et al. (2012) and

---

[1] The intra-class correlation coefficient (ICC) is the correlation between two observations within the same cluster. The higher the correlation *within* the clusters (i.e. the larger the ICC), the lower is the variability *within* the clusters, and consequently the higher the variability is *between* the clusters. For details and calculation, see Snijders and Bosker (2012).

Andersson et al. (2013) report 14–18%, (with Spörlein and van Tubergen (2014) somewhere in between), which may suggest that the context level does not carry the largest explanatory power. In fact, both Andersson and Olhsson and their colleagues argue that the ICC observed in their respective studies is low, prompting them to conclude that most of the variation is in fact explained by individual factors. In contrast, (only) Kogan (2006, 2007) reports approx. 50% of the variance explained by structural and institutional country-level factors.

A second point to be noted is that, despite increasingly good data available at both individual and contextual level, estimating macrolevel effects remains a challenge due to self-selection bias/endogeneity problems, which are particularly serious in the case of immigrants (for a discussion, see Oberwittler 2007). Needless to say, cross-sectional designs are more susceptible to this endogeneity risk, while longitudinal designs are better suited to mitigate it. These risks are not explicitly dealt with in the studies reviewed here, which may require their findings to be taken with a grain of salt, especially since they all rely on cross-sectional designs, as Table 2 indicates. Van Tubergen and colleagues (van Tubergen et al. 2004; Spörlein and van Tubergen 2014) and Fleischmann and Dronkers (2010) partly account for some of these potential biases by formulating composition effects hypotheses (and specifically theorizing the nonrandom distribution of individuals across contexts), and by including simultaneously destination, origin and community factors (for a discussion, see van Tubergen et al. 2004).

The third point to be noted is the fact that individuals are not confined to one context unit only, but that they have institutional and personal contacts outside the units where they reside, which poses particular challenges to multilevel analyses (see Goerres et al. 2019). Not accounting for this spillover of contacts may result in the overestimation of macrolevel effects. One qualification is needed though: Not all contexts and social interactions are equally likely to be affected. For example, it is more likely to observe contacts across neighborhoods than across countries. What is more, some explanations are more susceptible to this criticism than others. For example, social capital arguments assume that individuals develop networks in the place where they reside/work. Immigrants may however be living in one area, but their networks of support and information may spread across much larger areas, which they can reach via social media and technology.

The problem of overlapping contexts must be addressed at both the theoretical and methodological level. Researchers can rely on previous studies and good knowledge of the phenomenon of interest in order to appropriately theorize the cross-context connections. Methodologically, they have different techniques at their disposal to account for such complex interactions, for example by controlling for contacts outside of the context unit, or by specifically modeling affiliation to different context units with crossed-classified models (Snijders and Bosker 2012). As Table 2 indicates, among the reviewed studies, six implement cross-classified modeling techniques (Ohlsson et al. 2012; Andersson et al. 2013; Kesler and Hout 2010; Van Tubergen 2005; Fleischmann and Dronkers 2010; Spörlein and van Tubergen 2014), while the issue of the nonisolation of individuals in contexts units is not addressed/reflected upon in the remaining ones.

The fourth issue, namely the reciprocal effects between contexts and individual behaviors, is not specific to multilevel studies, but they are particularly vulnerable to it, especially because they aim to identify the effects of macrolevel units. For example, immigrants' protracted entry to the labor market or poor career mobility may prompt the adoption of antidiscrimination measures, which in turn may affect immigrants' labor market outcomes. Such reciprocal effects are more likely to occur in relation to certain outcomes and contexts. Therefore, whenever the likelihood of their occurrence is high, they should be elaborated upon theoretically and modeled statistically. This also presupposes that the datasets must ideally be supportive of such an endeavor. The reviewed studies employ cross-sectional designs, and thus they have difficulties in accounting for such potential reciprocal effects.

Last but not least, the role of theory must be addressed (see Kroneberg 2019). Some of the challenges mentioned above have suggested that a correct account of macrolevel effects is dependent on the way these effects are theorized. The power of contextual analyses depends on comprehensive theories, which outline not only which factors matter, but also how they combine and through which mechanisms they impact individuals' lives (Blalock 1984). Many of the studies reviewed here use existing theories to justify the expectations concerning the effects of selected macrolevel factors (while controlling for individual-level factors). While this is widely accepted as standard practice in hypothesis-testing studies, this strategy may not be enough in the particular case of studies that focus on context effects on individual outcomes, as the mechanisms through which the context *affects* individuals remain obscure or confusing. For example, based on various discrimination arguments and previous studies, van Tubergen et al. (2004) develop no fewer than five propositions which hypothesize both positive and negative effects of origin, destination and community characteristics on immigrants' employment, and the analysis confirmed almost all of them (additionally, six hypotheses based on human capital theory were formulated, with similar positive or negative expected effects). A similar approach is used by Fleischmann and Dronkers (2010) and Spörlein and van Tubergen (2014). While these studies illustrate that a multitude of factors is shaping immigrants' labor market outcomes, they cannot satisfactorily explain the situations when unexpected or divergent effects are observed. For example, van Tubergen et al. (2004) find that the mean educational level of co-nationals and the geographical distance between origin and destination have negative effects on labor force activity, but positive effects on likelihood of employment. Some of these findings support their hypotheses, but the divergent effects of the same factors on different outcomes remain unaccounted for theoretically.

Studies which focus on few context factors (for example on institutions only) are at risk of being criticized for failing to recognize the more complex constellations of factors shaping immigrants' outcomes. Adding breadth (by increasing the number of context variables tested) addresses this criticism, but fails to add depth (by developing a mechanism through which context factors affect individuals). The problem becomes particularly obvious when the context of interest is far removed from individuals, as is the case with country-level characteristics, and when the models include only *direct* macro effects, which is the strategy adopted by the vast majority of the studies reviewed for this article. The macrolevel factors produce

microlevel effects by affecting certain individual characteristics, and therefore they are more likely to operate via some individual-level psychological or ideological mechanisms. The theoretical arguments still fail to elaborate on the paths through which the context factors exercise their impact: Are they moderators or triggers of processes mediated by other individual factors?

As the Western European countries become more ethnically heterogeneous, new questions emerge about the life chances of immigrants and their descendants. In particular, how do the institutions that govern the labor markets and welfare states affect immigrants' entry to and performance on the labor market? Or, how do social and economic structures shape the opportunities that immigrants have to enter the labor market? Multilevel analyses enable researchers to address these questions, as this overview has illustrated. However, the effect of contexts has been researched with other statistical methods as well, such as OLS (ordinary least squares) or logistic regressions (for example, Grönqvist 2006; Musterd et al. 2008; Kanas et al. 2009; Tammaru et al. 2010), Heckmann selection models (Kanas et al. 2012), or two-stage regressions (Lancee 2012). The question which naturally emerges is how do the findings of multilevel studies compare to those of other studies which captured the effects of macrocharacteristics with different modeling techniques? It is difficult to argue that research based on multilevel modeling brings substantial new results. For example, multilevel studies corroborate the absence of significant effects of institutional factors observed by Lancee (2012), and, similarly to studies which use other techniques (see for example, Musterd et al. 2008; Grönqvist 2006; Kanas et al. 2012; Xie and Gough 2011) do not find clear evidence of the direction of the effects ethnic enclaves. They also corroborate findings which suggest that structural factors may have significant effects on immigrants' economic outcomes (Tammaru et al. 2010; Lancee 2012; Hedberg and Tammaru 2013). However, given that multilevel techniques avoid some of the pitfalls of other estimation methods (see also Diez-Roux 2000), it can be argued that the evidence which they provide stands on firmer ground.

By simultaneously estimating direct and moderating effects of contexts as well as individual effects, multilevel modeling techniques enable scholars to better grasp the complex reality of immigrants' labor market outcomes. The studies reviewed here have identified factors which seem to shape immigrants' economic outcomes, such as social capital, or structural factors, as well as factors which do not have a consistent impact, such as institutions. At the same time, several challenges have been identified that have been tackled by the literature on contextual determinants of immigrants' labor market outcome. Some are directly linked to multilevel analysis as a method, while others reflect more general concerns about how we theoretically account for complex reality. As this review has shown, there is still plenty of room for development when it comes to immigrants' labor market outcomes. These concluding paragraphs point out three directions which are likely to yield rewarding research output.

Longitudinal Studies The multilevel studies on immigrants' labor market outcomes reviewed for this study have adopted cross-sectional designs. They thus provide a snapshot of reality, but they cannot illuminate much beyond which factors are

associated with a given labor market outcome. However, if scholars conceptualize outcomes as non-static, or if processes leading to certain outcomes are of interest, or, as discussed earlier, reciprocal effects between labor market outcomes and context conditions must be disentangled, then longitudinal designs are more appropriate. Although more demanding in terms of data and technique, they are also likely to bring us a step closer in terms of understanding what factors shape immigrants' labor market outcomes, and how. For example, research can focus on changes in immigrants' labor market status (either between employment and unemployment, or up the career ladder), which can be determined by economic conditions, as well as by changes in the institutional environment such as adoption and enforcement of anti-discrimination measures or change to a merit-based immigration policy. Moreover, building on the distinction between long-term and short-term risk factors (borrowed from psychological literature, see Oberwittler (2007), citing Wikström and Sampson (2003)), researchers could explore how the former accumulate over time, and combine with the latter, which provide opportunity structures, resulting in desirable or undesirable labor market outcomes for immigrants.

Combinations of Subnational and National Contexts There is already a compelling body of evidence that national-level factors and local-level opportunity structures make a difference. This evidence comes from studies which capture their direct effects, such as all the studies included in this review. But how do local- and national-level factors combine, and what is their combined effect on individual outcomes? The increasing availability of subnational data opens up the possibility of exploring the effects of combined national and subnational contexts. Such analyses can be conducted in one country, but also cross-country. Germany, Netherlands or the Nordic countries have a large quantity of subnational level data available, which already creates a pool of countries whose subnational and national characteristics are sufficiently diverse to produce interesting results.

Theories and Mechanisms Similar to other statistical methods, multilevel analysis will help researchers to describe and quantify patterns in the data. What it will not do is to explain these patterns. An explanation emerges only from the reciprocal relationship between theory elaboration and theory testing. The availability of sophisticated modeling possibilities such as the ones offered by multilevel models allows in principle the testing of complex theoretical propositions. A first step is to move beyond hypothesizing *direct* contextual effects. Theorizing *interactions* between contexts and individual-level factors allows one to gauge much more nuanced relationships and processes, as the test of the queuing hypothesis by Kogan and Kalter (2006) has shown.

However, it is more important—and the challenge remains—to develop substantive theoretical explanations, which not only indicate which factors are likely to make a difference, but also elaborate on the mechanisms through which the contextual factors operate. The existing studies focus on *direct* effects, allowing us to learn a great deal about a web of simultaneous effects. However, especially in studies which use countries as macrocontexts, as the distance between country characteristics and individuals is large, it often remains unclear how the former influence immigrants'

life chances. One avenue of research in this direction could be to theorize and test several alternative mechanisms that link the same outcomes to context-related and individual factors. This would translate into a systematic scholarly endeavor, and would result in a much better understanding of the processes taking place in our societies. Without a systematic approach and an endeavor oriented towards theory-building, multilevel analysis remains a tool which gauges variations across contexts and which provides reliable estimates of context impact on individual outcomes, while we continue to find it difficult to make sense of the patterns it uncovers.

References

Adsera, Alicia, and Barry R. Chiswick. 2007. Are there gender and country of origin differences in immigrant labor market outcomes across European destinations? *Journal of Population Economics* 20:495–526.

Agudelo-Suarez, Andrés, Diana Gil-Gonzaléz, Elena Ronda-Pérez, Victoria Porthé, Gema Paramio-Pérez, Ana M. Garcia and Aitana Garí. 2009. Discrimination, work and health in immigrant populations in Spain. *Social Science and Medicine* 68:1866–1874.

Ainsworth, James W. 2002 Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces* 81:117–152.

Andersson, Lina, Mats Hammarstedt, Shakir Hussain and Ghazi Shukur. 2013. Ethnic origin, local labour markets and self-employment in Sweden: A multilevel approach. *The Annals of Regional Science* 50:885–910

Berthoud, Richard. 2000. Ethnic employment penalties in Britain. *Journal Ethnic and Migration Studies* 26:389–416

Blalock, Hubert M. 1984. Contextual-effects models: Theoretical and methodological issues. *Annual Review of Sociology* 10:353–372.

Blume, Karen, Björn Gustafsson, Peder J. Pedersen and Mette Verner. 2007. At the lower end of the table: Determinants of poverty among immigrants to Denmark and Sweden. *Journal of Ethnic and Migration Studies* 3:373–396.

Borjas, George. 1990. *Friends or strangers: The impact of immigrants on the U.S. economy*. New York: Basic Books.

Brand, Jennie E. 2015. The far-reaching impact of job loss and unemployment. *Annual Review of Sociology* 41:359–375.

Bratsberg, Bernt, Oddbjørn Raaum, Marianne Røed and Pål Schøne. 2014. Immigration wage effects by origin. *The Scandinavian Journal of Economics* 116:356–393.

Bryan, Mark L., and Stephen P. Jenkins. 2016. Multilevel modeling of country effects: A cautionary tale. *European Sociological Review* 32:3–22.

Büchel, Felix, and Joachim R. Frick. 2005. Immigrants' economic performance across Europe: Does immigration policy matter? *Population Research and Policy Review* 24:175–212

Constant, Amelie, and Klaus Zimmermann. 2006. The making of entrepreneurs in Germany: Are native men and immigrants alike? *Small Business Economics* 26:279–300.

Daigneault, Pierre-Marc, and Daniel Béland. 2015. Taking explanation seriously in political science. *Political Studies Review* 13:384–392.

De Jong, Gordon, and Michele Steinmetz. 2004. Receptivity attitudes and the occupational attainment of male and female immigrant workers. *Population Research and Policy Review* 23:91–116.

Demidenko, Eugene. 2004. *Mixed models. Theory and applications*. Hoboken, NJ: Wiley

Diez-Roux, Ana V. 2000. Multilevel analysis in public health research. *Annual Review of Public Health* 21:171–192.

Dustmann, Christian, and Arthur Van Soest. 2002. Language and the earnings of immigrants. *Industrial Labor Relation Review* 55:473–492.

Evans, M. D. R. 1989. Immigrant entrepreneurship: Effects of ethnic market size and isolated labor pool. *American Sociological Review* 54:950–962.

Fieldhouse, Edward A. 1999. Ethnic minority unemployment and spatial mismatch: The case of London. *Urban Studies* 36:1569–96.

Flap, Henk, Adem Kumcu and Bert Bulder. 2000. The social capital of ethnic entrepreneurs and their business success. In *Immigrant business: The economic, political and social environment*, ed. Jan Rath, pp. 142–161. New York: St. Martins Press.

Fleischmann, Fenella, and Jaap Dronkers. 2010. Unemployment among immigrants in European labour markets: An analysis of origin and destination effects. *Employment & Society* 24:337–354.

Galloway, Tarry A., and Rolf Aaberge. 2005. Assimilation effects on poverty among immigrants in Norway. *Journal of Population Economics* 18:691–718.

Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Goerres, Achim, Markus B. Siewert, and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Gold, Steven J. 2005. Migrant networks: A summary and critique of relational approaches to international migration. In *The Blackwell Companion to Social Inequalities*, eds. Mary Romero and Eric Margolis, p. 257. Hoboken, NJ: Blackwell Publishing.

Goldstein, Harvey. 2011. *Multilevel statistical models*. London: Edward Arnold.

Grönqvist, Hans. 2006. Ethnic enclaves and the attainments of immigrant children. *European Sociological Review* 22:369–382.

Hammarstedt, Mats. 2001. Immigrant self-employment in Sweden—its variation and some possible determinants. *Entrepreneurship & Regional Development* 13:147–161.

Heath, Anthony F., Catherine Rothon and Elina Kilpi. 2008. The second generation in Western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology* 34:211–235.

Hedberg, Charlotta, and Tiit Tammaru. 2013. 'Neighbourhood effects' and 'city effects': The entry of newly arrived immigrants into the labour market. *Urban Studies* 50:1165–1182.

Herzog-Punzenberger, Barbara. 2003. Ethnic segmentation in school and labor market? 40 year legacy of Austrian guestworker policy. *International Migration Review* 37:1120–1144.

Hooijer, Gerda, and Georg Picot. 2015. European welfare states and migrant poverty: The institutional determinants of disadvantage. *Comparative Political Studies* 48:1879–1904.

Hox, Joop. 2010. *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Hox, Joop, Rens van de Schot and Suzette Mattijsse. 2012. How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods* 6:87–93.

Huber, Joan. 1991. Introduction. In *Macro-micro linkages in sociology*, ed. Joan Huber, pp. 6–9. Newbury Park, CA: Sage.

Johnston, Ron, Ibrahim Sirkeci, Nabil Khattab, and Tariq Modood. 2010. Ethno-religious categories and measuring occupational attainment in relation to education in England and Wales: A multilevel analysis. *Environment and Planning A: Economy and Space* 42:578–591.

Kalter, Frank, and Nadia Granato. 2002. Demographic change, educational expansion and structural assimilation of immigrants: The case of Germany. *European Sociological Review* 18:199–216.

Kalter, Frank, and Irena Kogan. 2006. Ethnic inequalities at the transition from school to work in Belgium and Spain: Discrimination or self-exclusion? *Research in Social Stratification and Mobility* 24:259–274

Kanas, Agnieszka, Chiswick Barry, Tanja van der Lippe and Frank van Tubergen. 2012. Social contacts and the economic performance of immigrants: A panel study of immigrants in Germany. *International Migration Review* 46:680–709.

Kanas, Agnieszka, Frank van Tubergen and Tanja van der Lippe. 2009. Immigrant self-employment: Testing hypotheses about the role of origin- and host-country human capital and bonding and bridging social capital. *Work and Occupations* 36:181–208.

Kesler, Christel. 2015. Welfare states and immigrant poverty: Germany, Sweden, and the United Kingdom in comparative perspective. *Acta Sociologica* 58:39–61.

Kesler, Christel, and Michael Hout. 2010. Entrepreneurship and immigrant wages in US labor markets: A multi-level approach. *Social Science Research* 39:187–201

Kim, Harris Hyun-Soo. 2018. A liability of embeddedness? Ethnic social capital, job search, and earnings penalty among female immigrants. *Ethnicities* 18:385–411.

Kislev, Elyakim. 2017. Deciphering the 'ethnic penalty' of immigrants in Western Europe: A cross-classified multilevel analysis. *Social Indicators Research* 134:725–745.

Kloosterman, Robert. 2010. Matching opportunities with resources: A framework for analysing (migrant) entrepreneurship from a mixed embeddedness perspective. *Entrepreneurship & Regional Development* 22:25–45.

Kloosterman, Robert, and Jan Rath. 2001. Immigrant entrepreneurs in advanced economies: Mixed embeddedness further explored. *Journal of Ethnic and Migration Studies* 27:189–201.

Kogan, Irena. 2006. Labor markets and economic incorporation among recent immigrants in Europe. *Social Forces* 85:697–721.

Kogan, Irena. 2007. *Working thouh barriers. Host country institutions and immigrant labour market performance in Europe*. Wiesbaden: Springer.

Kogan, Irena, and Frank Kalter. 2006. The effects of relative group size on occupational outcomes: Turks and ex-Yugoslavs in Austria. *European Sociological Review* 22:35–48.

Kroneberg, Clemens. 2019. Theory development in comparative social research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00604-y.

Lancee, Bram. 2012. The economic returns of bonding and bridging social capital for immigrant men in Germany. *Ethnic and Racial Studies* 35:664–683.

de Leeuw, Jan, and Erik Meijer. 2008. *Handbook of multilevel analysis*. New York: Springer.

Maas, Cora J. M., and Joop Hox. 2005. Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 1:85–91.

Meuleman, Bart, and Jean Billiet. 2009. A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods* 3:45–58.

Midtbøen, Arnfinn H. 2015. Ethnic penalties in Western labour markets: Contributions, explanations, critiques. *Nordic Journal of Migration Research* 5:185–193.

Munoz-Comet, Jacobo. 2016. Potential work experience as protection against unemployment: Does it bring equal benefit to immigrants and native workers? *European Sociological Review* 32:537–551.

Musterd, Sako, Roger Andersson, George Galster and Timo Kauppinen. 2008. Are immigrants' earnings influenced by the characteristics of their neighbours? *Environment and Planning A* 40:785–805.

Oberwittler, Dietrich. 2007. *The effects of ethnic and social segregation on children and adolescents: Recent research and results from a German multilevel study*. Discussion Paper No SP IV 2007-603 WZB Social Science Research Center Berlin.

O'Campo, Patricia. 2003. Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology* 157:9–13.

Ohlsson, Henrik, Per Broomé and Pieter Bevelander. 2012. Self-employment of immigrants and natives in Sweden—a multilevel analysis. *Entrepreneurship & Regional Development* 24:405–423.

Parsons, Craig. 2007. *How to map arguments in political science*. Oxford: University Press.

Perreira, Krista M., Kathleen Mullan Harris and Dohoon Lee. 2007. Immigrant youth in the labor market. *Work and Occupations* 34:5–34.

Pichler, Florian. 2011. Success on European labor markets: A cross-national comparison of attainment between immigrant and majority populations. *International Migration Review* 45:938–978.

Portes, Alejandro, and Jószef Böröcz. 1989. Contemporary immigration: Theoretical perspectives on its determinants and modes of incorporation. *International Migration Review* 23:606–630.

Portes, Alejandro. Ed. 1995. *The economic sociology of immigration: Essays on networks, ethnicity, and entrepreneurship*. New York: Russell Sage Foundation.

Raudenbush, Stephen, and Anthony Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*. Thousands Oaks, CA: SAGE.

Sampson, Robert J., Jeffrey D. Morenoff and Thomas Gannon-Rowley. 2002. Assessing "neighborhood effects": Social processes and new direction in research. *Annual Review of Sociology* 28:443–478.

Sanders, Jimy M., and Victor Nee. 1996. Immigrant self-employment: The family as social capital and the value of human capital. *American Sociological Review* 61:231–249.

Scheller, Friedrich. 2017. *The ambiguous role of ethnic context: A multilevel analysis of the relationship between group size and labor market integration of three immigrant groups in Germany*. Duisburger Beiträge zur soziologischen Forschung, No. 2017-03.

Schmidt-Catran, Alexander W., Malcolm Fairbrother, and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Simon, Patrick. 2003. France and the unknown second generation: preliminary results on social mobility. *International Migration Review* 37:1091–119

Skrondal, Anders, and Sophia Rabe-Hesketh. 2010. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Snijders, Tom A.B., and Roel J. Bosker. 2012. *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. 2nd edition. London: SAGE.

Spörlein, Christoph, and Frank Van Tubergen. 2014. The occupational status of immigrants in Western and non-Western societies. *International Journal of Comparative Sociology* 55:119–143.

Stegmueller, Daniel. 2013. How many countries for multilevel modeling? A comparison for frequentist and bayesian approaches. *American Journal of Political Science* 57:748–761.

Tammaru, Tiit, Magnus Strömgren, Olof Stjernström and Urban Lindgren. 2010. Learning through contact? The effects on earnings of immigrant exposure to the native population. *Environment and Planning A* 42:2938–2955.

Todaro, Michael, and Lydia Maruszko. 1987. Illegal migration and US immigration reform: A conceptual framework. *Population and Development Review* 13:101–111

Van Tubergen, Frank. 2005. Self-employment of immigrants: A cross-national study of 17 Western societies. *Social Forces* 84:709–732.

van Tubergen, Frank, Ineke Maas and Henk Flapp. 2004. The economic incorporation of immigrants in 18 Western societies: Origin, destination, and community effects. *American Sociological Review* 69:704–727.

Vermeulen, Hans, and Joel Perlman. Eds. 2000. *Immigrants, schooling and social mobility. Does culture make a difference?* London: Macmillan

Waldinger, Roger, Howard Aldrich and Robin Ward. 2006. *Ethnic entrepreneurs: Immigrant business in industrial societies*. Thousand Oaks, CA: Sage.

Wang, Qingfang. 2009. Gender, ethnicity, and self-employment: a multilevel analysis across US metropolitan areas. *Environment and Planning A* 41:1979–1996.

Wikström, Peroh, and Robert Sampson. 2003. Social mechanisms of community influences on crime and pathways in criminality. In *Causes of Conduct Disorder and Juvenile Delinquency*. Eds. Benjamin B. Lahey, Terrie E. Moffitt and Avshalom Caspi, p. 118. New York: The Guilford Press.

Wolbers, Maarten H.J., and Geert Driessen. 1996. Social class or ethnic background? Determinants of secondary school career of ethnic minority pupils. *The Netherlands Journal of Social Sciences* 32:109–26

Xie, Yu, and Margaret Gough. 2011. Ethnic enclaves and the earnings of immigrants. *Demography* 48:1293–1315.

Van Zanten, Agnes. 1997. Schooling immigrants in France in the 1990s: Success or failure of the Republican model of integration? *Anthropology & Education Quarterly* 28:351–374

Zolin, Roxanne, Artemis Chang, Xiaohua Yang and Elly Yi-Hsuan Ho. 2015. Social capital or ethnic enclave location? A multilevel explanation of immigrant business growth. *Thunderbird International Business Review* 58:453–463.

Romana Careja 2009 PhD, is Associate Professor at the Political Science Department, University of Southern Denmark. Her main publications are in the area of immigration policies and attitudes towards immigrants in Western European countries. Her recent publications include a special issue on the topic of possibilities and challenges of population registers for sampling immigrant minorities (edited together with H.J. Andreß, and published in *Comparative Migration Studies*); Making good citizens: local authorities' integration measures navigate national policies and local realities. Journal of Ethnic and Migration Studies 2018); Institutions, culture and migrants' preference for state-provided welfare: Longitudinal evidence from Germany. Journal of European Social Policy

Varieties of Capitalism and Welfare Regime Theories: Assumptions, Accomplishments, and the Need for Different Methods

Martin Schröder

Abstract This article reviews assumptions, contributions, and impasses of the two most important theories of capitalist diversity: Esping-Andersen's (1990) welfare regime and Hall and Soskice's (2001) varieties of capitalism typology. It shows that each theory implies a multilevel structure that nests lower-level units in upper-level units. However, even though this multilevel structure is at the center of many debates in the comparative capitalism literature, it is rarely explicitly modeled, let alone tested. Identifying this as an important route forward for research on capitalist diversity, I will show how future research could employ multilevel models to answer some of the most important questions about capitalist diversity.

Keywords Regime typology · Methodology · Mixed models · Hierarchical models

Theorien von Kapitalismusvarianten und Wohlfahrtsregimen: Annahmen, Erfolge und die Notwendigkeit anderer Methoden

Zusammenfassung In dem vorliegenden Artikel werden Annahmen, Erfolge und Sackgassen der beiden wichtigsten Theorien kapitalistischer Vielfalt beschrieben: der Wohlfahrtsregimetheorie Esping-Andersens (1990) und der Theorie über Kapitalismusvarianten von Hall und Soskice (2001). Es wird dargelegt, dass jede Theorie eine Mehrebenenstruktur impliziert, welche untere Einheiten in oberen verschachtelt. Obwohl diese Mehrebenenstruktur im Zentrum vieler ungelöster Debatten in der vergleichenden Kapitalismusliteratur postuliert wird, wird sie selten explizit modelliert, geschweige denn getestet. Der Artikel zeigt deswegen, wie die explizite Modellierung von Einbettungsstrukturen im Rahmen von bisher nicht genutzten

M. Schröder (✉)
Institut für Soziologie, Philipps-Universität Marburg
Ketzerbach 11, 35037 Marburg, Germany
E-Mail: martin.schroeder@uni-marburg.de

 Springer

Mehrebenenmodellen einige der wichtigsten Fragen der international vergleichenden Kapitalismusforschung beantworten kann, besonders die Frage, welche Kapitalismus- und Wohlfahrtsregimevarianten es gibt.

Schlüsselwörter Regimetypologie · Methodologie · Mehrebenemodelle · Hierarchische Modelle

1 Introduction

This article reviews assumptions, contributions, and impasses of the two most important theories of capitalist diversity. Based on this, I will show how these macro theories of capitalist diversity could use multilevel modeling to find answers to their most important questions (for multilevel modeling see Meuleman 2019; Schmidt-Catran et al. 2019). First, I will illustrate how Esping-Andersen's (1990) welfare regime theory differentiates between three worlds of welfare: a Scandinavian social democratic, a Continental European conservative, and an English-speaking liberal welfare regime. Second, I will show how Hall and Soskice (2001a) differentiate a liberal type of capitalism in English-speaking countries from a coordinated type of capitalism in Continental European and Scandinavian countries. I will explain the theoretical assumptions that underlie each approach and the literature's critique of it. I will then show how each theory implies a multilevel structure that nests lower-level units in upper-level units. But even though this multilevel structure is at the center of many debates about Esping-Andersen and varieties of capitalism, it is rarely explicitly modeled, let alone tested. Identifying this as an important route forward for research on capitalist diversity, I will show how future research could employ multilevel models to answer some of the most important questions about capitalist diversity.

2 What Can Researchers Expect from Theories of Capitalist Diversity?

Theories of capitalist diversity provide useful guidelines about which country differences to expect and which not to expect. For such theories to be useful, however, one must understand that the welfare and capitalist regimes that they describe are ideal types, to which countries adhere to a greater or lesser degree. Keeping the ideal-typical nature of regimes in mind neutralizes some of the most important criticisms against macro theories of capitalist diversity.

A first obvious criticism is that no two countries are ever alike, so that grouping them into welfare or capitalist regimes feigns an unwarranted simplicity. While two countries are certainly never alike, this is no argument against theories of capitalist diversity. Think about how we generally use typologies. Everyone agrees that every tree is different from every other tree. Nevertheless, it still makes sense to group trees into families, in spite of their individual divergence. Similarly, even though every country is unique, it makes sense to test *to what degree* countries match with and diverge from a regime type. Theories of capitalist diversity therefore never deny that

there are differences between countries. In fact, by grouping countries into regimes, they highlight what divergence to expect, and which aspects to find exceptional. Far from denying the existence of differences between countries, this allows national differences to be systematized in the first place.

Second, critics of theories of capitalist diversity argue that it is unclear how many regime types they should distinguish (Saint-Arnaud and Bernard 2003; Amable 2003; Ferragina et al. 2015). Is it enough to split the coordinated market economies of European countries away from the liberal market economies of English-speaking countries (Hall and Soskice 2001b)? Or should one further subdivide Continental European countries with conservative welfare states from Scandinavian countries with social democratic welfare states (Esping-Andersen 1990; Thelen 2014)? The answer is that neither undermines the utility of theories of capitalist diversity. Again, take the example of how we classify trees. The fact that we distinguish trees from other flora, and thus use a coarse distinction, does not undermine the utility of a more fine-grained division, which further distinguishes one type of tree from another. Similarly, theories of capitalist diversity allow one to discern coarser or more fine-grained country differences, depending on what is needed, just as maps of different scales serve different purposes.

Third, some argue against theories of capitalist diversity because some cases such as Japan or Switzerland seem unclassifiable (Dore 1997; Bonoli 2003). But if we wish to see whether a specific case fits into an existing system of categorization, we need to establish such categories in the first place. The fact that individual countries are sometimes not readily classifiable into one or another regime type is therefore no reason not to classify those countries that *can* be classified, to test whether unclassifiable cases are new combinations of old regimes or to develop new regime types.

Fourth, one can argue that capitalist countries can be differentiated by how they differ on certain variables, rather than subsuming them to one or another regime type. This is true, but such variables are collinear, which gives rise to theories of capitalist diversity in the first place. For example, countries with more generous unemployment insurance also tend to have more generous pension and health insurance, as well as coordinated wage bargaining and stricter labor protection laws (Ebbinghaus and Manow 2001; Schröder 2013). In other words, if some institutions within countries decommodify workers, others tend to do so as well, and this needs an explanation, which is provided by theories of capitalist diversity. So again, this is a reason to classify, rather than to abandon classifications. Let us once more take the tree analogy: one could describe every tree based on its variable attributes. But it also makes sense to highlight how trees that differ from others in one regard also tend to differ in other regards.

Two systems of classification have proven particularly useful to distinguish between different advanced capitalist countries. These are the relatively parsimonious theories of Esping-Andersen (1990) and Hall and Soskice (2001b). The sections below introduce their main ideas, accomplishments, and finally their problems, to suggest that many of these problems could be resolved by the application of multi-level modeling.

3 Esping-Andersen's Welfare Regime Theory

3.1 Theory

The main idea of Esping-Andersen's (1990, p. 26) welfare regime theory is that "the welfare state variations we find are [...] clustered by regime types." He suggests that welfare states cluster into an English-speaking, a Continental European, and a Scandinavian type, because each has a different "history of political class coalitions" (Esping-Andersen 1990, p. 1). This argument draws on the power resources approach, according to which strong welfare states are based on strong labor movements (Korpi 1985, 2006). But Esping-Andersen (1990, p. 30) reasoned that parties that represented the organized labor movement rarely had the necessary parliamentary majority to construct a welfare state, so that they had to form coalitions. These class coalitions took three forms, which led to three welfare regimes.

3.2 Three Welfare Regimes, Three Logics

Esping-Andersen argues that in Scandinavia, social democratic parties formed a coalition with small, capital-intensive, and politically well-organized farmers. This "red–green" coalition sowed the seeds for a universal welfare regime, which was successively extended to the middle class by providing ever-improving social services and public service jobs. In English-speaking countries, a liberal welfare regime remained residual because progressive parties found no one to enter into a coalition with. A third, so-called conservative welfare regime evolved in Continental Europe, where labor-intensive large-scale farmers were historically bound up with conservatives, which isolated the labor movement, so that Christian democratic parties created the welfare state on the basis of group-based solidarity, rather than on a universalism of all workers (Esping-Andersen 1990, p. 30).

Because social democratic parties were able to build a broad coalition in Scandinavia, they could make benefits progressively universal—not aiming them at the poor, but basing them on citizenship. For this to happen, benefits had to be so generous that they also appealed to the middle class, making this welfare regime the most decommodifying, as its "service is rendered as a matter of right, and [allows a person to] maintain a livelihood without reliance on the market" (Esping-Andersen 1990, p. 22).

The conservative welfare regime, mainly found in Continental Europe, can be very decommodifying as well, so that welfare states of this regime type are not necessarily smaller. However, contrary to social democratic welfare states, they redistribute within, rather than between social groups. Conservative welfare states function like an insurance policy. Those who pay a lot into the welfare system get a lot out if they become unemployed or when they retire. The individual proportionality between payments and benefits means that welfare states within the conservative regime type do virtually nothing to reduce social stratification. Instead, they are conservative in keeping everyone in their place within the social hierarchy (Esping-Andersen 1990, p. 24).

Last, the liberal welfare regime does not decommodify people, nor does it reduce social stratification. Instead, it follows the doctrine of liberalism, with benefits so low that they exclusively target the poor (Esping-Andersen 1990, p. 27). This welfare regime is embodied in the institutions of English-speaking countries. The simple distinction between social democratic, conservative, and liberal welfare states has been hugely influential. Esping-Andersen's (1990) original *Three Worlds of Welfare Capitalism* has been cited 27,000 times on Google Scholar. It rivals classics such as Max Weber's *The Protestant Ethic* and has provided a concise depiction of how welfare states differ, making it the workhorse of comparative sociology, political economy, and even much of economics (Scruggs and Allan 2006, p. 69).

3.3 An Implicit Micro–Macro Link in Each Welfare Regime

But this theory also contains a much-overlooked micro–macro link. The social democratic welfare regime presupposes solidarity among individuals, which is mirrored by similarly solidarity-based institutions in the unemployment, pension, and health systems. Institutions not only rely on individual solidarity (an influence that goes from the micro to the macro level); they also create it by handing out benefits when people need them. Because individuals are aware of how much the welfare state does for them in times of need, they are—at least in principle—willing to pay high taxes. While the macro institutions of social democratic welfare states could not survive without individuals' commitment to egalitarian justice, this individual egalitarianism in turn results from the macro institutions of social democratic welfare states. Their institutions are not only similar *within* the welfare states of Denmark, Norway, Sweden, and Finland, but also *across* these countries, collectively constituting a cross-national welfare regime, which differs from conservative and liberal welfare regimes.

Conservative welfare states also contain a specific micro–macro link, as individuals are said to be motivated by solidarity toward their social group. Due to this fragmented solidarity, individuals contribute to unemployment, health, and pension insurance systems that are divided along class lines, so that micro-level limitations in solidarity influence macro-level institutions. However, conservative welfare states not only rely on the fragmented solidarity that is typical in these countries, but also engender it, as they redistribute within social groups, but not across them. Such conservative welfare state institutions are not only similar within the unemployment, pension, and health systems of each country, but also across different Continental European countries.

Last, the liberal welfare regime also contains a specific micro–macro link, as individuals are said to hold pro-market attitudes, which in turn lead to residual welfare institutions. That these welfare institutions only cater to the poorest teaches everyone else to expect nothing of the welfare state, and this again reinforces liberal attitudes. This race to the bottom of liberal attitudes and institutions takes place in all English-speaking countries that form this welfare regime. In this sense, the idea behind each welfare regime is that "institutions give rise to certain interests and norms, which in turn either reinforce or undermine the original institutions.

We observe, in other words, a bi-directional causal logic between institutions and interests/norms" (Rothstein 1998, p. 135).

3.4 Empirical Studies

Empirical studies confirmed that the populations of different welfare regimes indeed harbor social justice views that mirror Esping-Andersen's theory (Arts and Gelissen 2001, p. 297; see also the results in Mehrtens 2004; Breznau 2010, p. 479 f.; Svallfors 2012). They also showed that welfare regimes are connected to a number of important outcomes, such as gender pay gaps (Mandel and Shalev 2009), possibilities to reconcile work and family life (Esping-Andersen 2009), labor market inequalities (Palier 2010; Palier and Thelen 2010; Chauvel and Schröder 2014, 2015), and majoritarian versus proportional voting systems (Manow 2009). Cluster analyses also confirmed that Esping-Andersen's theory fits with empirical indicators behind welfare states (Saint-Arnaud and Bernard 2003; Ferragina and Seeleib-Kaiser 2011; Ferragina et al. 2015; Schröder 2017). In short, a plethora of studies has documented many of the similarities and differences between welfare states that Esping-Andersen postulated. However, others found important aspects where the theory was deficient.

3.5 The Literature's Critique of Esping-Andersen's Theory

One important criticism against Esping-Andersen's welfare regime theory stated that it should consider "new" policy fields, such as family policy, rather than merely the classical welfare state responsibilities of sickness, retirement, and unemployment. Esping-Andersen indeed enlarged his regime theory accordingly, showing that characteristic differences between welfare regimes also exist concerning family policy: conservative welfare states encourage women to stay at home. Social democratic welfare states provide a large public sector for women to work in, as well as publicly funded childcare, both of which encourage female work, but segregate it into a sheltered public sector. Liberal welfare states neither discourage nor support female employment, as both would interfere with free markets (Esping-Andersen 2009). Taking the role of the family into account also led some to suggest a fourth regime type for Southern European countries. It was argued that in this "Mediterranean" welfare regime, families were replacing the welfare state (Lessenich 1994; Ferrera 1996; Ferragina et al. 2015). Others suggested that there was a Japanese or Asian regime (Esping-Andersen 1997); yet others suggested a post-communist regime (Castles and Obinger 2008, p. 338 f.). Seeing how one regime after another was added, some took this criticism to its logical conclusion, arguing that each country is unique, so that no country can be subsumed under one regime type or another (Crouch and Streeck 1997). Another critique was that institutions within countries are so heterogeneous that it makes no sense to speak of country-wide policy styles, let alone regime-wide ones (Kasza 2002; Scruggs and Allan 2006). Recent reviews therefore suggest that "[w]hile some authors have indeed discerned regime-consistent country differences in welfare state support there is much variation also among countries belonging to the same regime" (Kumlin and Stadelmann-

Steffen 2014, p. 6). The same critique has been made against varieties of capitalism, which distinguishes not between welfare states, but between types of economies.

4 Hall and Soskice's Varieties of Capitalism

4.1 Theory

Similar to Esping-Andersen's welfare regime theory, varieties of capitalism not only separates countries into types of capitalism, but also provides a theory to explain these differences. Varieties of capitalism starts from the opposite theoretical premise of Esping-Andersen's welfare regime theory, by suggesting that it is not the power of labor but employers' production models that lead to two types of capitalism: a market-based and a "coordinated" one. But why would firms ever militate against market arrangements in the economy? Hall and Soskice (2001a) argue that firms that continually improve existing products have an interest in coordinating with competitors and their employees. This is not easy because while both are better off if such coordination takes place, each has a short-term interest in cheating the other. This draws on the theory of neo-corporatism, which essentially postulates that economies are most efficient when workers and employers are either highly organized, because then they can help companies to work together, or highly unorganized, so that they cannot constrain companies. In contrast, labor and employer organizations with intermediate strength are too weak to help companies cooperate in game-theoretical dilemmas, but too strong to permit companies to enjoy the benefits ensuing from free markets (see Crouch 1993, p. 12; Lehmbruch and Schmitter 1979, 1982; Calmfors and Driffill 1988).

4.2 Two Types of Capitalism, Two Logics

According to varieties of capitalism, two antithetic types of capitalism are therefore efficient, each in their own right. Essentially, the theory differentiates between the same countries as Esping-Andersen's welfare theory, juxtaposing the liberal market economies (LMEs) of English-speaking countries to the coordinated market economies (CMEs) of Continental Europe and Scandinavia. But the varieties of capitalism approach added the revolutionary twist to Esping-Andersen's theory that firms themselves can have an interest in promoting non-liberal coordinated economic institutions which nurture "competitive advantages that depend on high levels of regulation" (Hall and Soskice 2001a, p. 63). Yet only companies that incrementally improve existing products need a high level of regulation. Consider automobile production, a typical industry with incremental innovation, as each new car model is intended to improve upon the former. Such long-term continual improvement needs workers to dedicate their working lives to improving one car. Varieties of capitalism argues that workers are unwilling to learn the specific skills necessary for this if they cannot be sure that they will stay with the same company for life (Estevez-Abe et al. 2001). Therefore, if companies want workers to learn company-specific skills, they have to insure them against unemployment. This gives companies a rational interest

in generous unemployment insurance. Hall and Soskice also argue that firms which continually improve existing products need cooperative workforce relations, so that they favor employee codetermination. To facilitate long-term cooperation with their workforce, firms also have an interest in outsourcing the conflictual wage bargaining process to all-embracing employer associations and trade unions. Varieties of capitalism therefore suggests that companies in coordinated countries have a comparative advantage when it comes to incremental innovations, while companies in liberal countries are better at radical innovations that profit from the flexibility of free markets, rather than from long-term cooperation with the workforce (Hall and Soskice 2001a, p. 40 ff.).

Hall and Soskice also argue that more innovative countries grow at a faster rate, so that coherently coordinated and coherently liberal institutions are complementary in furthering economic growth. For example, strict layoff protection is complementary to employee codetermination, which is complementary to long-term financing, all of which supports a coordinated type of economy that brings incremental innovation and thereby economic growth. Hall and Soskice (2001a, p. 18) therefore suggest "that nations with a particular type of coordination in one sphere of the economy should tend to develop complementary practices in other spheres as well," so that over time, countries develop types of capitalist institutions that are either fully coordinated or fully liberal. Firms should even lobby for "less market" in coordinated types of capitalism, as they "attempt to preserve arrangements in one sphere of the economy in order to protect complementary institutions or synergies with institutions elsewhere" (Hall and Soskice 2001a, p. 64). Companies that pursue incremental innovations may therefore push for more employee codetermination, stricter protection against being laid off, generous unemployment insurance and stronger trade unions in an attempt to improve their competitiveness.

While long-term cooperation may help companies to continually improve products, it does little to foster radical innovation, which flourishes best on free markets with little regulation that allow capital and labor to flow where and when companies need them. The needs of radically innovating companies therefore clash with those of incrementally innovating ones. Companies that specialize in radical innovations do not require long-term cooperation with their workforce, so they oppose employee codetermination. They eschew long-term capital, needing venture capital to finance their ideas. They suffer from rigid labor laws, as they have to hire and fire workers for new projects. Firms that specialize in radical innovations therefore push countries to liberalize labor law and financing. This led to the revolutionary idea that liberal market economies are bound to become more liberal, while coordinated market economies are bound to become more coordinated in order to increase their competitiveness in global markets (Hall and Soskice 2001a, p. 57 f).

Based on this argument, the varieties of capitalism school could claim that extensive welfare states followed the wishes of capitalists, rather than the power of organized labor (Estevez-Abe et al. 2001; Swenson 2002; see also Mares 2001, p. 184). It thereby explained the large welfare states of coordinated and social democratic countries through efficiency rather than class struggle, contradicting the concept that the power resources of labor led to non-liberal types of capitalism (Esping-Andersen 1985; Korpi 2006). Thus, while the power resources school claimed

that coordinated countries were built against the interests of capitalists where labor was powerful, the varieties of capitalism school claimed that coordinated countries catered to the interests of capitalists, and indeed were constructed at their behest. This seemingly irreconcilable debate was resolved by the Solomonic judgment that employers were not the farsighted enthusiastic supporters of welfare legislation that varieties of capitalism first portrayed them as. However, employers objected less strongly to social policy when they could integrate it into their production model by specializing in long-term incremental innovations (Paster 2011). Proponents of the varieties of capitalism approach accepted this, arguing that the power resources view provides an accurate picture of the low-wage/low-skill sector, which mainly produces for the domestic market. They maintained, however, that in the high-wage/high-skill sector that produces to satisfy international demand, the insurance function of the welfare state is better explained by "politics for markets," i.e., through the desire of employers, rather than the "politics against markets" of the labor movement (Iversen and Soskice 2015).

That economic regulation could promote competiveness, rather than impede it, was probably the most important insight of varieties of capitalism. It was largely unheard of in the comparative political economy literature, but it answered some of its most pressing questions, such as why German and Swedish firms were so different from American or British ones, and why this did not make them less but in fact more successful. Varieties of capitalism, for the first time, gave a good functional reason as to why liberal and coordinated types of capitalism are bound to become more diverse on increasingly competitive world markets.

The insights provided by varieties of capitalism were also applied to many fields that did not seem obvious at first. It explained how liberal market economies make it easier for women to combine work and family by arguing that "CME companies that invest heavily in company-specific skills will be worried about hiring women who may leave the company" (Soskice 2005, p. 173; see also Estevez-Abe 2005; Iversen and Rosenbluth 2010). Its "skill specificity" approach could show why CMEs have an interest in upholding a vocational training system that teaches specific skills, while LMEs support a larger university sector that teaches general skills (see Busemeyer 2009; Busemeyer and Trampusch 2011). Further developments of Varieties of capitalism included a cultural approach, which showed that the same pro-market attitudes that promote liberal welfare states also promote liberal economic regulation. Conversely, market skepticism not only promoted coordinated economic institutions, but also larger welfare states (Ebbinghaus and Manow 2001; Schröder 2009, 2013; Ahlborn et al. 2016).

4.3 The Micro–Macro Link of Varieties of Capitalism

Varieties of capitalism also proposes a very strong link from macro institutions to micro behavior, as it argues that "[i]n any national economy, firms will gravitate toward the mode of coordination for which there is institutional support" (Hall and Soskice 2001a, p. 9). But the link goes back from the micro to the macro level, as firms in turn support the institutions that they have learned to use to their competitive advantage (Hall and Soskice 2001a, p. 63). Because both links prevent

change, institutions and companies are locked into a production model that is either based on widespread coordination that leads to and relies on incremental innovation, or widespread market arrangements that lead to and rely on radical innovation. This implies that individual companies even support institutions that restrict their freedom, as long as these institutions help them to cooperate with others.

4.4 Empirical Studies

Empirical studies were indeed able to confirm many of varieties of capitalism's central tenets. Countries with purely coordinated or purely liberal institutions indeed seemed to enjoy more rapid economic growth (Hall and Gingerich 2009; but see also Kenworthy 2006). Thelen (2012, 2014) traced how the most liberal countries (the US and the UK) became even more liberal, while Scandinavian countries became more inclusive and Continental European countries split workers further into insiders and outsiders. She argued that each variety of capitalism reinforces the logic that already marks its production system. Others found that varieties of capitalism explains why English-speaking countries push for free market regulation in multilateral negotiations, while coordinated market economies promote international social protection laws (Fioretos 2001; see also Hall 2014).

4.5 The Literature's Critique Against Varieties of Capitalism

Many voices were however critical of the theory's extremely parsimonious separation into only two types of capitalism. Some suggested an additional regime type for France (Schmidt 2002), others for China (Redding and Witt 2009; Peck and Zhang 2013), Italy (Trigilia and Burroni 2009), and Spain (Molina and Rhodes 2007). Nölke and Vliegenthart (2009) singled out Eastern European countries as dependent market economies which manufacture products that liberal and coordinated market economies invent. Schneider (2009) suggested that Latin American countries are hierarchical market economies, whilst Witt and Redding (2013) thought that Asian countries deserved their own variety of capitalism. Boyer (2005) subdivided the dual distinction of liberal and coordinated market economies into four varieties. Amable (2003) even proposed five types of capitalism, and recent tests found nine types of business systems among 61 countries that account for more than 90% of the world's gross domestic product (GDP; Witt et al. 2018). Another critique went in the opposite direction. It did not argue that countries within one regime type are more heterogeneous than previously assumed, but that major institutional heterogeneity exists even within countries, with companies creating coordinated islands in liberal market economies and liberal islands in coordinated market economies (Crouch et al. 2009a, b; Schröder and Voelzkow 2016).

However, neither have such debates fostered agreement on whether more regime types indeed explain greater cross-country variation, nor on which countries deserve a third, fourth, or fifth regime type, or whether greater variation lies at the level of firms, or of institutions that vary within countries or between countries, or even between groups of countries. I will argue below that the unresolved questions in Esping-Andersen's and Hall and Soskice's typologies are therefore structurally sim-

ilar, and that they could therefore benefit from conceptualizing their open questions as debates about embedding structures.

5 New Methods to Test Capitalist Variety

Existing methodologies were unable to answer the most pressing questions about the most important macro typologies. While cluster analyses showed which country clusters exist in the first place, they were less useful when it comes to confirming or disconfirming the assumption that lower-level units such as individuals or firms are embedded in clusters at the country level (Amable 2003; Schröder 2009, 2013; Witt et al. 2018). While qualitative case studies showed how policymaking is similar within countries of the same regime type, they were less useful in delivering guidelines about how much variation is explained through similarities within regimes versus differences between them (Hall 1986, 1993; Dobbin 1994; Streeck 2009; Thelen 2014).

I therefore argue that Esping-Andersen's welfare regime theory and varieties of capitalism implicitly discuss embedding structures and that multilevel methodology could therefore resolve some of the most important questions that plague these typologies of capitalist diversity. While it is conceptually important to understand this, let me be clear that existing multilevel models rely on structures that are not always present in typologies of capitalist diversity. For example, two or three regimes types are too few in number for commonly used multilevel modeling methods (see Schmidt-Catran et al. 2019). With this in mind, I urge scholars to develop a multilevel methodology in order to answer the most important questions about capitalist diversity, and I urge scholars who study capitalist diversity to understand that many of their debates discuss embedding structures. The next section shows how this could answer some of the most pressing questions that currently plague macro typologies.

5.1 Testing the Multilevel Structure of Esping-Andersen's Regime Theory

In varieties of capitalism and Esping-Andersen's regime theory, individual behavior and attitudes can be seen as a level-1 (micro) variable. For Esping-Andersen, individuals are embedded in pension, unemployment, and health systems, institutions which can be understood as level-2 variables. Databases such as the *Comparative Welfare Entitlements Dataset* measure such variables, showing how generously the health, unemployment, and pension systems replace wages in case of need for different kinds of individuals. Cross-embedding each person in the health, pension, and unemployment systems of which he or she is a member can show the degree to which the generosity of each system varies within a country, but also between countries. For example, people in one country may be in a generous health system, but at the same time in an ungenerous pension system. If the appropriate micro-level data were available, one could show how generously, e.g., the retirement system replaces the wages of one individual compared to another. This shows how internally coherent a country's institutions are, and this can be compared to the internal coherence of another country's institutions. To measure this, one needs to nest individuals on

level 1, the welfare institutions in which they are embedded on level 2, and both in countries as level 3. By embedding the third level of countries in a fourth level of regimes, one can even show whether the attitudes and institutions of different countries are more homogeneous within welfare regimes than between them. Keep in mind that the current methodology may not allow the clustering of lower-level units in regimes as a level, since there are only three regime types. A simple alternative may be to use dummy variables at the level of countries to measure regimes. It may also not always be possible (or indeed necessary) to use all four levels in actual multilevel regressions.

Nonetheless, much of the above critique against Esping-Andersen could be tested by modeling, and thus by testing (parts of) the multilevel structure that these theories postulate. For example, scholars who argue that new subfields such as family policy must be analyzed argue that people should not only be embedded in institutions that are more or less decommodifying in case of sickness, unemployment, or retirement, but also upon founding a family. Those who argue that more than three regime types exist technically simply argue that a level-4 regime variable explains much greater variation when it takes on more than three distinct values. As I mentioned above, scholars have also criticized that the institutions of countries do not conform perfectly to "their" regime type. Proponents of Esping-Andersen's regime theory technically simply argue that much between-country variation is explained through the clustering of individuals, institutions, and countries into regimes. Multilevel regressions can show how much variation of, e.g., decommodification or stratification, is indeed explained through regimes rather than through countries alone. One would therefore suppose that multilevel models have been routinely used in the regime modeling business. But in fact, multilevel models have rarely been used to answer the most pressing questions that arise in the welfare regime literature.

Jaeger (2006) explains individual support for redistribution by a two-level model that nests individuals in countries. However, this does not tell us how much individual variation (level 1) is explained through the institutions in which people are embedded (level 2), the countries in which these institutions are embedded (level 3), and the regimes in which the countries are embedded (level 4). Deeming and Hayes (2012) use multilevel regressions to show that happiness mainly varies between individuals rather than between countries. However, this does not show how much variation is explained through individual-level versus country-level, versus regime-level variation, so it remains unclear whether national- or regime-level characteristics explain happiness. Dallinger (2010, p. 344 f.) shows that two thirds of the variation in support for redistribution lies at the level of welfare regimes. However, this fails to explain how much of this variation is at the level of countries (level 3). Therefore, such studies cannot tell us how much individual variation (on level 1) is explained through the welfare institutions in which people are embedded (level 2), the countries in which the institutions are embedded (level 3), and the welfare regimes in which countries are embedded (level 4).

Existing studies indicate that such a design is feasible in principle. For example, Kallio and Niemelä (2014, p. 14 f.) show that residual cross-country variation on attitudes towards the poor declines by 36–56% after welfare regimes have been included. Similarly, Richter et al. (2012, p. 8) calculate that country variation in

self-rated health is about 30% lower after including welfare regimes in a multilevel model. Others even suggest a reduction of about 50% (Rathmann et al. 2015, p. 419). This indicates that variation between countries can indeed be meaningfully explained when individuals, institutions, and countries are grouped into regimes. Some studies even use a three-level design that clusters individuals (level 1) into regions (level 2) and into countries (level 3). This indicates that individual differences explain about 90% of variation in health, with only about 9% explained through countries, and almost no variation explained through regions. Including welfare regimes as an additional explanatory variable (not clustering level) reduces the country-level variation from 9 to about 5% (Eikemo et al. 2008, p. 2288 f.). This suggests that while most individual differences in health can be explained through actual individual variation (level 1), about half of the remaining variation may be explained through differences in welfare regimes, while idiosyncratic country differences explain the remaining half.

Calculating such effects using multilevel models could pacify bitter debates about whether countries are more or less internally homogeneous, by answering fairly technical questions about which level explains the greatest variation for a variable of interest. However, existing studies virtually never employ such a design, so that some of the most important questions in the field of welfare regimes remain unanswered. Reviews of the field therefore argue that some studies test regime and country variation, while others test "individual-level variation within one country in policy experiences," while scholars should go beyond "one or the other research strategy and [to apply] what we may call cross-level thinking" (Kumlin and Stadelmann-Steffen 2014, p. 319). This has not yet happened, so that many of the most important questions about welfare states remain unanswered. Such a misfit between unresolved questions and approaches to answer them also characterizes the debates on varieties of capitalism.

5.2 Testing the Multilevel Structure of Varieties of Capitalism

Technically speaking, varieties of capitalism also suggests a specific type of multilevel structure: firms (as level-1 units) are embedded in institutions that are more or less coordinated, such as wage setting, training systems, and workplace representation. Such institutional variables, as for instance those coded in the *Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts* (Visser 2015, 2016), show whether employees are represented on the board of directors, how far-reaching their codetermination rights are, how all-embracing wage negotiations are, and how employer and labor organizations coordinate. It would be preferable to have these variables on the company level, but most of them only exist as country-level aggregations. Varieties of capitalism posits that such institutions are relatively similar within countries (level 3) and within a regime of coordinated and liberal capitalism (level 4).

Many debates about the validity of the varieties of capitalism relate implicitly to this embedding structure. For example, scholars who argue for more than two regime types suggest that more regimes explain much greater variation, while those who wish to retain only two regime types counter that the increase in explained variation

through more regime types is negligible. Multilevel regressions can calculate who is right. Keep in mind, however, that since varieties of capitalism only distinguishes between two types of capitalism, it probably makes more sense to model the regime level with dummy variables on the country level, rather than as an actual level-4 variable.

Other critics have not merely suggested reclassifying countries or subdividing regimes. More radically, they have assailed varieties of capitalism's allegedly "pervasive tendency to methodological nationalism and spatial archetyping, in which the coherence of national regulatory configurations is presumed rather than demonstrated" (Peck and Theodore 2007, p. 750; similarly, see Allen 2004, p. 105). Technically, such critics argue that little variation is explained through level 4 (the clustering of countries into regimes) or level 3 (the clustering of institutions within countries). Others suggested that institutional arrangements vary with economic sectors, rather than between national types of capitalism (Hollingsworth 1991; Kitschelt 1991; Hollingsworth et al. 1994). Others still have posited that the institutions which companies use depend strongly on their embeddedness in subnational *regional* production systems (Piore and Sabel 1984; Storper 1997; Cooke et al. 2004; Parker and Tamaschke 2005; Crouch et al. 2001, 2004, 2009a, 2009b; Schröder and Voelzkow 2016). Technically, all of these critiques simply argue that a great deal of variation lies on level 2: institutions that vary within countries. This can be tested through multilevel regressions. One simply needs data on the institutions that firms use, such as whether a firm uses works councils, collective wage agreements, etc. One would then have to measure how much variation in the use of such institutions lies on level 1 (between firms), level 2 (intra-country variation between regions or economic sectors), level 3 (between countries), and level 4 (between regime types).

However, critics not only claim that greater variation lies on lower levels than varieties of capitalism is ready to concede. They also argue that varieties of capitalism underestimates variation on higher levels, ignoring that all countries may become neoliberal (for this critique, see Cerny et al. 2005; Soederberg et al. 2005; Bohle and Greskovits 2009). In its most extreme variant, this critique suggests that "political economy might have to abandon entirely the idea of national varieties of capitalism," as all types of capitalism become neoliberal (Streeck 2010, p. 38). Multilevel models could test this because such critiques—technically speaking—argue that no meaningful variation exists at level 3 (countries) and level 4 (regimes), but that all variation instead resides in a level-5 supercluster (capitalism), so that change within countries over time is more important than enduring differences between countries and regimes. In other words, these critics suggest that the within-variation, which is supposedly similar in all types of capitalism, explains greater variation than the between-variation, which durably differentiates between types of capitalism. Multilevel models could show whether the within-change of a clustering of all countries in years explains greater variation than do the between-differences through clustering in countries and regimes at each point in time. In this sense, many of the most important questions about varieties of capitalism can be answered through indicators that multilevel regressions calculate routinely, so that an important "route for future research is the analysis of within-country variation of institutional configurations" (Schneider and Paunescu 2012, p. 748).

However, multilevel approaches have rarely been used to answer some of the most important questions about varieties of capitalism, even though some studies show how this could resolve important questions. Bechter and Brandl (2015, p. 433) use data from 27 states and 18 economic sectors in order to show that "the sector context matters more nowadays than the country context." However, as they lack company-level data, they can only show whether the deviation from the mean for an institutional configuration is lower on the level of sectors or countries. Thus, they cannot tell how variation in company arrangements is distributed between companies, institutions, countries, and regimes. To my knowledge, no study uses multilevel regression to show how much variation lies on which level, making this a promising avenue for future research.

6 Conclusion

I have reviewed assumptions, accomplishments and impasses of varieties of capitalism and Esping-Andersen's welfare regime theory. This allowed me to show how discussions about macro typologies are often discussions about embedding structures, as scholars question how lower-level units such as firms and individuals are embedded in national institutions, how these institutions are embedded in countries, and how these countries are embedded in welfare and production regimes. But even though both critics and proponents of theories of welfare and production regimes discuss typical multilevel problems, they seem unaware that multilevel models may provide answers to some of their most gridlocked debates. For example, the welfare regime literature wants to know how much variation in individual behavior, attitudes, and outcomes can be explained through institutional arrangements in unemployment, pension, and health systems. The literature on varieties of capitalism wants to know how much variation between individual firms can be explained by the clustering of firms in sectoral and regional arrangements. Varieties of capitalism could use multilevel regressions to calculate how much variation is explained when clustering companies (level 1) in sectors or regions within countries (level 2), in different countries (level 3), and in types of capitalism (level 4). Welfare regime theory could test whether people (level 1) behave similarly or hold similar attitudes depending on the institutions in which they are embedded (level 2), the countries in which they live (level 3), and the country regimes in which they are embedded (level 4). As I mentioned, however, it may not always be possible to actually model an institutional level 2, as institution-specific data for individuals may not be comparable across countries and time, or because micro-level data that show in which health, pension, and unemployment systems people are embedded are unavailable. Generally, the levels that I am speaking of should be treated as conceptual, analytical devices. For example, the fourth level is only an analytical one, since there are no actors or institutions at the level of regimes. Note also that the typical number of three or four regimes at the fourth level may not be enough to treat them as actual levels, so they may have to be modeled using dummy variables.

Notwithstanding such problems, multilevel modeling could show how much variation of lower-level units can be explained when more regimes are distinguished,

e.g., through a fourth type of welfare regime or a third type of capitalism. They could also test which clustering of level 3 units in level 4 explains the greatest variance. For example, is greater variance explained when Ireland is conceptualized as a liberal market economy, a coordinated one, or even as a third type? Multilevel models could also test whether the within-change of a supercluster over time (level 5, all capitalist countries) explains greater variation of lower-level units than does the variation at level 3 (between countries) and level 4 (between regimes) at each point in time. In other words, a panel dataset could show whether change that is common to all countries and country regimes explains greater variation than enduring differences between countries and country regimes do. Note again, however, that perfect data will be hard to come by. It would be ideal to have firm- or person-level data that also show the different national institutions in which individuals or firms are embedded. Such data would allow one to see whether, e.g., bargaining within companies, decommodification, or stratification of individuals depends on individual variation, within-country variation, between-country variation, or transnational differences.

Results of such studies will very much depend on what one is trying to explain. Since varieties of capitalism tries to explain innovativeness, usually measured through patents, the most important question could be whether differences between firms, between the institutions in which firms are embedded, between the countries in which institutions are embedded, or between country regimes explain the variation in innovativeness. Second, varieties of capitalism argues that countries with either highly liberal or highly coordinated institutions have a higher growth rate. It would therefore be important to see whether countries with highly coordinated or highly liberal institutions—and thus little variation between institutions on level 2—indeed have more rapid economic growth (Hall and Soskice 2001a, p. 37 ff.; Hall and Gingerich 2009).

For Esping-Andersen's theory, it would be most relevant to test whether decommodification and stratification vary between individuals, between the institutions in which individuals are embedded, between the countries in which these institutions are embedded, or between the regimes in which countries are embedded. Variables such as poverty, health, social trust, and altruism have been explained through welfare regimes. Again, a multilevel approach could show the degree to which they vary between individuals, the different institutions in which individuals are embedded, the countries that embed institutions, and the regimes that embed countries. If results differ widely depending on the dependent variable, then this indicates where macro typologies are useful, and where they are not.

It is notably indicators such as the intraclass correlation coefficient (ICC), which show the degree to which clustering occurs at each level, while the variance partition coefficient (VPC) shows how variation is explained through different levels. As some of the most important debates about welfare regimes and varieties of capitalism are about these questions, such indicators can give empirical answers to some of the most hotly debated topics in these fields. It is important to keep in mind, however, that using multilevel regressions and their typical indicators, poses some problems in itself. One problem, which has already been mentioned, is that the typical number of regimes that varieties of capitalism and welfare regime theory propose (two to five) is too small to be used as separate clusters. One could therefore instead model

clusters as dummy variables (1 = belonging to a cluster, 0 = not), enter them into a regression model, and test how much variation remains on lower-level units, e.g., how much between-country variation is still explained through lower-level units when countries are grouped into regimes vs. when they are not. Note also that while multilevel regressions provide such indicators, researchers still have to decide whether, e.g., explaining 5% more variation by adding one more variety of capitalism is worthwhile. Judgment calls still have to be made, but at least they can then be made based on hard data.

Some researchers "have been always hostile to classifications, [arguing that] they constitute artificial constructions," suggesting that countries can only be understood through "detailed case analysis" (Ferragina and Seeleib-Kaiser 2011, p. 584). Such scholars are unlikely to be satisfied with any attempt to improve theories of capitalist diversity. However, if much variation can be explained through capitalist regimes, proponents of macro theories of capitalist diversity know that their assumptions rest on a solid footing in spite of such criticism. Conversely, critics of macro-typologies have a point if little variation can be explained through country regimes. So far, debates between country specialists and macro-regime scholars have been dialogues of the deaf, with one side showing that a country is different from another, and the other showing that both countries are nonetheless part of the same regime. Instead of arguing about the proverbial half-full glass, multilevel regressions can show precisely how full the glass is, that is, how much variation can be explained through the embedding of companies and individuals in institutions, countries, and country regimes.

In this sense, some of the most important questions about welfare theories of capitalist diversity must no longer be the topic of philosophical debates. Modeling the multilevel structure that these theories imply but cannot test could put a number on who is (how) right and who is (how) wrong. This could yield concrete results, such as that, e.g., 70% of the explicable variation in firms' strategies or individual welfare attributes depends on the actual firm or individual (level 1), while another 10% depends on the variation between institutions within countries (level 2), another 10% depends on institutions that are similar within countries (level 3), and yet another 10% depends on institutions that are similar across countries and thus within regimes (level 4). No doubt scholars will then find a reason why a given percentage is a lot, or is not, but at least debates could then rely on empirical facts, rather than on opinions.

References

Ahlborn, Markus, Joachim Ahrens and Rainer Schweickert. 2016. Large-scale transition of economic systems—Do CEECs converge toward western prototypes? *Comparative Economic Studies* 58: 430–454.

Allen, Matthew. 2004. The varieties of capitalism paradigm: Not enough variety? *Socio-Economic Review* 2:87–108.

Amable, Bruno. 2003. *The diversity of modern capitalism.* Oxford: Oxford University Press.

Arts, Wil, and John Gelissen. 2001. Welfare states, solidarity and justice principles: Does the type really matter? *Acta Sociologica* 44:283–299.

Bechter, Barbara, and Bernd Brandl. 2015. Measurement and analysis of industrial relations aggregates: What is the relevant unit of analysis in comparative research? *European Political Science* 14:422–438. https://doi.org/10.1057/eps.2015.65.

Bohle, Dorothee, and Béla Greskovits. 2009. Varieties of capitalism and capitalism "tout court". *Archives Européennes de Sociologie* 3:355–386.

Bonoli, Giuliano. 2003. The welfare state in Switzerland. In *Welfare capitalism around the world,* ed. Christian Aspalter, 179–197. Taichung: Casa Verde Publishing.

Boyer, Robert. 2005. How and why capitalisms differ. *Economy and Society* 34:509–557.

Breznau, Nate. 2010. Economic equality and social welfare: Policy preferences in five nations. *International Journal of Public Opinion Research.*

Busemeyer, Marius. 2009. Asset specificity, institutional complementarities and the variety of skill regimes in coordinated market economies. *Socio-Economic Review* 7:375–406.

Busemeyer, Marius, and Christine Trampusch. Eds. 2011. *The political economy of collective skill formation.* Oxford: Oxford University Press.

Calmfors, Lars, and John Driffill. 1988. Bargaining structure, corporatism and macroeconomic performance. *Economic Policy* 3:13–61.

Castles, Francis G., and Herbert Obinger. 2008. Worlds, families, regimes: Country clusters in European and OECD area public policy. *West European Politics* 31:321–344.

Cerny, Philip G., Georg Menz and Susanne Soederberg. 2005. Different roads to globalization: neoliberalism, the competition state, and politics in a more open world. In *Internalizing globalization: The rise of neoliberalism and the decline of national varieties of capitalism,* eds. Susanne Soederberg, Georg Menz and Philip Cerny, 1–32. New York: Palgrave.

Chauvel, Louis, and Martin Schröder. 2014. Generational inequalities and welfare regimes. *Social Forces* 92:1259–1283.

Chauvel, Louis, and Martin Schröder. 2015. The impact of cohort membership on disposable incomes in West Germany, France, and the United States. *European Sociological Review* 31:298–311.

Cooke, Philip, Martin Heidenreich and Hans-Joachim Braczyk. 2004. *Regional innovation systems: The role of governance in a globalized world.* London: Routledge.

Crouch, Colin. 1993. *Industrial relations and European state traditions.* Oxford: Oxford University Press.

Crouch, Colin, and Wolfgang Streeck. Eds. 1997 *Political economy of modern capitalism: Mapping convergence and diversity.* London: Sage.

Crouch, Colin, Patrick Le Galès, Carlo Trigilia and Helmut Voelzkow. Eds. 2004. *Changing governance of local economies: Responses of European local production systems.* Oxford: Oxford University Press.

Crouch, Colin, Patrick Legalès, Carlo Trigilia and Helmut Voelzkow. Eds. 2001. *Local production systems in Europe: Rise or demise?* Oxford: Oxford University Press.

Crouch, Colin, Martin Schröder and Helmut Voelzkow. 2009a. Regional and sectoral varieties of capitalism. *Economy and Society* 38:654–678.

Crouch, Colin, Martin Schröder and Helmut Voelzkow. 2009b. Conclusions: Local and global sources of capitalist diversity. In *Innovation in local economies. Germany in comparative context,* eds. Colin Crouch and Helmut Voelzkow, 169–188. Oxford: Oxford University Press.

Dallinger, Ursula. 2010. Public support for redistribution: What explains cross-national differences? *Journal of European Social Policy* 20:333–349.

Deeming, Christopher, and David Hayes. 2012. Worlds of welfare capitalism and wellbeing: A multilevel analysis. *Journal of Social Policy* 41:811–829.

Dobbin, Frank. 1994. *Forging industrial policy. The United States, Britain, and France in the railway age.* Cambridge: Cambridge University Press.

Dore, Ronald. 1997. The distinctiveness of Japan. In *Political Economy of Modern Capitalism: Mapping Convergence and Diversity,* eds. Colin Crouch and Wolfgang Streeck, 19–32. London: Sage.

Ebbinghaus, Bernhard, and Philipp Manow. Eds. 2001. *Varieties of welfare capitalism.* London: Routledge.

Eikemo, Terje Andreas, Clare Bambra, Ken Judge and Kristen Ringdal. 2008. Welfare state regimes and differences in self-perceived health in Europe: A multilevel analysis. Social *Science & Medicine* 66:2281–2295.

Esping-Andersen, Gøsta. 1985. *Politics against markets: The social democratic road to power.* Princeton, NJ: Princeton University Press.

Esping-Andersen, Gøsta. 1990. *The three worlds of welfare capitalism.* Cambridge: Polity Press.

Esping-Andersen, Gøsta. 1997. Hybrid or unique? The Japanese welfare state between Europe and America. *Journal of European Social Policy* 7:179–189.

Esping-Andersen, Gøsta. 2009. *The incomplete revolution.* Cambridge: Polity.

Estevez-Abe, M. 2005. Gender bias in skills and social policies: The varieties of capitalism perspective on sex segregation. *Social Politics* 12:180–215.

Estevez-Abe, Margarita, Torben Iversen and David Soskice. 2001. Social protection and the formation of skills: A reinterpretation of the welfare state. In *Varieties of capitalism: The institutional foundations*

of comparative advantage, eds. Peter Hall and David Soskice, 145–183. Oxford: Oxford University Press.

Ferragina, Emanuele, and Martin Seeleib-Kaiser. 2011. Thematic review: Welfare regime debate: Past, present, futures? *Policy & Politics* 39:583–611.

Ferragina, Emanuele, Martin Seeleib-Kaiser and Thees Spreckelsen. 2015. The four worlds of 'welfare reality'—social risks and outcomes in Europe. *Social Policy and Society* 14:287–307.

Ferrera, Maurizio. 1996. The 'Southern Model' of welfare in social Europe. *Journal of European Social Policy* 6:17–37.

Fioretos, Orfeo. 2001. The domestic sources of multilateral preferences: "Varieties of capitalism" in the European community. In *Varieties of capitalism": The institutional foundations of comparative advantage,* eds. Peter Hall and David Soskice, 213–247. Oxford: Oxford University Press.

Hall, Peter. 1986. *Governing the economy: The politics of state intervention in Britain and France.* Cambridge: Polity Press.

Hall, Peter. 1993. Policy paradigms, social learning, and the state. The case of economic policymaking in Britain. *Comparative Politics* 25:275–296.

Hall, Peter. 2014. Varieties of capitalism and the Euro crisis. *West European Politics* 37:1223–1243.

Hall, Peter, and Daniel W. Gingerich. 2009. Varieties of capitalism and institutional complementarities in the political economy: An empirical analysis. *British Journal of Political Science* 39:449–482.

Hall, Peter, and David Soskice. 2001a. An introduction to varieties of capitalism. In *Varieties of capitalism: The institutional foundations of comparative advantage,* eds. Peter Hall and David Soskice, 1–71. Oxford: Oxford University Press.

Hall, Peter, and David Soskice. Eds. 2001b. *Varieties of capitalism: The institutional foundations of comparative advantage.* Oxford: Oxford University Press.

Hollingsworth, Rogers. 1991. The logic of coordinating american manufacturing sectors. In *Governance of the American economy,* eds. John Campbell, Rogers Hollingsworth and Leon Lindberg, 35–73. Cambridge: Cambridge University Press.

Hollingsworth, J. Rogers, Philippe C. Schmitter and Wolfgang Streeck. 1994. *Governing capitalist economies: Performance and control of economic sectors.* New York: Oxford University Press.

Iversen, Torben, and Frances Mccall Rosenbluth. 2010. *Women, work, and politics: The political economy of gender inequality*. New Haven, CT: Yale University Press.

Iversen, Torben, and David Soskice. 2015. Politics for markets. *Journal of European Social Policy* 25:76–93.

Jæger, Mads Meier. 2006. Welfare regimes and attitudes towards redistribution: The regime hypothesis revisited. *European Sociological Review* 22:157–170.

Kallio, Johanna, and Mikko Niemelä. 2014. Who blames the poor? Multilevel evidence of support for and determinants of individualistic explanation of poverty in Europe. *European Societies* 16:112–135.

Kasza, Gregory, J. 2002. The illusion of welfare 'regimes'. *Journal of Social Policy* 31:271–287.

Kenworthy, Lane. 2006. Institutional coherence and macroeconomic performance. *Socio-Economic Review* 4:69–91.

Kitschelt, Herbert. 1991. Industrial governance structures, innovation strategies, and the case of Japan: Sectoral or cross-national comparative analysis? *International Organization* 45:453–493.

Korpi, Walter. 1985. Power resources approach vs. action and conflict: On causal and intentional explanations in the study of power. *Sociological Theory* 3:31–45.

Korpi, Walter. 2006. Power resources and employer-centered approaches in explanations of welfare states and varieties of capitalism: Protagonists, consenters, and antagonists. *World Politics* 58:167–206.

Kumlin, Staffan, and Isabelle Stadelmann-Steffen. 2014. How welfare states shape the democratic public: Borrowing strength across research communities. In *How welfare states shape the democratic public. Policy feedback, participation, voting, and attitudes,* eds. Staffan Kumlin and Isabelle Stadelmann-Steffen, 311–325. Cheltenham: Edward Elgar.

Lehmbruch, Gerhard, and Philippe Schmitter. 1979. *Trends toward corporatist intermediation.* Beverly Hills, CA: Sage.

Lehmbruch, Gerhard, and Philippe Schmitter. 1982. *Patterns of corporatist policy-making.* Beverly Hills, CA: Sage.

Lessenich, Stephan. 1994. Three worlds of welfare capitalism – oder vier? Strukturwandel arbeits- und sozialpolitischer Regulierungsmuster in Spanien. *Politische Vierteljahresschrift* 35:224–244.

Mandel, Hadas, and Michael Shalev. 2009. How welfare states shape the gender pay gap: A theoretical and comparative analysis. *Social Forces* 87:1873–1911.

Manow, Philip. 2009. Electoral rules, class coalitions and welfare state regimes, or how to explain Esping-Andersen with Stein Rokkan. *Socio-Economic Review* 7:101–121.

Mares, Isabela. 2001. Firms and the welfare state: When, why, and how does social policy matter to employers? In *Varieties of capitalism: The institutional foundations of comparative advantage,* eds. Peter Hall and Soskice David, 184–212. Oxford: Oxford University Press.

Mehrtens, F. John. 2004. Three worlds of public opinion? Values, variation, and the effect on social policy. *International Journal of Public Opinion Research* 16:115–143.

Meuleman, Bart. 2019. Multilevel Structural Equation Modeling for Cross-National Comparative Research. In Cross-national comparative research – analytical strategies, results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00605-x.

Molina, Óscar, and Martin Rhodes. 2007. The political economy of adjustment in mixed market economies. A study of Spain and Italy. In *Beyond varieties of capitalism,* eds.Bob Hancké, Martin Rhodes and Mark Thatcher, 223–252. Oxford: Oxford University Press.

Nölke, Andreas, and Arjan Vliegenthart. 2009. Enlarging the varieties of capitalism: The emergence of dependent market economies in east central Europe. *World Politics* 61:670–702.

Palier, Bruno Ed. 2010. *A long goodbye to Bismarck? The politics of welfare reforms in continental Europe.* Amsterdam: Amsterdam University Press.

Palier, Bruno, and Kathleen Thelen. 2010. Institutionalizing dualism: Complementarities and change in France and Germany. *Politics & Society* 38:119–148.

Parker, Rachel, and Louise Tamaschke. 2005. Explaining regional departures from national patterns of industry specialization: Regional institutions, policies and state coordination. *Organization Studies* 26:1787–1807.

Paster, Thomas. 2011. *The role of business in the development of the welfare state and labor markets in Germany: Containing social reforms.* New York: Routledge.

Peck, Jamie, and Nik Theodore. 2007. Variegated capitalism. *Progress in Human Geography* 31:731–772.

Peck, Jamie, and Jun Zhang. 2013. A variety of capitalism … with Chinese characteristics? *Journal of Economic Geography* 13:357–396.

Piore, Michael, and Charles Sabel. 1984. *The second industrial divide.* New York: Basic Books.

Rathmann, Katharina, Viktoria Ottova, Klaus Hurrelmann, Margarethe De Looze, Kate Levin, Michal Molcho, Frank Elgar, Saoirse Nic Gabhainn, Jitse P Van Dijk and Matthias Richter. 2015. Macro-level determinants of young people's subjective health and health inequalities: A multilevel analysis in 27 welfare states. *Maturitas* 80:414–420.

Redding, Gordon, and Michael Witt, 2009: China's business system and its future trajectory. *Asia Pacific Journal of Management* 26:381–399.

Richter, Matthias, Katharina Rathman, Saoirse Nic Gabhainn, Alessio Zambon, William Boyce and Klaus Hurrelmann. 2012. Welfare state regimes, health and health inequalities in adolescence: A multilevel study in 32 countries. *Sociology of Health & Illness* 34:858–879.

Rothstein, Bo. 1998. *Just institutions matter: The moral and political logic of the universal welfare state.* Cambridge: Cambridge University Press.

Saint-Arnaud, Sébastien, and Paul Bernard. 2003. Convergence or resilience? A hierarchical cluster analysis of the welfare regimes in advanced countries. *Current Sociology* 51:499–527.

Schmidt, Vivien. 2002. *The futures of European capitalism.* Oxford: Oxford University Press.

Schmidt-Catran, Alexander W., Malcolm Fairbrother, and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schneider, Ben Ross. 2009. Hierarchical market economies and varieties of capitalism in Latin America. *Journal of Latin American Studies* 41:553–575.

Schneider, Martin R., and Mihai Paunescu. 2012. Changing varieties of capitalism and revealed comparative advantages from 1990 to 2005: A test of the Hall and Soskice claims. *Socio-Economic Review* 10:731–753.

Schröder, Martin. 2009. Integrating welfare and production typologies. How refinements of the varieties of capitalism approach call for a combination with welfare typologies. *Journal of Social Policy* 38:19–43.

Schröder, Martin. 2013. *Integrating varieties of capitalism and welfare state research: A unified typology of capitalisms.* New York: Palgrave.

Schröder, Martin. 2017. Varieties of capitalisms and welfare states. Eroding diversity? In *Handbook of European social policy,* eds. Patricia Kennett and Noemi Lendvai, 45–59. Cheltenham: Edward Elgar.

Schröder, Martin, and Helmut Voelzkow. 2016. Varieties of regulation: How to combine sectoral, regional and national levels. *Regional Studies* 50:7–19. https://doi.org/10.1080/00343404.2014.904040.

Scruggs, Lyle, and James Allan. 2006. Welfare-state decommodification in 18 OECD countries: A replication and revision. *Journal of European Social Policy* 16:55–72.

Soederberg, Susanne, Georg Menz and Philip Cerny. Eds. 2005. *Internalizing globalization: The rise of neoliberalism and the decline of national varieties of capitalism.* New York: Palgrave.

Soskice, David. 2005. Varieties of capitalism and cross-national gender differences. *Social Politics* 12:170–179.

Storper, Michael. 1997. *The regional world: Territorial development in a global economy.* New York: Guilford Press.

Streeck, Wolfgang. 2009. *Re-forming capitalism: Institutional change in the German political eonomy.* Oxford: Oxford University Press.

Streeck, Wolfgang. 2010. *E Pluribus Unum? Varieties and commonalities of capitalism.* MPIfG Discussion Paper 10/12. Cologne. Max Planck Institute for the Study of Societies.

Svallfors, Stefan. 2012. Welfare attitudes in context. In *Contested welfare states: Welfare attitudes in Europe and beyond,* ed. Stefan Svallfors, 1–21. Stanford, CA: Stanford University Press.

Swenson, Peter. 2002. *Capitalists against markets: The making of labor markets and welfare states in the United States and Sweden.* Oxford: Oxford University Press.

Thelen, Kathleen. 2012. Varieties of capitalism: Trajectories of liberalization and the new politics of social solidarity. *Annual Review of Political Science* 15:137–159.

Thelen, Kathleen. 2014. *Varieties of liberalization and the new politics of social solidarity.* Cambridge: Cambridge University Press.

Trigilia, Carlo, and Luigi Burroni. 2009. Italy: Rise, decline and restructuring of a regionalized capitalism. *Economy and Society* 38:630–653. https://doi.org/10.1080/03085140903190367.

Visser, Jelle. 2015. Codebook. *Data base on institutional characteristics of trade unions, wage setting, state intervention and social pacts, 1960–2014 (ICTWSS). Version 5.0.* Amsterdam. Amsterdam Institute for Advanced Labour Studies AIAS.

Visser, Jelle. 2016. *ICTWSS database, version 5.1.* Amsterdam. Amsterdam Institute for Advanced Labour Studies AIAS.

Witt, Michael A., and Gordon Redding. 2013. Asian business systems: Institutional comparison, clusters and implications for varieties of capitalism and business systems theory. *Socio-Economic Review.*

Witt, Michael A., Luiz Ricardo Kabbach De Castro, Kenneth Amaeshi, Sami Mahroum, Dorothee Bohle and Lawrence Saez. 2018. Mapping the business systems of 61 major economies: A taxonomy and implications for varieties of capitalism and business systems research. *Socio-Economic Review* 16:5–38. https://doi.org/10.1093/ser/mwx012.

Martin Schröder Prof. Dr., Professor for Economic Sociology, University of Marburg. Research focus: social inequality, comparative capitalism and welfare state research, inequality, gender differences on labor markets, life satisfaction research. Publications: Integrating varieties of capitalism and welfare state research: A unified typology of capitalisms. New York 2013; Varianten des Kapitalismus. Die Unterschiede liberaler und koordinierter Marktwirtschaften. Wiesbaden 2014; Welfare states and their inequality as a result of cultural differences instead of varieties of capitalism. Wiesbaden 2016; Varieties of capitalisms and welfare states. Eroding diversity? Cheltenham 2017; Varieties of regulation: How to combine sectoral, regional and national levels. Regional Studies 50, 2016 (with H. Voelzkow).

**Kölner Zeitschrift
für Soziologie und
Sozialpsychologie**

ABHANDLUNGEN

Values in Life Domains in a Cross-National Perspective

Loek Halman · John Gelissen

Abstract The increase in international survey research projects investigating basic orientations reveals that the grand sociological theories fall short in explaining the often large differences between populations in contemporary societies that continue to exist. There is more than modernization to explain varieties. Institutions, culture, history, policies, all appear to affect people's values. Our review of current state-of-the-art cross-national research activities demonstrates that increasingly multilevel analysis techniques are applied to address either modernization or institutionalism, or both. From various theoretical perspectives, hypotheses are extracted about the impact of the context on values in a specific domain. The selected studies include very different contextual features to explain varieties in domain-specific value orientations, which clearly proves that there are different mechanisms at work in the distinct value domains. The studies reveal that "context matters," but also that it is essential to include individual-level characteristics, at least as controls. Quite often, the individual attributes appear differently distributed in different countries, which may be the main reason why differences in value orientations between countries remain.

L. Halman (✉) · J. Gelissen
Department of Methodology and Statistics, TS Social and Behavioral Sciences, Tilburg University
Warandelaan 2, PO Box 90153, 5037 AB Tilburg, The Netherlands
E-Mail: loek.halman@uvt.nl

J. Gelissen
E-Mail: J.P.T.M.Gelissen@uvt.nl

 Springer

Werte in unterschiedlichen Lebensbereichen in einer ländervergleichenden Bestandsaufnahme

Zusammenfassung Die Zunahme an internationalen Umfrageforschungsprojekten, die grundlegende Wertorientierungen untersuchen, zeigt, dass die großen soziologischen Theorien nicht ausreichen, um die oft großen Unterschiede zwischen den Bevölkerungen in heutigen Gesellschaften zu erklären. Es gibt mehr als nur Modernisierung, um die Unterschiede zu erklären. Institutionen, Kultur, Geschichte und Politik scheinen alle die Werte der Menschen zu beeinflussen. Die vorliegende Überprüfung der aktuellen internationalen länderübergreifenden Forschungsaktivitäten zeigt, dass zunehmend Mehrebenenanalysen verwendet werden, um entweder Modernisierung oder Institutionalismus oder beide theoretischen Ansätze gemeinsam zu untersuchen. Aus verschiedenen theoretischen Perspektiven werden Hypothesen über die Auswirkungen des Kontexts auf Werte in einem bestimmten Lebensbereich generiert. Die ausgewählten Studien untersuchen sehr unterschiedliche kontextuelle Merkmale, um die Varianz domänenspezifischer Wertorientierungen zu erklären, was eindeutig belegt, dass unterschiedliche Mechanismen in den verschiedenen Lebensbereichen wirken. Die Studien zeigen, dass „Kontext" von Bedeutung ist, aber auch, dass es notwendig ist, Merkmale auf individueller Ebene zu berücksichtigen, zumindest als Kontrollvariablen. Häufig sind die individuellen Merkmale in den verschiedenen Ländern unterschiedlich verteilt, was der Hauptgrund dafür sein kann, dass Unterschiede in den Wertorientierungen zwischen den Ländern bestehen.

Schlüsselwörter Moralvorstellungen · Europa · Modernisierung · Institutionalismus · Multilevel · Mikro- und Makroeffekte

1 Introduction

This paper takes us on a journey into some recent empirical studies on moral and religious beliefs, political views, family and gender issues, and work values. All of these regard people's values to result from influences of characteristics of the national context in which people live and of their sociological background characteristics. Efforts have been made in recent years to test theoretically informed hypotheses about micro and macro effects on values, and combining micro- and macro-level determinants has become the benchmark in many sociological studies, most of which focus on one or more specific domain or domains of values.

Although there have been attempts to find overarching value systems covering values in all life spheres, most studies show that modern society appears to be highly fragmented regarding its fundamental social values. The various value domains do not constitute clear patterns of values: Values appear domain specific (Halman and de Moor 1987, 1994; Kerkhofs 1997). This finding is not surprising. Some authors have argued that due to modernization processes such as differentiation and specialization, life domains have become independent sectors in society with their own values and independent of religion (Durkheim 1964; see also Smelser 1973; Meulemann 1983).

Because value patterns appear fragmented, value domains can be explored separately, which is something that we will be doing in this paper. We focus on orientations regarding morality, religion, relationships, politics, and work. The selected studies regard values as consequences of individual and macro features, and test related theoretically informed hypotheses using state-of-the-art multilevel models (Fox 2016; Hox et al. 2017). In particular, all selected studies use random intercept models or—if cross-level interaction hypotheses are considered—random intercept-and-slopes models to investigate the complex interplay between macro-level and individual-level determinants of values. The data analytical design of the studies that we discuss takes the country level as the macro level, although we also include studies that additionally—or rather alternatively—consider lower-level regional contextual effects rather than, or in addition to, country-level effects (see also Meuleman and Gorres, Siewert and Wagemann in this volume). The studies demonstrate that there is more than modernization, understood as economic advancement, to explain differences in values, and that institutions are significant determinants of values as well.

Before we present the studies, we start this paper with a brief discussion of the modernization perspective, followed by a reflection on some of the methodological issues that studies on explaining country differences in value domains have to cope with.

2 Theoretical and Methodological Perspectives

2.1 Modernization Theory

Survey projects, such as the European Values Study (EVS) and World Values Surveys (WVS), European Social Survey (ESS), and the International Social Survey Programme (ISSP), reveal significant similarities but above all dissimilarities in values across European countries. Attempts to understand the variation in basic value orientations often refer to modernization theory. The central claim of this theory is that socioeconomic development goes hand in hand with coherent and—to some extent—predictable societal changes (Marsh 2014). A second common element is that less developed societies acquire characteristics that are common to more developed societies (e.g., Lerner 1968, p. 386). Thus, "if and as the level of modernization increases, the level of structural uniformity among relatively modernized societies continually increases" (Marsh 2014, p. 279). This would imply that pre-modern and modern societies differ from each other much more than do the many varieties of modern society (Schmidt 2010, p. 516). Hence, convergence in values is to be expected.

The claims of modernization theory have been contested and refuted. For example, Gundelach (1994) found that institutional factors appeared to constitute far better explanations of variations in basic orientations than economic development did. Proponents of modernization theory defend their theory by stating that modernization is not deterministic but probabilistic. Socio-economic development seems to bring predictable cultural and political changes, and economic collapse tends

to bring predictable changes in the opposite direction (Inglehart and Welzel 2005, p. 20). However, the idea that modernity as it has developed in the Western world will be taken over in all modernizing societies should be qualified. Modern societies may have many features in common, but differences persist between them regarding mentalities, institutions, and other factors (Sachsenmaier 2002, p. 42). This naturally resembles the idea of path dependency. Although societies seem to develop in similar and predictable directions, the actual trajectory is dependent on historical backgrounds, persisting traditions, and other country-specific factors (Inglehart and Welzel 2005, pp. 19, 20). Modernization is taking place all over the world, but in different shapes and very specific and unique ways. There is no single version of modernization (Preyer and Sussman 2016, p. 10; Eisenstadt 2002).

2.2 Within-Country Variation

Studies that analyze values in different countries confirm the ideas of multiple modernities. The *Atlas of European Values* (Halman et al. 2011) reveals that Europe is far from homogeneous on many issues, and that country differences seem to persist. Inglehart's (2018) cultural map of the world also provides evidence for rejecting the idea of convergence in orientations.

Countries' positions are, however, inferred on such maps from the countries' means calculated from individuals' values. In doing so, it is neglected that countries can have the same mean value for a characteristic, but differ considerably in the variance of the characteristic. Societies with similar means on a specific orientation can be homogeneous, but also very heterogeneous. Within-country distributions should therefore not be overlooked.

Taking into account distributions within countries is also important because countries may differ significantly in the composition of important individual-level explanatory variables. If such characteristics are distributed unevenly across countries, then such compositional differences may explain not only differences between individuals, but also differences between countries (Diez-Roux 2002). Such composition effects are often not controlled for in aggregate-level analyses.

Investigating relationships between macro-level characteristics of societies often results in high percentages of explained variances in value orientations. However, Robinson's (1950) article on "ecological fallacy" made it clear that one should avoid making inferences about individual behavior from aggregate data because strong associations between macro characteristics cannot necessarily be generalized to lower levels. It can therefore be highly misleading to investigate value patterns and their changes at the macro level only, and expect similar associations to occur at the individual level. One cannot fully understand the internal logic of societies or cultures in the terms used for individuals' personality dynamics. Coleman's (1990) famous "boat" illustrates the complex connections between the micro and macro levels. As Coleman showed, associations can exist between variables at the macro level because such variables either relate directly to each other or because a macro-level variable exerts an influence on another macro-level variable via mechanisms that run via individual-level characteristics, or both.

It is obvious that people do not live in isolation, but in a context which constrains or stimulates, and, as such, the cultural and social climate in a society codetermines people's attitudes, values, and behaviors. People adjust to what they experience and what other people around them are doing. Thus, despite individualization and increased levels of personal autonomy in modern societies, individuals are not entirely free but constrained, not only by their context, but also by the way in which they are raised and socialized. Parents, relatives, friends, and peer groups mold and determine people's feelings, ideas, beliefs, and values.

Individuals are naturally not deterministic victims of social pressure and circumstances, but such factors make it more likely that specific outcomes for these individuals will occur. People carry around in their heads a basic stock of knowledge that is the result of living in a particular country, culture, or subculture, but they do not mechanically respond to external stimuli. People have an inner mental life and highly subjective experiences. People's actions, attitudes, and values are therefore a *joint product* of individuals' needs, traits, temperament, culture, socialization, and personal experiences (Mead 1934). Institutions, the social context, and history all play a role, though it is the individual who makes the final decision in the end. People's responses can be in line with these norms, but they can also deviate from the group's standards. To investigate this, a multilevel approach is needed in which both levels—individual features and context characteristics—are accounted for simultaneously.

2.3 Addressing Multilevel Issues

The recent increase in international comparative (survey) data sources, together with the rapid development of advanced statistical methods to analyze micro and macro levels simultaneously, enable researchers to address these multilevel issues. No wonder, then, that numerous articles and papers have appeared in scientific journals and at conferences investigating research problems combining micro-level with macro-level explanations of sociological phenomena.

The idea that both individual features and context characteristics influence people's values and actions is of course far from new in the sociological discipline. What is, however, somewhat of a recent development is that sociologists use the possibility to investigate micro and macro determinants simultaneously. Before the advent of multilevel statistical modeling, analyses combining micro and macro features were often done country by country. The effects of individual characteristics were investigated for each country separately, and compared with the effects that such characteristics had within other countries. Such a country-specific approach was applied in *The Civic Culture* (Almond and Verba 1965) to examine whether or not effects were similar, or whether different effects were reported country by country in countries with different political-historical experiences and findings.

Country-by-country analyses were also presented in more contemporary works. Inglehart (1977) confirmed his theory of *The Silent Revolution* by testing his hypotheses in each of the selected countries separately and in *Political Action* (Barnes and Kaase 1979), five Western democracies were analyzed separately to explore the political action repertoires and their antecedents. A repeat study was carried out in

a more limited number of countries, also including country characteristics, testing the expectation that "the process of modernization will lead to more or less similar individual responses of the people in these societies" (Van Deth and Jennings 1989, p. 17). And although the contributors to the volume edited by Van Deth and Scarbrough on the *Impact of Values* (1995) applied somewhat more sophisticated statistical methods, they organized and presented their analyses country by country.

An early attempt to investigate context effects was made by De Moor (1987) in analyzing the European Values Study data on religion and morality. Country dummies were included in regression analyses in order to explain individual differences in religious and moral values. It appeared that the country in which people live does indeed have an effect, but which country attribute or attributes causes this effect remained unclear because country dummies do not reveal what country-level factor has an effect. Referring to countries without defining or clarifying what country features explain differences adds virtually no new information to the observation that there appear to be differences across people in different countries. The problem is, however, that it has hardly been thoroughly theorized what these features are or can be. As such, notions such as country, nation, and state are merely "black boxes" hiding many features, which might and will be important (see Kroneberg 2019 in this special issue).

The enormous amount of (survey) data available to social scientists these days makes it painfully clear that there is a problem of how to cope with all the varieties and similarities, and how to interpret and explain them. The grand theories of modernization appear far too general to explain the variations that exist between populations in different countries. De Moor (1994, p. 232) therefore concluded that "empirically founded partial theories are needed." The challenge for contemporary comparative sociology is to develop theories to understand and explain how similarities and differences at the macro-level lead to micro-level differences and similarities.

Value researchers should test such theories by replacing country dummies with precisely defined country characteristics (Przeworski and Teune 1970), but even then ordinary OLS (ordinary least squares) regression analyses risk overestimating the significance of country effects (i.e., they provide too liberal a test) because calculations are based on the numbers of individual cases and not on the number of countries. To limit the risk that one too quickly concludes that "context characteristics matter," multilevel analysis techniques should be applied, and these have become standard practice in contemporary sociological journals to investigate multilevel issues. It means that theories are developed to elaborate mechanisms for effects at both levels and sometimes combine both levels and extract cross-level interactions from these theories to explain that individual-level effects may be different in certain contexts/circumstances.

3 Making Sense of the Variety in Values

We selected studies for inclusion in this contribution that were conducted in the domains of morality, religion, politics, family, and work which explored the im-

pact of both individual and macro characteristics and applied multilevel analyses to understand and explain the value differences and similarities in European countries.

3.1 Moral Views: Personal and Civic Morality

Several empirical studies on moral views which rely on survey data from the European Values Study distinguish between two moral dimensions "personal morality" and "civic morality." The first refers to relational or interpersonal behaviors and sexual conduct, whereas the second deals with minor cheating or dishonesty or petty fiddling and activities which contravene the law (see e. g., Phillips and Harding 1985; De Moor 1987; Halman 1991; Halman and Vloet 1994; see Table 1).

Letki (2006) explored civic morality, i.e., compliance with the law and public order or respect of and obedience to the norms and rules. She formulated hypotheses about the effects of individual levels of trust, regional levels of social capital, and macro or country levels of institutional configurations (Letki 2006, p. 311), and analyzed data from WVS in 38 countries. The quality of a country's government, which was measured by government effectiveness, regulatory quality, the rule of law, and control of corruption (from the World Bank Governance Indicators Dataset), and economic performance (Gross Domestic Product (GDP) growth and level of unemployment; Letki 2006, p. 321) appeared to be important attributes. At the individual level, determinants of civic morality are socio-economic characteristics, but the hypothesized effects of generalized trust and the vibrancy of associational life, indicating regional levels of social capital, were not confirmed. The finding that "the dimensions of institutional quality which are relevant to ordinary people's lives, to competence and transparency as well as to the efficiency of the civil service, the level of contract enforcement, the incidence of corruption, levels of crime and the level of unemployment" (Letki 2006, p. 321) affect civic morality is groundbreaking, since researchers often measure institutional quality by general variables such as GDP growth or the level of democracy. Such findings reveal that the institutional context matters for people's attitudes towards compliance.

Apart from hypothesizing that more highly religious people will hold more conservative views on morality (see Siegers 2019 in this special issue), Finke and Adamczyk (2008) also explored the effect of religiosity at the macro level, arguing that morality will be stronger in more religious nations (Finke and Adamczyk 2008, p. 619). They also formulated a cross-level interaction hypothesis, holding that the impact of an individual's religiosity will be stronger in more religious contexts, and they additionally expected differential effects of religiosity for different kinds of morality, with effects being strongest for moral issues which are not sanctioned by the state. They relied for the analyses on ISSP and WVS data, and their measures of morality correspond to the dimensions that had been distinguished earlier by Stoetzel (1983) and by Phillips and Harding (1985): Morality that is sanctioned by legal codes and morality that is not sanctioned by the state. At the country level, they included both an aggregated measure of the combination of church attendance and religious importance in ISSP, and other variables such as religious concentration (Herfindahl index), migration, corruption, dominant religion, and homosexuality regulation. At the individual level, religiosity was not measured by church attendance alone, but

Table 1 Overview of Macro- and/or Meso-Level Explanatory Characteristics included in the selected Values Studies

Study	Dependent variable(s)	Explanatory variables at country level	Explanatory variables at regional level	Data source(s), timespan, no. of countries
Values domain: Moral orientations				
Letki (2006)	Civic morality	Level of democracy; quality of government; GDP growth per year; level of unemployment	Social trust; vibrancy associational life	EVS 1999–2002 38 countries
Finke and Adamczyk (2008)	Civic morality (morality sanctioned by the state)	(a) Religiosity From ISSP: aggregate from factor scale combining measures of attendance and religious importance From WVS: aggregate measure of religious importance	–	EVS-WVS 1997 46 countries
	Sexual morality	(b) Religious concentration; migration; corruption; dominant religion; homosexuality regulation	–	ISSP 1998 32 countries
Sieben and Halman (2015)	Civc morality	Ex-Soviet; democracy index; GDP	–	EVS 23 countries time
Storm (2016)	Autonomy vs. tradition; self-interest vs. social norms	Religiosity; confidence in institutions	–	EVS 1981, 1990, 1999, 2008 46 countries
Values domain: Religious beliefs and practices				
Hirschle (2013)	Church attendance; religious beliefs	Economic performance (GDP)	–	Eurobarometer; ISSP; EVS; WVS; ESS; ALLBUS
Te Grotenhuis et al. (2015)	Church attendance	Social security expenditure; GDP; average level of education	–	The Mannheim Eurobarometer Trend File containing trend items from Eurobarometers between 1970 and 2002
Schwadel (2015)	Standardized measure of prayer, belief in God, self-identification as religious and religious service attendance	Education: mean university degree; communist country; government regulation index; GDP; mean no. religion	–	ISSP 2008; IMF
Storm (2017)	Church attendance; self-rated degree of religiosity	GDP (log of); social security benefits	–	ESS 2002–2014

Table 1 (Continued)

Study	Dependent variable(s)	Explanatory variables at country level	Explanatory variables at regional level	Data source(s), timespan, no. of countries
Interpersonal and social trust				
Olson and Li (2015)	Generalized trust	Percent religious; religious heterogeneity; religious culture; economic inequality; national wealth; government quality	–	EVS/WVS 1999–2004, 2005–2007, 2008–2010; Religion and State Project; World Development Indicators
Rapp (2016)	Generalized trust	Dispersion; bimodality; constraint; economic prosperity; economic inequality; ethnic diversity; predominant religion	–	WVS 2005–2008
Hooghe et al. (2009)	Generalized trust	26 measurements of diversity; GDP; income inequality	–	ESS 2002–2003; OECD 2005; UNHCR 2000, 2004
Beugelsdijk and Klasing (2016)	Generalized trust	Value diversity; ethnic fractionalization; linguistic fractionalization; religious fractionalization; genetic diversity; ethnic segregation; linguistic segregation; ethno-linguistic diversity	Political ideology; GINI; GDP; population; per cent Protestants; per cent Catholics; Eastern Germany	WVS/WVS
Values domain: Gender issues				
André et al. (2013)	Support for traditional female roles	Female labor force participation; parental leave (years); % GDP on childcare	–	ISSP 1988, 1994, 2002; ILO 2010 2011; OECD 2012
Naldini et al. (2016)	Women reduce working time or quit their jobs to take care of their needy parents	% of GDP on care for elderly; level of service coverage; % of public spending on LTC; familialist care culture; GDP; unemployment rate; incidence of part-time employment on total employment; % of access to early retirement pensions age	–	Eurobarometer 2007
Dotti Sani and Quaranta (2017)	Attitudes towards gender roles	Gender inequality index; years of democracy	–	Civic and Citizenship Education Study 2009; UNDP 2010

Table 1 (Continued)

Study	Dependent variable(s)	Explanatory variables at country level	Explanatory variables at regional level	Data source(s), timespan, no. of countries
Value domain: Work orientations				
Gesthuizen and Verbakel (2011)	Intrinsic and extrinsic job preferences	Human capital investments; risk-reducing socio-economic factors; quality of labor force and jobs	–	EVS 1999/2000; Eurostat; ILO
Esser and Lindh (2018)	Extrinsic and intrinsic job preferences; job autonomy	GDP; Gini; female labor force participation; family policy; social insurance generosity; labor market organization	–	ISSP 1989, 1997, 2005, 2015; OECD 2017
Pichler and Wallace (2009)	Job satisfaction	GDP; wage level; unemployment rate; unionization; inequality; average working time	–	European Quality of Life Survey; Eurostat New Cronos

also by personal religious beliefs, while denomination was included as a control variable, in just the same way as marital status, gender, age, and education. Their multilevel analyses show that the religious context matters, as do a person's age and religious beliefs and practices, at least in the case of moral issues not sanctioned by the state. Context and individual-level features do not have much explanatory power for moral issues which are sanctioned by legal codes.

Sieben and Halman (2015) also addressed public good morality. Their interest was the impact of the communist legacy on the justification of issues which are defined as illegal. They focused on post-socialist Europe, and found differences between individuals and countries in the degree to which behaviors that harm the public good are considered justified. They expected that the experience of repressive communist rule would induce lower levels of public good morality, which would be the case for people who had lived under communist rule for a prolonged period. This hypothesis is only partly confirmed, and, remarkably, the younger inhabitants of that part of Europe appear to be the most lenient on public morality issues. They found no support for the expectation that the process of democratization results in higher levels of public good morality and that individuals would be stricter as they would endorse democratic values more strongly.

Storm (2016) investigated the relationship between religiosity and moral values in a secularizing Europe longitudinally, using the data from the EVS waves in 1981, 1990, 1999, and 2008. Next to civic morality, or morality sanctioned by the state, she also addresses morality "that is not uniformly sanctioned by legal codes" (see Finke and Adamczyk 2008, p. 622) and for which religion often appears to be a better predictor than it is for state-sanctioned civic morality (Storm 2016, p. 113). She also formulated hypotheses about generational value change, the differential impact of countries' levels of religiosity on the two moral dimensions, and about the dependency of these moral orientations on weak or illegitimate governing institutions. The religious context, measured by the average level of religiosity in a country, matters, especially for religious people. The religious context matters, which supports the social network hypothesis of Putnam and Campbell (2010) "that the availability of religious fellow citizens increases the impact of religiosity on moral values" (Storm 2016, p. 133). The evidence that she found once again demonstrates the importance of context for explaining differences in moral views in Europe.

3.2 Religious Beliefs and Practices

Davie's (1994) striking characterization of the British religious situation in the early nineties echoed Durkheim (2001 [1912]) that religion is about beliefs and practices. She described Britain's religious situation as "believing without belonging," meaning that, as in many European countries, church attendance is on the decline, but religious beliefs (i. e., individuals' levels of religiosity) seem to persist.

Economic development is considered to be one of the leading causes of this situation. Inglehart (1997) contended that due to increasing levels of economic security, indicated by, for instance, GDP or related measures such as social security expenditures, people no longer need the church for salvation and reassurance (Inglehart 1997, p. 80; Norris and Inglehart 2004, p. 18). It does not imply that religiosity

as such is declining, but that the once dominant role of the churches is gradually diminishing.

That is also what was found by Hirschle (2013), who investigated the impact of economic growth on church attendance and religious beliefs. Based on the classic secularization idea that economic growth leads to disenchantment and to increasing levels of existential security, he expected a decline in religious values, resulting in turn in a drop in church attendance (Hirschle 2013, p. 412). The alternative explanation was that, with economic growth, people's needs are increasingly satisfied by secular goods and less by religious products. Hence, one can expect a decline in church attendance, leading in turn to a decline in religious beliefs, because the churches lose their socializing capacity. Using data from ISSP, EVS, WVS, ESS, and ALLBUS, he found that an increase in GDP is associated with declining church attendance, while religious values persist. When economic growth is considered as "a proxy for the expansion of a market for alternative goods that meet religious needs" (Hirschle 2013, p. 422), the alternative explanation is presumably substantiated.

Hirschle's contribution concerned the macro level only. Te Grotenhuis et al. (2015) considered both macro- and individual-level explanations. Like Hirschle, they argued that higher levels of wealth and security, both at the individual and the country level, will be conducive to lower levels of church attendance (2015, p. 644). Increasing rationalization, and hence higher levels of education, will also undermine traditional religious worldviews, both at the individual and the country level. Finally, when social ties are weak and heterogeneous (as is the case in urbanized areas), the normative pressure to conform to the environment is weaker and social control less severe. Analyzing the Mannheim Eurobarometer Trend File (GESIS Datenarchiv, Köln. ZA3521 Datenfile Version 2.0.1), they found that it was rising levels of GDP that reduce church attendance and not social security expenditures or rising levels of education and urbanization. This conclusion does not hold for individual levels of wealth and security. Hence, the context appears to matter more than the individual level of income, education, and urbanization.

The importance of the context is also demonstrated by Schwadel (2015), who elaborated on the effect of education on religiosity and found that national context mitigates the effect of education. Although he also argued that a higher level of schooling should negatively relate to religiosity at both individual and macro levels, he also expected that this association would differ across countries: Any negative effect will be strongest in (former) communist and religious countries, and weak(er) in highly educated countries. Analyzing ISSP 2008, he concluded that an individual's education has the predicted effect, although sex, marital status, and age appear to have stronger effects on religiosity (Schwadel 2015, p. 414). A national effect of education, which was "assessed with the mean of the university degree variable in each nation" (Schwadel 2015, p. 407) is not confirmed, but GDP, included as a control variable, appeared to fully mediate the association between nation-level education and individual religiosity. He also found support for the hypothesis that the schooling effect varies across countries, but he could not substantiate the hypothesis that the association will be strongest in (ex-)communist countries for which he created a dummy variable and in more highly-educated countries. He finds support,

however, for the hypothesis that the effect of education is strongest in more religious countries. Schwadel measured the level of religiosity in a nation with "the mean with no religious affiliation (...), standardized to adjust for the positive skew" (Schwadel 2015, p. 407). Once again, the results demonstrate that the context matters, and above all that the context may moderate the effect of individual characteristics, in this case the level of education. Also, the importance of GDP in this analysis underscores the idea that existential security is an important driver of secularization. Government regulation of religion, for which Schwadel relied on Grim and Finke's (2006, p. 7) Government Regulation Index assessing "restrictions placed on the practice, profession, or selection of religion by the official laws, policies, or administrative actions of the state," was also included as a control variable because it should affect religiosity negatively, but the effect was negligible.

Storm (2017) explored the effects of GDP and economic security on self-rated degrees of religiosity and tested several individual-level and macro-level hypotheses. In the same way as Hirschle, she distinguished attending religious services from religious beliefs, and expected that these would differentially associate with existential security issues. She hypothesized first that individual and national economic prosperity would negatively relate to religiosity. Second, she conjectured that social protection expenditures would negatively impact on the individual's religious participation because social protection is an alternative to religion in providing security. When religion is considered a buffer or substitute in stress situations, the church as a faith community may offer comfort and support, but one may also find comfort and meaning in religious beliefs and worldviews. Hence, Storm expected that religious people would evaluate their financial situation more positively than less religious or nonreligious people. Finally, she tested the well-known hypothesis that economic growth leads to a decline in religiosity. She used GDP for economic prosperity, and welfare expenditure was measured with the percentage of GDP spent on social benefits.

Storm's analysis of ESS data covering the years 2002–2014 showed that individual and macro levels of prosperity are indeed associated with lower levels of church attendance and self-rated religiosity, but that over time, changes in GDP do not explain the religious changes. Her religious substitution hypothesis was confirmed only for church attendance, but not for self-rated religiosity.

All in all, these studies show that distinguishing religious beliefs from practices makes sense, and that it is not only macro-level factors that determine levels of religiosity; religious orientations are also determined to a considerable extent by individual characteristics which studies on secularization and religious change should not ignore.

3.3 Political Values: Generalized and Social Trust

Classical political values refer to security, order, respect for authority, and conformity, whereas more contemporary values stress personal autonomy, independence, and emancipation (Van Deth 1995, p. 2; Inglehart 1997; Halman 2007). Some political values may turn out to be more persistent than others, and have not vanished. For example, in *The Civic Culture* (Almond and Verba 1965), several democratic

attitudes were distinguished that were already identified as important by de Toc-queville, and that are considered (again or still) highly relevant today. Such attitudes of trust, political partisanship, and societal involvement are vital concepts of what the sociological literature recognizes as social capital, a notion that has regained prominence since the works of, for instance, Putnam (1993) and Fukuyama (1995). Here, we will survey publications on interpersonal or social trust, which is considered a fundamental orientation for democracies. Numerous studies have addressed the impact of diversity on generalized or social trust, and we selected a few of them.

We know a lot about the greater trust that religious people have, but we know less about the effect of a religious setting on trust among both religious and nonreligious people. Olson and Li (2015) measured religious context not only as the percentage of the population that is religious, but also as the heterogeneity of religious groups among religious people measured by the well-known and commonly used Herfind-ahl index (Olson and Li 2015, p. 757). They expected that "because religious people are likely more engaged in trustworthy and trust-fostering interactions" (Olson and Li 2015, p. 758), the more people in a country are religious, the higher the level of trust will be. From religious heterogeneity theory (see Siegers' 2019 in this special issue), they predicted that a more diverse religious landscape emphasizes religious differences and creates boundaries between religious groups. Hence, levels of mis-trust will be lower in homogeneously religious societies. Their study is innovative in that they argued about the combined effect of these two context characteristics. The interaction hypothesis that they formulated was that an increase in religious heterogeneity makes the predicted positive effects of percent religious less positive. "Increases in the percent religious make the negative predicted effects of religious heterogeneity on trust even more negative" (Olson and Li 2015, p. 760). The study included individual-level variables such as church attendance and religious tradition as controls for compositional differences between countries.

The results showed that religious context indicators did not affect trust, but also that, when combined, there is a substantial effect as predicted in their hypothesis that: "Nations that are both highly religious and religiously heterogeneous have much lower levels of trust" (Olson and Li 2015, p. 769). This finding is all the more interesting since it contradicts what is usually found, and confirmed in Olson and Li's study, namely that religious people usually appear more trusting than nonreligious people. This result underscores the fact that individual-level effects can differ from macro-level effects, hence the need to include both levels in the analyses.

Also, Rapp (2016) included characteristics at both levels in testing the hypothesis that levels of trust are lower in more morally polarized societies. Her study captured opinion polarization in three ways: "dispersion, bimodality, as well as issue con-straint between our three moral issues of homosexuality, abortion, and euthanasia" (Rapp 2016, p. 37). Rapp based her hypothesis on the idea of perceived similar-ity: People tend to associate more, feel more comfortable with, and put more trust in those whom they perceive as similar. Rapp explores this idea for moral opin-ions, arguing that people trust each other less in morally more polarized societies. Analyses of the 2005–2008 WVS data confirmed the negative relationship between moral opinion polarization and social trust. Having said that, the individual-level characteristics also had significant effects on an individual's trust, demonstrating

the necessity to not only rely on macro-level analyses, but also to control for such compositional effects.

Diversity was the primary concern of Hooghe et al. (2009). They explored the impact of no fewer than 26 measures of diversity regarding immigration, ranging from the well-known and often applied fractionalization index taken from Alesina and La Ferrara (2002) to the inflow of foreign workers to asylum requests, for which they relied on OECD figures (Alesina and La Ferrara 2002, p. 219). Apart from such country characteristics, they also included individual-level variables as controls. Analyses of these control variables substantiated other studies, but the expected country-level effects were not confirmed. Their key hypothesis was based on the perceived threat thesis, and stated "that the population of the host society will be less trusting when it faces a rapid rise in the immigrant population over time and when the perceived cultural and religious distance or economic differences between immigrants and the majority group are larger" (Alesina and La Ferrara 2002, p. 204). This hypothesis found no support; hence the often-assumed negative impact of ethnic diversity on generalized trust does not hold in Europe (Alesina and La Ferrara 2002, p. 218).

To explore diversity's presumed impact on trust, Beugelsdijk and Klasing (2016) focused on the role of shared values or "conversely, the degree to which society is polarised along such values" (Beugelsdijk and Klasing 2016, p. 524). They calculated the degree of value polarization for 17 questions in EVS-WVS, and took the averages across these 17 scores for each country and wave to calculate the degree of value diversity (Beugelsdijk and Klasing 2016, p. 526). Using arguments from social identity theory, they verified the idea that value diversity negatively affects generalized trust. It is worth noting that they analyzed at three levels, with individuals being nested not only in countries, but also in regions. Their analyses confirm the expectation that value diversity, especially concerning government intervention in markets and income redistribution, induces lower levels of trust. "This relationship holds at various levels of aggregation: the country level, the sub-national (regional) level, and the individual level" (Beugelsdijk and Klasing 2016, p. 538).

These studies on generalized trust yield mixed results. Cultural diversity appears to affect trust negatively, whereas ethnic diversity does not seem to have much impact on trust. Perhaps the latter did not have much effect because the nation as the context may be too abstract and too distant to affect people in their daily lives; the region and its characteristics may come closer to people's experiences and determine their possibilities and ultimate choices. Beugelsdijk and Klasing show the usefulness of including the sub-national level in the analyses of social or generalized trust.

3.4 Family and Gender Division of Tasks

The domain of family life has changed fundamentally in recent years, and a concept such as family "is becoming more fluid and changeable" (Chambers 2012, p. 1). Demographic developments are seen as a consequence of progressing individualization (Lesthaeghe 2014), and changes in family life, primary relationships, and parenthood are regarded as expressions of the growing emphasis placed on personal autonomy, self-expression, and emancipation. The domain of family life includes

many issues ranging from personal relationships and family types to gender roles and gender equality, adulthood, and same-sex intimacies (see Hank et al. 2019 and Grunow 2019 in this special issue). This variety of topics appears for instance in the 2012 ISSP module "Family and Changing Gender Roles." Apart from attitudes and behaviors on female employment over the life cycle, the successive ISSP modules address attitudes towards marriage, the way a partnership organizes income, the gendered division of household chores, preferred and actual division of paid and unpaid labor, and alternative family forms (www.issp.org). Gender roles, gender equality, and combining work and family life are topical issues in modern societies and national and European policies. Equality between men and women is considered a fundamental human right and is defined as one of the Sustainable Development Goals by the United Nations. Despite progress, gender gaps appear to persist in Europe, albeit to different degrees. The empowerment of men and women has been uneven across regions and within countries (EU 2018; ESB 2017; see also Halman et al. 2011). This unevenness is also recognized by scholars who assume that differences in social, economic, and political contexts determine opportunities and constraints, and hence determine to no minor degree the choices that people make and the values to which they adhere on family- and gender-related issues.

Voicu and Constantin (2014) tried to explain European country differences in support of equal gender roles by modernization and institutional theories. In the same way as Inglehart (2018), they argued that modernization is only part of the explanation of differences in values among European populations, with a country's religious traditions and political heritage being major explanations of a country's trajectory. Their analyses of the 44 countries included in 2008 EVS confirmed that not only socio-economic development, but also main religious denomination, the rate of women's employment, and recent past are important for understanding differences in adherence to the traditional gender roles and attitudes towards gender equality on the labor market. Although not explicitly investigated in the article, some individual-level characteristics, such as religiosity, age, denomination, and gender, which were included as control variables, appear to be important attributes as well.

André et al. (2013) argued that using socialization theory, women would be more directly affected in their roles than men by family policies aiming at combining paid work and family care. As contextual factors, they included female labor market participation, years of parental leave that can be used by both partners together (both found in ILO data), and governmental childcare expenditures for which they included public expenditure on daycare and home help services as a percentage of GDP in 2000. Statistics came from OECD statistics 2012 (André et al. 2013, p. 460). The individual-level data came from 2002 ISSP data in 32 countries, and included women aged 40–60 years in paid work or who had worked in the past, having or having had frail elderly parents in the past ten years. All their individual-level hypotheses were confirmed, meaning that men are more traditional than women, higher-educated people are less traditional compared to the lower educated, employed people are less traditional, and the more children the respondent has, the more he or she supports traditional female roles. The study also found positive evidence for the adolescence hypothesis stating that when someone grows up in a situation where the mother was employed, he or she will be less supportive of the traditional female role. Their

analyses made it clear that subsidized childcare and enhancing female labor market participation are probably the best means to reduce the gender gap.

Naldini et al. (2016) also investigated female employment. They investigated the impact of the institutional and cultural context on women's decision to reduce working hours or quit their jobs to take care of their needy parents. They focused in particular on the effects of a country's care policies and (intergenerational) family care culture (Naldini et al. 2016, p. 609). Eurobarometer data from 21 European countries were analyzed, and their main conclusion was that when there is limited formal care in a country, women are more eager to reduce or quit their jobs to take care of their frail parent. More women remain in a paid job in countries where formal care is well organized and where societal norms concerning informal help are weak. Indicators for formal care policies were expenditure on care for the elderly, taken from Eurostat. Level of service coverage from the Multilinks database measures the percentage of over-65s receiving home care to the percentage of total public spending on long-term care; cash allowances were calculated as a percentage of total expenditure on long-term care (Naldini et al. 2016, pp. 612–615). Societal norms referred to familial care culture, which combined the proportions of individuals in each country agreeing with four items in the Eurobarometer survey (Naldini et al. 2016, p. 615). Their study revealed the importance of both policies and culture, but also that it is difficult to disentangle the specific influence of each (Naldini et al. 2016, p. 627). At the individual level, it turned out that occupational class, the presence of a partner, and type of caregiver play a role in determining whether women decide to reduce working hours or give up their work.

The advantages of multilevel analysis also appear in the study by Dotti Sani and Quaranta (2017), who investigated attitudes toward gender roles at three distinct levels of analysis. Next to country differences to explain differences in these attitudes, they expected gender differences to co-vary with the socio-economic background of the family of origin (Dotti Sani and Quaranta 2017, p. 31). Based on dependency theory, they expected that more egalitarian views would prevail in countries with higher levels of gender equality (Dotti Sani and Quaranta 2017, p. 32). From social learning theory, they derived hypotheses about gender differences, and the ideas of social diffusion of innovations theory led to hypotheses about the impact of the background of the family of origin and a cross-level interaction hypothesis about the differential impact of the family of origin for sons and daughters (Dotti Sani and Quaranta 2017, p. 33). The authors formulated two additional cross-level interactions about the differential impact of gender and maternal education in countries with higher or lower levels of gender equality (Dotti Sani and Quaranta 2017, p. 33). For gender equality, they included the Gender Inequality Index 2008 developed by UNDP, which "is a summary indicator accounting for gender-based disadvantages in reproductive health, empowerment, and the labor market" (Dotti Sani and Quaranta 2017, p. 34). The individual-level data comes from the International Civic and Citizenship Education Study 2009 (Schulz et al. 2010), containing data from eight-grade students in 36 countries. The analyses yielded evidence of their hypotheses, not only showing that the family of origin plays a determining role in views about gender equality, but also the countries' level of equality.

Overall, also in this domain, researchers increasingly make use of the opportunity to compare the attitudes and values between different societies and to study the impact of context characteristics. As seen above, there is evidence that differences in social, economic, and political contexts determine a person's opportunities and constraints, and hence the choices that people make and the values to which they adhere.

3.5 Intrinsic and Extrinsic Work Orientations and Job Satisfaction

The Meaning of Work (MOW) project that was carried out in the eighties revealed that work is a core aspect of the lives of large majorities of people in the industrialized world. After "family," it is mentioned as the second most important life domain, a finding that is substantiated by the EVS, which has been asking the same question since 1990 (Zanders 1994, p. 133; Halman et al. 2008, pp. 13–18).

People nevertheless have a variety of different reasons for regarding work as so profoundly important. One reason for people to work is to have income. As the MOW International Research Team (1987, p. 250) stated: "the dominant underlying reason why people work is to secure and maintain an income to purchase needed/or desired goods and services." However, people also engage in paid work for other reasons (see also Grunow and Erlinghagen in this volume). A distinction between extrinsic and intrinsic aspects of work has become widely accepted here. Extrinsic or material aspects refer to the economic rewards such as pay and job security, while intrinsic aspects emphasize the importance of autonomy and personal development (e.g., Jutz et al. 2018, p. 100).

Gesthuizen and Verbakel (2011) investigated whether differences between individuals and countries in extrinsic and intrinsic job preferences are attributable to socialization, economic deprivation, and job qualities. They predicted that people in nations that invest heavily in human capital (percentage of GDP spent on acquisition of human resources), that have low risks of economic deprivation (generous welfare states), and have a high-quality labor market (percentage working in first-digit ISO 88 groups 1, 2, and 3), combined with level of autonomy in a job (aggregated from EVS; Gesthuizen and Verbakel 2011, p. 673), will emphasize intrinsic job qualities more and extrinsic job qualities less. At the individual level, they argued that self-development is considered more important by more highly educated people, who therefore endorse intrinsic work qualities. People who suffer from economic deprivation presumably find extrinsic or material job features more important than intrinsic ones, and while those in low-quality jobs will seek satisfaction outside of work, people in high-quality jobs will presumably find satisfaction in work because it provides opportunities for self-development and autonomy, and therefore the latter will emphasize intrinsic job characteristics. Analyses of the EVS 2008 confirmed their hypotheses.

Esser and Lindh (2018) also analyzed both work orientations, using the work modules in ISSP. They explored the developments in the two work orientations between 1989 and 2015 in 19 Western countries. From modernization theory, they deduced hypotheses about the effects of economic development, increasing levels of education, and female labor participation, while the institutional theory about

welfare states led to hypotheses about the economic inequality within countries, social protection, and labor market regulations (Esser and Lindh 2018, p. 144). Their explanations of differences in work values are also (see Gesthuizen and Verbakel above) based on socialization theory, economic deprivation theory, and work quality. However, they also explored the effect of gendered socialization and a gendered labor market, "with women employed in the (public) care services promoting what is known as dare-rational motivations ..., which would be reflected in stronger intrinsic valuations among women" (Esser and Lindh 2018, p. 145).

At the country level, modernization theory predicts that higher levels of economic development are conducive to placing a stronger emphasis on intrinsic work aspects and weaker emphasis on extrinsic ones. Modernization and institutional theory predict that female workforce participation will induce stronger intrinsic work values and weaker extrinsic ones. Finally, institutional theory predicts that lower income inequality, more generous welfare provisions, gender equality policies, and regulated labor markets lead to stronger intrinsic work values and weaker extrinsic work values (Esser and Lindh 2018, p. 148). Neither contrasting perspective enjoys consistent support, but changes over time can be generally attributed to growing economic development and female labor market participation, thus underpinning the modernization perspective. The institutional perspective—stressing the importance of equality—was confirmed, but the limited effect of social policies was unexpected (Esser and Lindh 2018, p. 164). It was, however, found once more that both work values act in tandem and that both are more important in more unequal societies.

Work orientations are also significant predictors of job satisfaction. If a job has work characteristics that one considers to be important, it is likely that an individual is satisfied with the job. Indeed, "job satisfaction is regarded as the result of some perceived job characteristics including intrinsic and extrinsic rewards ..." which "are seen as strong factors in the experience of work, which heavily influence job satisfaction" (Pichler and Wallace 2009, pp. 536–537). According to Pichler and Wallace, job satisfaction depends on "the particular country a person lives in (...) because of the specific constellation of work, gender, social, and economic relations in a given context" (Pichler and Wallace 2009, p. 535). At the individual level, they investigated the impact of the type of occupation, supervision responsibilities, working hours, and extrinsic and intrinsic work characteristics. Such issues are known to affect job satisfaction, and since countries' composition of these characteristics may vary, differences in job satisfaction may (also) be a by-product of such individual differences (Pichler and Wallace 2009, p. 538). The primary hypotheses were about the effects of country-level institutional factors. They expected job satisfaction to be higher in countries with higher GDP, higher wage levels and shorter working hours, lower levels of unemployment, and higher levels of both equality and unionization (Pichler and Wallace 2009, pp. 538–539). A counter argument was that where unemployment is high, people can also be happy with their jobs because they have a job at all.

The data were from the European Quality of Life Survey in 28 European countries (European Foundation for the Improvement of Living and Working Conditions 2004). The results indicated that individual-level and compositional effects are the strongest predictors of job satisfaction. Objective working conditions (e. g., occu-

pational level, type of contract, job-related training) are important determinants of job satisfaction, but above all, both intrinsic and extrinsic job qualities promote job satisfaction. Institutional explanations work less well, but average wage differences mainly determine country differences in job satisfaction. The latter provides better explanations than the unemployment rate, the degree of unionization, and inequality (Pichler and Wallace 2009, p. 546).

Such studies signify the importance of both individual and macro levels when it comes to understanding varieties in work values. Equally importantly, they emphasize that alternative theoretical views should be developed to explain varieties in work orientations and job satisfaction.

4 New Directions

The increase in the number of international survey research projects investigating basic orientations makes it painfully clear that the grand theories in sociology fall short when it comes to explaining the often considerable differences between populations in various countries. Modernization theories do not tell the whole story, and cannot explain all the variety in these orientations. The central claim of modernization theory that socioeconomic development solely incurs value differences seems obsolete. It is about more than economy, and institutions, culture, history, and policies all help explain the sometimes major differences in fundamental human values.

So, modernization claims are insufficient; other theoretical notions must supplement them, of which institutionalism appears a good candidate. Our review of current state-of-the-art cross-national research activities, summarized in Table 1, makes it clear that many studies indeed focus on either modernization or institutionalism, or both. The studies use various theoretical perspectives to extract hypotheses about the impact of the context on values in a specific domain. They include very different contextual features to explain the varieties in the domain-specific value orientations, which proves that there are different mechanisms at work in the distinct value domains. The studies reveal that "context matters," but also that it is essential to include individual-level characteristics, at least as controls. Quite often, the individual attributes appear differently distributed in different countries, which may be the main reason why societies vary in certain value orientations. Multilevel analysis is the appropriate tool for separating such composition effects from true contextual effects.

Some of the reviewed studies (e.g., Storm 2016; Finke and Adamczyk 2008; Schwadel 2015; André et al. 2013; Naldini et al. 2016; Dotti Sani and Quaranta 2017; Pichler and Wallace 2009) demonstrated the fruitfulness of theorizing and testing hypotheses about cross-level interactions, arguing that individual-level effects depend on the contexts. Sociologists should be challenged to develop theories combining the macro level with the micro level and reflect on how the macro level may influence micro-level effects.

Two of the studies, namely Letki, and Beugelsdijk and Klasing, examined regional differences. Such studies are not only empirically but also theoretically vital, because

they draw our attention to the fact that various contextual levels offer the structural and institutional contexts that may shape people's values simultaneously. Both the national and regional level provide conditions and more or less stable frames of reference for behaviors and beliefs. National characteristics capture broader structural and institutional influences that will affect every citizen of a country, and in that sense, they are more distal factors influencing people's everyday actions and beliefs. People base their values on the cultural, political, and economic climate in society. Contextual effects resulting from characteristics at the regional level may be narrower in scope and impact, as they are somewhat more closely related to people's immediate surroundings and daily lives. What is happening at the micro level is nearly always embedded in institutionalized social networks at the meso level, such as markets, organizations, and communities (Arts 2011, p. 29). Fortunately, statistical agencies such as Eurostat are increasingly collecting statistics on regional structural characteristics, and these allow for regional analyses. Expanding the country-level studies to include regional-level characteristics as well would mean a first new direction in comparative sociology.

Many studies rely on comparing mean scores of countries or groups of people with specific characteristics to test hypotheses about varieties in values. However, as we have argued before, the same mean may be based on very different frequency distributions. Frequency distributions may be homogeneous or heterogeneous, and when investigating varieties across countries and groups, it can be fruitful to explore this within-country homogeneity/heterogeneity. This approach can also be interesting from a theoretical point of view because hypotheses about the distributions of orientations can be deduced from individualization and globalization ideas, which are considered critical processes of recent and contemporary social change. For example, individualization would mean that people's choices, actions, and orientations are no longer controlled by their social position, but are increasingly based on personal convictions and individual preferences. Apart from that, the individual in (post-)modern society faces a multitude of alternatives because of internationalization, transnationalization, and globalization. The world has become compressed, and the consciousness of the world has intensified tremendously (Robertson 1992, p. 8). Societies are currently interconnected (Genov 2015, p. 205), and people in this "global village" encounter many alternative cultural habits, lifestyles, and modes of conduct. Because individualized people have been liberated from (traditional) controls and constraints, and given that globalization implies that people can choose from the global cultural marketplace, the likelihood that people will choose the same will decline, and hence the heterogeneity of people's value preferences will increase. Value orientations are therefore not only likely to shift away from traditional views towards individualistic values and declining acceptance of traditional authorities, these orientations and opinions are also likely to become more diverse.

Of course, there may be many more reasons to expect certain populations and categories of people to be more homogeneous or heterogeneous. Theorizing about such explanations and empirically testing hypotheses about varying distributions would be another new direction in comparative sociology. Methods have already been developed for carrying out such an "inverted" multilevel analysis for group-level outcomes (Croon and van Veldhoven 2007).

The small number of studies presented here not only demonstrate that the combination of individual-, regional-, and country-level explanations offers greater explanatory power, but also that differences in orientations are determined (sometimes to a large extent) by differences in countries' composition of these individual characteristics. When studying the impact of macro- and meso-level characteristics on value orientations, the advice of Wil Arts (2011, p. 15) to "always control for the composition of the population by including individual-level variables" should be taken seriously.

References

Alesina, Alberto, and Eliana La Ferrara. 2002. Who trusts others? *Journal of Public Economics* 85:207–234.

Almond, Gabriel, and Sidney Verba. 1965. *The civic culture*. Boston: Little, Brown & Company.

André, Stéphanie, Maurice Gesthuizen and Peer Scheepers. 2013. Support for traditional female roles across 32 countries: Female labour market participation, policy models and gender differences. *Comparative Sociology* 12:447–476.

Arts, Wil. 2011. Explaining European value patterns: Problems and solutions. *Studia UBB Sociologica* 56:7–31.

Barnes, Samuel, and Max Kaase. 1979. *Political action: Mass participation in five western democracies*. Beverly Hills, CA: Sage.

Beugelsdijk, Sjoerd, and Mariko Klasing. 2016. Diversity and trust: The role of shared values. *Journal of Comparative Economics* 44:522–540.

Chambers, Deborah. 2012. *A sociology of family life*. Cambridge, Malden: Polity Press.

Coleman, James. 1990. *Foundations of social theory*. Cambridge, MA: Belknap Press of Harvard University Press.

Croon, Marcel, and Marc van Veldhoven. 2007. Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods* 12:45–57.

Davie, Grace. 1994. *Religion in Britain since 1945.*Oxford: Blackwell.

De Moor, Ruud. 1994. Epilogue. In *The individualizing society,* eds. Peter Ester, Loek Halman and Ruud de Moor, 229–232. Tilburg: Tilburg University Press.

De Moor, Ruud. 1987. Religieuze en morele waarden [Religious and moral values]. In *Traditie, secularisatie en individualisering*, eds. Loek Halman, Felix Heunks, Ruud de Moor and Harry Zanders, 15–49. Tilburg: Tilburg University Press.

Diez-Roux, Ana. 2002. A glossary for multilevel analysis. *Journal of Epidemiology and Community Health* 56:588–594.

Dotti Sani, Giulia, and Mario Quaranta. 2017. The best is yet to come? Attitudes toward gender roles among adolescents in 36 countries. *Sex Roles* 77:30–45.

Durkheim, Émile. 1964. *The division of labour in society*. 1893. New York: Free Press

Durkheim, Émile. 2001. *The elementary forms of religious life.1912.* Oxford: Oxford University Press.

Eisenstadt, Shmuel. 2002. Some observations on multiple modernities. In *Reflections on Multiple Modernities*, eds. Dominic Sachsenmaier, Jens Riedel and Shmuel Eisenstadt, 27–41. Leiden, Boston, Cologne: Brill.

Esser, Ingrid, and Arvid Lindh. 2018. Job preferences in comparative perspective 1989–2015: A multidimensional evaluation of individual and contextual influences. *International Journal of Sociology* 48:142–169.

European Commission. 2017. *Special eurobarometer 465 gender equality 2017*. Brussels: European Union.

European Commission. 2018. *Report on equality between women and men in the EU*. Luxembourg: Publications Office of the European Union.

European Foundation for the Improvement of Living and Working Conditions. 2004. *Quality of Life in Europe. First European Quality of Life Survey 2003*. Luxembourg: Office for Official Publications of the

European Communities, available from: http://www.eurofound.eu.int/publications/files/EF04105EN.pdf.

Finke, Roger, and Amy Adamczyk. 2008. Cross-national moral beliefs: The influence of national religious context. *The Sociological Quarterly* 49:617–652.

Fox, John. 2016. *Applied regression analysis & Generalized linear models*. Third Edition. Thousand Oaks, CA: Sage Publications.

Fukuyama, Francis. 1995. *Trust. The social virtues and the creation of prosperity*. London: Penguin Books.

Genov, Nikolai. 2015. Challenges of Individualisation. *International Social Science Journal* 64:197–209.

Gesthuizen, Maurice, and Ellen Verbakel. 2011. Job preferences in Europe: Tests for equivalence and explanations for cross-national variation. *European Societies* 13:663–686.

Grim, Brian, and Roger Finke. 2006. International religion indexes: Government regulation, government favoritism, and social regulation of religion. *Interdisciplinary Journal of Research on Religion* 2:1–40.

Grunow, Daniela. 2019. Comparative analyses of housework and its relation to paid work: Institutional contexts and individual agency. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Gundelach, Peter. 1994. National value differences. Modernization or Institutionalization? *International Journal of Comparative Sociology* 35:37–58.

Halman, Loek. 1991. *Waarden in de Westerse wereld* [Values in the Western world]. Tilburg: Tilburg University Press.

Halman, Loek. 2007. Political values. In *The Oxford handbook of political behaviour,* eds. Russell Dalton and Hans-Dieter Klingemann, 305–322. Oxford: Oxford University Press.

Halman, Loek, and Ruud de Moor. 1987. Waarden: een inleiding [Values: an introduction]. In *Traditie, secularisatie en individualisering*, eds. Loek Halman, Felix Heunks, Ruud de Moor and Harry Zanders, 1–14. Tilburg: Tilburg University Press.

Halman, Loek, and Ruud de Moor. 1994. Comparative research on values. In *The Individualizing Society*, eds. Peter Ester, Loek Halman and Ruud de Moor, 21–36. Tilburg: Tilburg University Press.

Halman, Loek, and Astrid Vloet. 1994. *Measuring and comparing values in 16 countries of the western world in 1990 and 1981*. Tilburg: WORC.

Halman, Loek, Ronald Inglehart, Jaime Díez-Medrano, Ruud Luijkx, Alejandro Moreno and Miguel Basáñez. 2008. *Changing values and beliefs in 85 countries*. Leiden: Brill.

Halman, Loek, Inge Sieben and Marga van Zundert. 2011. *Atlas of European values. Trends and traditions at the turn of the century*. Leiden: Brill.

Hank, Karsten, and Anja Steinbach. 2019. Families and their institutional contexts: The role of family policies and legal regulations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Hirschle, Jochen. 2013. "Secularization of Consciousness" or alternative opportunities? The impact of economic growth on religious belief and practice in 13 European countries. *Journal for the Scientific Study of Religion* 52:410–424.

Hooghe, Marc, Tim Reeskens, Dietlind Stolle and Ann Trappers. 2009. Ethnic diversity and generalized trust in Europe. *Comparative Political Studies* 42:198–223.

Hox, Joop, Mirjam Moerbeek and Rens van de Schoot. 2017. *Multilevel analysis: Techniques and applications. Third Edition*. London & New York: Routledge.

Inglehart, Ronald 1977. *The silent revolution*. Princeton, NJ: Princeton University Press.

Inglehart, Ronald 1997. *Modernisation and postmodernisation*. Princeton, NJ: Princeton University Press.

Inglehart, Ronald 2018. Modernisation, Existential security and cultural change: Reshaping human motivations and society. In *Handbook of advances in culture and psychology*, volume 7, eds. Michele Gelfand, Chi-yue Chiu and Ying-yi Hong. Oxford: Oxford University Press.

Inglehart, Ronald, and Christian Welzel. 2005. *Modernisation, cultural change and democracy*. Cambridge: Cambridge University Press.

Jutz, Regina, Evi Scholz, Michael Braun and Markus Hadler. 2018. The ISSP 2015 Work Orientations IV Module. *International Journal of Sociology* 48:95–102.

Kerkhofs, Jan. 1997. *De Europeanen en hun waarden* [Europeans and their values]. Leuven: Davidsfonds.

Kroneberg, Clemens. 2019. Theory development in comparative social research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für*

Soziologie und Sozialpsychologie. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Lerner, Daniel. 1968. Modernisation. In *International encyclopedia of the social sciences*, eds. David Sills and Robert Merton, 386–395. New York: Macmillan.

Lesthaeghe, Ron 2014. The second demographic transition: A concise overview of its development. *PNAS* 111:18112–18115.

Letki, Natalia. 2006. Investigating the roots of civic morality: Trust, social capital, and institutional performance. *Political Behaviour* 28:305–325.

Marsh, Robert. 2014. Modernisation theory, then and now. *Comparative Sociology* 13:261–283.

Mead, George Herbert. 1934. *Mind, self, and society: From the standpoint of a social behaviourist.* Chicago: University of Chicago Press.

Meulemann, Heiner 1983. Value change in West Germany, 1950–1980: Integrating the empirical evidence. *Social Science Information* 22:777–800.

MOW International Research Team. 1987. *The meaning of working.* New York: Academic Press.

Naldini, Manuela, Emmanuele Pavolini and Christina Solera. 2016. Female employment and elderly care: The role of care policies and culture in 21 European countries. *Work, Employment and Society* 30:607–630.

Norris, Pippa, and Ronald Inglehart. 2004. *Sacred and secular*. Cambridge: Cambridge University Press.

Olson, Daniel, and Miao Li. 2015. Does a nation's religious composition affect generalized trust? The role of religious heterogeneity and the percent religious .*Journal for the Scientific Study of Religion 54*:756–773.

Phillips, David, and Stephen Harding. 1985. The structure of moral values. In *Values and social change in Britain*, eds. Mark Abrams, David Gerard and Noel Timms, 93–108. London: MacMillan.

Pichler, Florian, and Claire Wallace. 2009. What are the reasons for differences in job satisfaction across Europe? Individual, Compositional, and Institutional Explanations. *European Sociological Review* 25:535–549.

Preyer, Gerhard, and Michael Sussman. 2016. Introduction on Shmuel N. Eisenstadt's sociology: The path to multiple modernities. In *Varieties of multiple modernities. New Research Design*, eds. Gerhard Preyer and Michael Sussman, 1–29. Leiden, Boston: Brill.

Przeworski, Adam, and Henry Teune. 1970. *The logic of comparative social inquiry.* New York: John Wiley & Sons.

Putnam, Robert. 1993. *Making democracy work*. Princeton, NJ: Princeton University Press.

Putnam, Robert, and David Campbell. 2010. *American grace: How religion divides and Unites Us*. New York: Simon and Schuster.

Rapp, Carolin. 2016. Moral opinion polarization and the erosion of trust. *Social Science Research* 58:34–45.

Robertson, Roland 1992. *Globalization.* London: Sage

Robinson, William 1950. Ecological correlations and the behaviour of individuals. *American Sociological Review* 15:351–357.

Sachsenmaier, Dominic. 2002. Multiple modernies. The concept and its potential. In *Reflections on multiple modernities,* eds. Dominic Sachsenmaier, Jens Riedel and Shmuel Eisenstadt, 42–67. Leiden, Boston, Cologne: Brill.

Schmidt, Volker. 2010. Modernity and diversity: Reflections on the controversy between modernisation theory and multiple modernists. *Social Science Information* 49:511–538.

Schulz, Wolfram, John Ainley, Julian Fraillon, David Kerr, and Bruno Losito. 2010. *ICCS 2009 International report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Schwadel, Philip. 2015. Explaining cross-national variation in the effect of higher education on religiosity. *Journal for the Scientific Study of Religion* 54:402–418.

Sieben, Inge, and Loek Halman. 2015. Morality on the public good in post-socialist European states. *Studies of Transition States and Societies* 7:7–26.

Siegers, Pascal. 2019. Is the influence of religiosity on attitudes an behaviors stronger in less religious or more religious societies? A review of theories ana contradictory. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Smelser, Neil 1973. Toward a theory of modernisation. In *Social change, sources, patterns, and consequences*, eds. Eva Etzioni-Halevy and Amitai Etzioni, 268–284. New York: Basic Books.

Stoetzel, Jean. 1983. *Les Valeurs du temps présent: Une enquête Européenne*. [Today's values. A European survey] Paris: Presses Universitaires de France.

Storm, Ingrid. 2016. Morality in context: A multilevel analysis of the relationship between religion and values in Europe. *Politics and Religion* 9:111–138.

Storm, Ingrid. 2017. Does economic insecurity predict religiosity? Evidence from the European social survey 2002–2014. *Sociology of Religion: A Quarterly Review* 78:146–172.

Te Grotenhuis, Manfred, Marijn Scholte, Nan-Dirk de Graaf and Ben Pelzer. 2015. The between and within effects of social security on church attendance in Europe 1980–1998: The danger of testing hypotheses cross-nationally. *European Sociological Review* 31:643–654.

Van Deth, Jan. 1995. Introduction: The impact of values. In *The impact of values,* eds. Jan van Deth and Elinor Scarbrough, 1–18. Oxford: Oxford University Press.

Van Deth, Jan, and Kent Jennings. 1989. Introduction. In *Continuities in political action* eds. Jan van Deth and Kent Jennings, 3–21. Berlin, New York: Walter de Gruyter.

Van Deth, Jan, and Elinor Scarbrough. 1995. *The impact of values*. Oxford: Oxford University Press.

Voicu, Mălina, and Andreea Constantin. 2014. Attitudes towards gender roles in Europe: Modernisation and social institutions. In *Value contrasts and consensus in present-day Europe,* eds. Wil Arts and Loek Halman, 311–328. Leiden, Boston: Brill.

Zanders, Harry. 1994. Changing work values. In *The individualizing society*, eds. Peter Ester, Loek Halman and Ruud de Moor, 129–154. Tilburg: Tilburg University Press.

Loek Halman 1956, PhD., Associate Professor at Tilburg University. Areas of research: Values, comparative studies, sociology. Publications: The Atlas of European Values, Leiden, 2011 (with I. Sieben and M. van Zundert); Religion and Civil Society, Dordrecht, 2013 (ed. with J. de Hart and P. Dekker); Value Contrasts and Consensus in Present-Day Europe, Leiden (ed. with W. Arts).

John Gelissen 1972, PhD., Endowed Professor at Tilburg University. Areas of research: Sociological and methodological research on cross-cultural differences in public opinion. Publications: Testing Measurement Equivalence of Experienced Holiday Quality: Evidence on Built-in Bias in the Flash Eurobarometer survey data, 2018; Clusters of cultures: Diversity in meaning of family value and gender role items across Europe, 2017 (with E. van Vlimmeren and G. Moors); Exploring the response process of culturally differing survey respondents with a response style: A sequential mixed methods study, 2013 (with M. Morren and J.K. Vermunt).

Kölner Zeitschrift
für Soziologie und
Sozialpsychologie

Check for
updates

Internationally Comparative Research Designs in the Social Sciences: Fundamental Issues, Case Selection Logics, and Research Limitations

Achim Goerres · Markus B. Siewert · Claudius Wagemann

Abstract This paper synthesizes methodological knowledge derived from comparative survey research and comparative politics and aims to enable researches to make prudent research decisions. Starting from the data structure that can occur in international comparisons at different levels, it suggests basic definitions for cases and contexts, i. e. the main ingredients of international comparison. The paper then goes on to discuss the full variety of case selection strategies in order to highlight their relative advantages and disadvantages. Finally, it presents the limitations of internationally comparative social science research. Overall, the paper suggests that comparative research designs must be crafted cautiously, with careful regard to a variety of issues, and emphasizes the idea that there can be no one-fits-all solution.

Keywords International comparison · Comparative designs · Quantitative and qualitative comparisons · Case selection

A. Goerres (✉)
Fakultät für Gesellschaftswissenschaften, Institut für Politikwissenschaft, Universität Duisburg-Essen
Lotharstr. 65, 47057 Duisburg, Germany
E-Mail: achim.goerres@uni-due.de

M. B. Siewert · C. Wagemann
Fachbereich Gesellschaftswissenschaften, Institut für Politikwissenschaft, Goethe-Universität Frankfurt
Theodor-W.-Adorno Platz 6, 60323 Frankfurt am Main, Germany

M. B. Siewert
E-Mail: siewert@soz.uni-frankfurt.de

C. Wagemann
E-Mail: wagemann@soz.uni-frankfurt.de

🖄 Springer

International vergleichende Forschungsdesigns in den Sozialwissenschaften: Grundlagen, Fallauswahlstrategien und Grenzen

Zusammenfassung Dieser Beitrag bietet eine Synopse zentraler methodischer Aspekte der vergleichenden Politikwissenschaft und Umfrageforschung und zielt darauf ab, Sozialwissenschaftler zu reflektierten forschungspraktischen Entscheidungen zu befähigen. Ausgehend von der Datenstruktur, die bei internationalen Vergleichen auf verschiedenen Ebenen vorzufinden ist, werden grundsätzliche Definitionen für Fälle und Kontexte, d. h. die zentralen Bestandteile des internationalen Vergleichs, vorgestellt. Anschließend wird die gesamte Bandbreite an Strategien zur Fallauswahl diskutiert, wobei auf ihre jeweiligen Vor- und Nachteile eingegangen wird. Im letzten Teil werden die Grenzen international vergleichender Forschung in den Sozialwissenschaften dargelegt. Der Beitrag plädiert für ein umsichtiges Design vergleichender Forschung, welches einer Vielzahl von Aspekten Rechnung trägt; dabei wird ausdrücklich betont, dass es keine Universallösung gibt.

Schlüsselwörter Internationaler Vergleich · Vergleichende Studiendesigns · Quantitativer und qualitativer Vergleich · Fallauswahl

1 Introduction

This article deals with the challenges and pitfalls that researchers frequently have to face when engaging in cross-national comparative analyses. Such a discussion is not an easy task. Both methodologists and practitioners conducting cross-national analyses at the *macro* level use different terminologies and emphasize different criteria of comparison than their colleagues who work at the *individual* level. This is complicated even further by similar communication deficits across qualitative and quantitative methods (Brady and Collier 2004, 2010; Goertz and Mahoney 2012; King et al. 1994). Against this backdrop, we seek to inform a heterogeneous readership about the terminology and various strands of argumentation, as well as of potentials and pitfalls related to carrying out cross-case international comparisons.

We take a pluralistic stance on methods by bringing together insights from various strands of methodological schools of thought on how to design and conduct comparative research. We hence present a concise summary concerning the state-of-the-art of doing comparisons in the social sciences, but most certainly do not seek to propose a specific recipe for how to carry out cross-country comparisons, or multilevel research. This article is a cookbook with many recipes fitting different occasions rather than just one recipe. This also means that we do not take sides on the methodological debates or propose a fixed set of rules in terms of what comparative research should look like. Instead, we would rather start from the assumptions (i) that the application of methods should be question driven (Shapiro 2002), (ii) that a research design can, and even must, undergo necessary adjustments during the course of research (Schmitter 2008), (iii) and that, at the end of the day, every researcher should be her/his own methodologist (Wright Mills 1959, p. 224). The overall goal of the article is, therefore, to provide orientation about the state

of important debates and discussions in the field of comparative research without ascribing a higher value to one specific approach.

Our focus lies on *international* comparisons. Not every comparison necessarily has to be internationally oriented, since we can also compare city structures, parties, social movements, government action, etc., within a single country, using similar lines of logic. However, cross-country comparisons usually show certain complications, compared to an otherwise similar mono-country project: concepts need to be applicable across different cases; analytically important differences need to exist across cases to be explained; practical problems can come to the fore, such as planning fieldwork in a foreign country or experiencing a language barrier. In short, the cross-national perspective poses challenges and pitfalls which are different from comparisons within the same country context (see Snyder 2001 on the issue of subnational comparison). This means for our purpose that we deal with comparison as such, but always with an eye to the specific challenges for *international* comparison.

The article is structured as follows: first, we place the internationally comparative design into a broader methodological perspective, discuss different data structures, and then elaborate what they mean for a project, before defining cases and contexts as the basic concepts. Second, we give a comprehensive overview with guidelines on different selection strategies for international cases. Third, we discuss the limitations of the internationally comparative design before, fourth, concluding the paper.

2 The Basics of Comparative Analysis: Cases, Contexts, and Data Structure

2.1 Comparative Research in the Social Sciences

The etymological origin of the word "comparison" comes from Latin and points to the identification of similarities and differences, shaping the labels of scientific subdisciplines such as comparative macro-sociology or comparative politics (Goldthorpe 1997; Powell et al. 2014). At the same time, the term has also had a methodological career, most famously through Arend Lijphart's (1971) seminal article on the "Comparative Method", which seemed to identify a whole study field with a method—or, as we would say, a design. However, reading Lijphart carefully, one detects a clear rank order of methodological approaches that still holds today (see Lijphart 1971, p. 684 et seqq.). First and foremost, the experimental study continues to be the gold standard due to the possibility for the researcher to manipulate the values of the independent variable while controlling for possible moderating factors. Lijphart defines the "statistical method" as the weaker variant of the experiment, keeping in place at least one of the central principles of experiments, namely to select cases randomly. Finally, the "comparative method" is presented as the weakest variant and "a very imperfect substitute" (Lijphart 1971, p. 685) of experimental and statistical methods. It is notable at this point that Lijphart identifies the comparative method with a small-*N* analysis, i.e. an analysis of just a few cases. This then subsequently implies the main limitation: "The number of cases it

deals with is too small to permit systematic control by means of partial correlations" (Lijphart 1971, p. 684).

Comparative research designs are hence not free from criticism. If we compare countries, the number of available cases is often not only limited for the desired sample, but also for the theoretical reference population. Applying specific theoretical lenses creates research situations where only a limited range of countries are available—the often labeled "theories of the middle range" (Merton 1957) from a perspective of research design. But when we study, for instance, industrialized advanced economies or countries' responses to natural disasters, we usually end up with numbers which do not allow for the application of standard statistical techniques, given that since basic assumptions, such as questions of distribution, unit homogeneity, or causal independence (see also King et al. 1994), are not met.

Quantitative researchers are quick to worry about an indeterminate research design when comparing countries, i.e. that there are more variable constellations than observations. This perspective reflects one of the reasons for not trying to engage in international comparisons since the luxury of having enough observations at the international level is rarely found in the available data. Apart from having such a rare abundance of international data, only the quasi-experimental design is not subject to the problem since it is based on an ex post construction of artificial treatment and control groups of international cases (see below).

One proposed way to circumvent this is to engage in small-*N* comparisons with only two, four, or a few more cases under observation (Mahoney 2003; Skocpol and Somers 1980). Others also subsume longitudinal designs within a case over time (often marked through historical ruptures and embedded in temporal sequences) as a "comparison" (Gerring 2007, p. 28). More recent techniques such as Qualitative Comparative Analysis (QCA) even allow one to work on designs focusing on a mid-sized number of cases through the use of set-theoretic relations (Ragin 2008; Schneider and Wagemann 2012). What all these proposals have in common is that they, first, do not reach the case numbers which are typical for most surveys and other quantitative approaches, can therefore, second, not rely on probabilistic approaches or techniques which are based on randomization, and are, third, accused of not meeting the standards for scientific inference which are typical of quantitative approaches (see e.g. Brady and Collier 2004, 2010; Goertz and Mahoney 2012).

When speaking about comparative research, we thus quickly touch upon the debates between qualitative and quantitative methods, or more specifically between macro-level comparativists using (comparative) case study logics *versus* quantitative researchers who apply the large-*N* logic of individual analysis to the country level (see e.g. Brady and Collier 2010; Collier 2014; della Porta and Keating 2008; Goertz and Mahoney 2012; Mahoney and Goertz 2006; Ragin 2004). This bifurcation within the methodological world has however engendered various strands of literature that are virtually or even completely isolated from each other. Just think of the proposals from the (comparative) case study design literature (Blatter and Haverland 2012; Gerring 2007; Ragin 2008; Rohlfing 2012), or the methodological pieces about complex survey studies with international survey data (Steenbergen and Jones 2002), which ignore each other to put it mildly. However, both approaches are intended for comparisons at the international level.
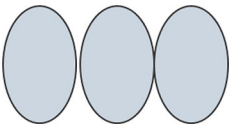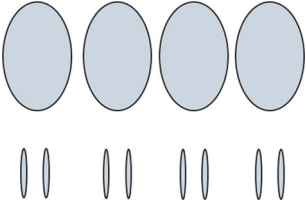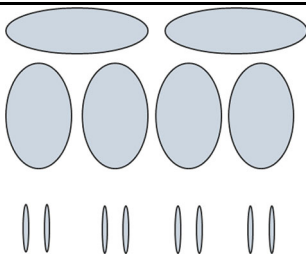
2.2 International Comparison at Different Levels of the Data Structure

Let us start by locating where the international component can be found in the data structure. The simplest data structure in terms of international comparison is non-hierarchical, as is illustrated in scenario I in Tab. 1. Only one level of variance exists here, namely, country cases. Researchers focus completely on one level of comparison, and only strive to make statements at one level of international analysis. Such an international analysis would not be considered to warrant any kind of multilevel modelling strategy due to the nonhierarchical nature of the data. A prominent example of such a design can be found in a volume edited by Robert Dahl (1966), which entails contributions comparing political oppositions in Western democracies without any further hierarchy in the data. Country-level national oppositions in a given region (here: Western Europe) are considered equally, without any reference to levels above (such as supranational regions) or below (such as subnational oppositions).

Once we have hierarchical data—i.e. a data structure with different levels of aggregation—international comparisons can be a part of the overall design which targets different levels. Scenario II in Tab. 1 depicts this situation where the international comparison is at the highest level of aggregation, with the actual units of analysis being nested in country contexts. A typical design in this respect is internationally comparative survey studies, where individuals are the units of analysis, and the contexts in which individuals are embedded are subject to an international comparison. An example of such a data structure can be found in Achim Goerres and Markus Tepe (2010), who examine in which country contexts older people are supportive of state structures providing public childcare. In this study, individuals as cases are grouped in country contexts that stand for different political, socioeconomic and cultural characteristics with regard to both societal and political aspects. Based on the analysis of surveys in twelve countries, the authors then identify direct as well as moderating effects from the macro towards the micro level. In such a data scenario, researchers must take at least two decisions for case selection (see also below): one for the country comparison at the top and one for the units of analysis within the country contexts. Researchers also strive to make analytical statements about the meaning of the international variance for the unit of analysis (macro–micro

Table 1 Three forms of international comparison in the data structure (authors own work)

Scenario I: No hierarchy	Scenario II: Hierarchical data, international comparison at the highest level of aggregation	Scenario III: Hierarchical data, international comparison at more than one level in the data structure

effects) and about the contextualization (or moderation) of subnational effects (here micro-level effects) through the macro-level effects.

Scenario III in Tab. 1, finally, illustrates a data structure in which the international comparison comes in on several layers. The work carried out by Gary Marks et al. (2006) is a prime example of such a multilayered design. Investigating patterns of national party competition across Eastern and Western Europe, their cases are political parties that are nested in countries, while the countries again are nested in the country groups of Eastern and Western Europe with different historical traditions. Country cases are thus combined further in analytical groups. In the example, the selection of countries is justified with reference to the additional country groups that are relevant for the project. The two upper levels of international comparison are integrated with one another. It is possible to have one logic for comparison for the supranational country group and a second logic for the actual country cases, and further ones for the subnational units. Researchers thus have at least three opportunities for selection and can make inferences about the impact of the supranational region on their subnational unit of analysis—here: parties—of the national context on the subnational, of the supranational on the national, and of all these causal arrows as contextualizing factors in moderation analysis.

Researchers who are confronted with the question of how to define their research design will have an easy choice between scenario I on the one hand or scenarios II or III on the other. Scenario I does not entail any interest in subnational variation, thus making the data structure and design decisions on international aspects less complicated. If researchers are not interested in scenario I, they can thus choose between scenarios II and III. They should opt for scenario III if the number of country cases is sufficiently large to warrant further grouping in country groups, and if they have theoretical reasons to argue for a supranational layer of causal dynamics.

Only scenarios II and III allow the modelling of causal relationships between different levels of aggregation. There are many effects on individuals from the country contexts, and individuals as a whole can influence the country context. There are also macro–macro causal relationships, such as the diffusion of environmental problems across states and its subsequent influence on individuals.

2.3 Cases

We have already used a key term, namely "case", that we need to define properly. The terminological clarification of what a case is starts with a confusion: if we compare internationally, it seems quite clear that a country constitutes a case. However, other terms are often used synonymously, such as "unit of analysis" or "unit of observation" (for some examples, see Gerring 2007, p. 17; 19 et seqq.; Seawright and Collier 2010a, p. 315, p. 357), even though their meaning is not always unambiguous. In order to be more illustrative, one could say that the discussion of what a case is can be abbreviated as the need to describe the entities which define the rows in a spreadsheet. In an international comparison, cases are most prominently identical to countries and other geographical entities, but also to societies, markets, organizations (e.g. political parties, unions, businesses, schools), events (e.g. wars, natural disasters, scandals), processes (democratization, deprivation, mobilizations,

radicalization), etc. Depending on the level at which we operate, even individuals between whom we might want to further differentiate, e. g. according to their various life phases, gender, etc., can qualify as cases.

The discussion of what a case actually is becomes relevant due to two important implications: first, the definition of what constitutes a case also comprises the question of what it is a case *of,* i.e. to which reference population it can be attributed (Collier and Mahoney 1996, p. 4, 38; Ragin 2000, p. 43 et seqq., 2004). It is indispensable to render the reference population explicit, since inferences can only be made to that reference population, if at all. In contrast to standard statistical techniques, the problem of comparative research operating with small or midsized case numbers is not only (or perhaps not even so much) the question of case selection or sampling, but that of the researcher carefully defining the population (Mahoney and Goertz 2004).

Second, and connected to the first issue, is the discussion of "casing" (Ragin and Becker 1992; see also Rohlfing 2012, p. 23–28). The issue of what constitutes a case is usually not naturally given, but rather needs creative construction on the part of the researcher. For instance, while country borders might lend themselves as natural identifiers of countries as cases, the endeavor is made more difficult if the units of observation are organizations. The more formalized organizational structures are, the better defined are their borders, and the easier it is to define it as a case. However, when comparing, for instance, organizational fields in a given economic sector, the establishment of where the field starts and where it ends is anything but trivial. The same holds, for instance, for the social movement organizations which are characterized by fluid structures and memberships. Before comparing social movements, scholars therefore have to define what a movement is. Note that, although the definition of "country cases" seems to be clearcut, the problem of casing can also occur at the country level. While just three years separate Germany in 1988 from Germany in 1991—which is the same time distance as between 1978 and 1981—only few would suggest that Germany after the fall of the Berlin Wall constitutes the same case as preunification Germany. This change was certainly also accompanied by territorial changes (growth) and a new legal situation (full sovereignty), which might also have led to a different country in structural terms (despite the continuation of the Basic Law [*Grundgesetz*] and the main institutional structures). Examples of such temporal "before–after" gaps constituting new cases abound in the social sciences—another illustrative examples is the world pre-9/11 and post-9/11.

"Casing", however, draws our attention to a further special asset of "defining what a case is", going back to the fact that cases can also be seen as configurations of their properties—a perspective which is largely inspired by the works of Charles Ragin (2000, p. 64 et seqq.), but also by Paul Lazarsfeld's (1937) idea of a property space. Depending on the actual research question, different aspects of cases might be analytically important. Studying the United States of America from the perspective of the migration research means that the researcher understands the U.S. case differently, as if (s)he were studying religious pluralism, resistance to welfare reform, executive politics, or the polarization of politics. The more vaguely defined a case is (i.e. the less clear its borders are), the more room for "casing"

opens up—think about such creative concepts as the idea of a "European society". While vaguely defined concepts have the advantage of allotting considerable scope to the individual researchers' decisions with regard to casing, they usually come at the price of ambiguous conceptual definitions (Collier and Adcock 1999; Goertz 2006; Sartori 1970).

In fact, cases can only be compared if they share at least enough characteristics in order to belong to the same group of research objects. While Germany is a country, San Francisco is a city, which means that Germany and San Francisco should not be compared if this fundamental difference in territorial constitution is relevant for the research interest. If we compare, for instance, Liechtenstein and Würselen, the city in which the 2017 SPD candidate for Chancellor, Martin Schulz, was Mayor before starting his EU career, we will see that both territories have a more or less similar number of inhabitants, varying between 35,000 and 40,000. If we are only interested in structures of social networks in communities of that size, then the two settings might be comparable, but otherwise not. We can see that this again takes up the issue of casing from above: the comparability strongly depends on the properties at which we look when we execute the comparison. Liechtenstein and Würselen might not be comparable in many respects, but they are comparable in terms of the population size. So, if the population size is a decisive category, and if we can make convincing arguments that all the other differences between the two cases do not influence or are not relevant to the phenomena that we want to study (something which is hard to imagine for this example), then a comparison of these two units can make sense.

Casing is thus closely connected to the idea of case properties. Comparability is ensured through a configuration of case properties in which some properties are held constant in such a way that they form a *species* (in the sense of a higher-order concept), while other properties are defined as being irrelevant. If we understand every case as a configuration of its properties, then comparability is ensured by having sufficient subsets of shared properties. The old idea of *genus et differentiam*, which is used for defining concepts, comes back in here: while much has to be equal, or at least sufficiently similar, between two (or more) cases so that the same *genus* can be ascertained, other properties must be different so as not to compare two equal cases.

Note that cases in international comparison are more often than not dependent on one another, and arguably increasingly so. This certainly also has repercussions on questions of inference which will be addressed later. Indeed, the *in*dependence of country cases should not be taken for granted and is difficult to achieve in our current times of international exchanges of knowledge and experiences—an issue which is usually referred to as Galton's problem. For instance, the spatial dependency of countries can lead to the diffusion of policy ideas that can be traced through international policy diplomacy, i.e. policy experts travel to the neighboring country to learn about public policy issues and can then try to implement their insights back home (Simmons and Elkins 2004). Another example refers to the Arab Spring, which was strongly characterized by the spill-over and imitation processes. This mutual dependency can also arise out of temporal dependency between geographically, culturally, or otherwise close countries (Jahn 2006). In some studies, this mutual dependency of cases is captured in an analysis of the relationship between international units

themselves. To address these issues, Lundsgaarde et al. (2010), for instance, employ dyadic data of foreign aid and trade flows to directly estimate the mutual influences of countries and money flows. What remains a task for all researchers is to identify and take into an account possible dependencies between cases in an international comparison.

2.4 Contexts

In order to systematically study the dependency of cases, the concept of *contexts* is relevant. We understand contexts as those environmental conditions into which cases are embedded, i.e. cases are sorted in groups whose characteristics can be analytically described. Cases belonging to certain contexts share elements of the context, and because of this they are similar and thus more comparable than if we worked with random samples from a universe of cases. For example, Germany belongs to the context of rich countries (defined through the GDP level, for example), and being embedded in such a context renders Germany different from those cases which are not embedded in the same context. Attention must be paid to this similarity of cases that are embedded in contexts and it can be explicitly used in the international comparison.

The similarity of cases within a context is usually connected to the characteristics of data collection. For instance, in international surveys with random samples in each country, two randomly drawn respondents from one country are more similar to each other than two randomly drawn individuals from two countries. The embeddedness of cases in a context can be addressed by using variables to describe the contextual characteristics at the case level, thus bringing the context dependency to the level of the case. For instance, in the volume edited by Cees van der Eijk and Mark Franklin (1996), the contributors pool international survey data and measure all country characteristics as individual-level variables. Yet, going back to Coleman's concern with different levels of causal paths and problems of aggregation in his bathtub heuristic (Coleman 1990), one might wish to explicitly model the differences between a case and its contexts, as these are set at different levels of aggregation and rely on different causal mechanisms.

3 Selecting Cases for Comparative Research

It should have become clear that choosing the right cases for each level is a crucial task for any comparative research design. We therefore next address different logics of sampling, as most users of quantitative individual-level techniques would say that case selection has become the central term in the comparative case literature. We start by describing the very low-key logic of contrasting empirics from different countries. We then address quasi-experimental logics of selecting country contexts. After that, we talk about random selection of country cases, and finally, and most extensively, about theoretical sampling. Table 2 provides an overview of the identified case selection logics and summarizes their defining features, as well as highlighting both potentials and pitfalls.

Table 2 Selection logics for comparisons: potentials and pitfalls (Based on the authors' on compilation)

Sampling logic	Defining feature	Potentials	Pitfalls
Contrasting	At least two country cases are used in order to describe cases analytically	Typically, very low demand on selection; some variance suffices	Inferring from the results is not possible; no known application in the multilevel world
Census	All countries in a theoretically defined universe are contained in the sample; limitations only arise from a lack of data availability	More data is always better if high quality is assured; no uncertainty due to sampling	High demand on data availability and data quality; any census can be seen as a sample from a theoretical superpopulation that needs to be defined
Quasi-experiments	Country-period cases are compared with themselves or other country-period cases in order to evaluate the impact of an ex post constructed treatment with an artificial control group	Gives high leverage on causal effects of the treatment; can be combined with hierarchical data modeling	High demand on availability of comparison; main variable of interest must be identifiable and constructed as an exogenous factor
Random sampling of country contexts	Countries are sampled randomly, usually in a very small N, in order to collect further data in a resource-rich manner	The resource-rich data collection is whitewashed as to antecedent factors, and is thus unbiased	Data analysis that places high emphasis on country-level effects will be influenced by problems of inference with small N
Theoretical case selection	Various substantiated reasons, often derived from theory or previous empirical research, are used in order to arrive at a purposeful set of cases	Relevant factors can be identified more easily; middle-range explanations are possible; explanatory narratives can be achieved more easily	Generalizability is limited; difficult to find the "correct" rule for selection

3.1 Contrasting Cases

On the simplest level, an international comparison can just be an exercise in contrasting two different case experiences of the phenomenon in question. It is a relatively shallow design as far as the international selection strategy is concerned but is applied relatively frequently in the published work. As mentioned earlier, international comparisons usually involve countries as cases for which researchers then explore differences and/or similarities between them. Analytically, such exercises have a very low-hung goal, namely to demonstrate that there is variance across countries—or that there is no such variance—and to use this insight in order to enhance the analytical description of what is happening in the various settings. There are numerous examples of such a contrasting approach. For example, Weisskopf (1975) contrasts the ways in which political leadership dealt with issues of economic development in India and China without being very explicit about why he chose these countries.

If anything, two country cases suffice in order to show similarities or differences. In principle, such comparative designs are not restricted to two cases, but can involve several cases. Researchers who have a main interest in analytically describing one

case—maybe because it is the context of a follow-up study—could use this technique of contrasting in order to analytically describe their main case in comparison with another one. In most comparative research projects, however, it seems to make more sense to select by theoretical sampling or to create a census of all available international cases in a theoretical universe. The contrasting approach usually does not have a very nuanced strategy for case selection but is likely to refer to a general argument of "these are interesting countries" and/or "we know them well".

3.2 A Census of Cases

Another relatively simple rationale in terms of selection logic is to opt for a full census of cases, given a certain theoretical definition (Berk et al. 1995). For instance, the Comparative Party Manifesto Project is a data collection for all political parties in any political system since 1945. This project has been ongoing since 1979 and successively extended the scope of available countries and years across four decades with a full census (e. g. Merz et al. 2016).

Researchers should always opt for this selection logic if they have a reasonable chance of actually realizing this census and if the data quality is similarly high across all cases and points in time. When applying this kind of logic, researchers should ask themselves whether their universe is in fact not a sample from a theoretical superpopulation. The data for countries are always restricted to a certain time period, leading to the question as to what the data for these country-period cases mean for other periods of the same countries. Some social scientists thus suggest that statistical analyses of census data should still include uncertainty measures in order to reflect that kind of inference about a theoretical superpopulation (Behnke 2005; Broscheid and Gschwend 2005).

3.3 Quasi-experimental Logic

A rather demanding way of conceptualizing a comparative design is to follow quasi-experimental logic. This means that cases are selected that have experienced some kind of treatment, i. e. an exogenous variable exerting a certain effect on them. A "sibling" case is then chosen for each treated country that mirrors the first case "as if" the treatment had not occurred.

We describe two variants of this approach. Carporaso and Pelowski (1971) conducted an analysis of the effects of membership in the European Community in its early phase. They applied interrupted time-series analysis in order to compare countries with themselves before and after significant changes in EC membership regulation. The change in various outcome variables is compared against the hypothetical value of Y that is estimated based on the past trend. In another example, Sebastian Galiani et al. (2017) compare countries against themselves, once shortly before they cross an external set threshold for receiving foreign aid by the International Development Association (the development aid agency of the World Bank) and once shortly after. Thus, a country's economic development is compared with receiving aid and without receiving aid.

This quasi-experimental logic is very powerful in terms of causal inference, as it comes close to an experimental study. There are, however, many circumstances in which such a design is not feasible, as cases of the artificial control group are not available in such a comparison, or because there is no pattern that can be operationalized as a clear treatment. It is the only international comparative design in which there is no danger of an indeterminate research design, i.e. where there are too many country-level variables and too few observations at the international level.

3.4 Random Sampling

Random selection has two general advantages. It allows the use of classic frequentist statistics in order to make inferences about the population from which the random sample was drawn. This feature is not relevant for an international comparison since the population of feasible countries or country-time points is typically not that big. Moreover, random selection blurs any differences between elements that come into the sample and those that are not drawn into the sample. No antecedent factor determines which element gets in and which one does not. That latter feature, in contrast, is very helpful. Researchers who are mainly interested in subnational units and have limited resources might choose a random sample of countries with a relatively small N because they do not want their resource-intensive research at the subnational unit to be distorted by the preselection of countries. For example, Franklin (2008) studies the reaction of governments to challenges of their human rights violations in the media. Since he uses extensive media sources in each country to identify episodes of human rights violations and reactions or nonreactions in the public media, he drew a random sample of seven Latin American countries, so that his findings are unbiased as to country characteristics. The fact that he draws inferences from a random sample of n = 7 is of no relevance to him.

The more common usage of random selection (Fearon and Laitin 2008), also with regard to large-N scenarios, takes place in numerous comparative survey studies, sometimes with surprisingly practical implications. An international consortium of researchers very often defines a country sample here (usually with some rough definitional characteristics such as liberal democracies), and then negotiates with country teams and national funding agencies as to who gets in and who stays out. Random selections of respondents are then executed within each country that allow for inferences about the population with regard to each country context. Researchers confronted with such a design have to be aware that—at the country level—the sample is not random (but typically a theoretically defined sample that is furthermore subject to feasibility aspects), and that they have at their disposal a series of equal random samples from countries for which classic frequentist techniques can be applied. Researchers very often apply random-effects models to such data sets where the statistical technique actually assumes that the country sample is also a random sample. There are some more recent methodological studies that explore how to best apply statistics in such a context (please, see also other articles in special issue).

3.5 Theoretical Case Selection

Random samples are not always appropriate in the international comparative research (Fearon and Laitin 2008; Gerring 2007; Seawright and Gerring 2008). Beyond the much too low case numbers, which place their usefulness in doubt, there are (at least) two more reasons why random selection is not particularly encouraged when it comes to comparing cases. First, as mentioned above, casing strategies include a great deal of theorizing. There might be good reasons (connected to our research questions) why we want to study both large and small countries, or why we want to make sure to look at as many developing countries as industrialized ones. Our theoretical frames might therefore induce us into a particular case selection which is predominantly theory-guided and less automatic.

Second, and again related to theory, cases are not just configurations of their properties for which configurations are interchangeable. Moreover, cases have proper names with capital letters which sometimes identify paradigm cases. Just think about a study of welfare states which would exclude any Scandinavian country, just because none of these countries "made it" into the random selection. A similar example is to study processes of transitional justice after peace agreements and their societal consequences without looking at South Africa. Certainly, there might be comparative studies in which it does not count *which* analyzed case is the actual one, but this follows more of a large-*N* logic, and thus renders randomization possible.

Alternative methods of randomization are hence available within the framework of comparative designs. Indeed, there are a number of proposals which can be grouped under this title. Most famously, Jason Seawright and John Gerring (2008) provided their readers with a typology of cases to be selected (for the following, see mainly the table in Seawright and Gerring 2008, p. 297 et seq.). One option is to study so-called *typical* cases which correspond to *on*-liers. Such a case is representative of the population or the supposed causal (or descriptive) relationship. In contradistinction to the other options discussed here, the typical case scenario is also possible for an $n = 1$ study, although both the existence and the desirability of such a situation might be doubted (Rueschemeyer 2003).[1] The study of *diverse* cases, on the other hand, is only possible if the researcher looks at more than one case (something which is in any case standard in international comparison). The idea behind this strategy is to isolate parts of the cases, and therefore to explore the variation of potential patterns. Note that diversity can be defined through both the independent and the dependent variables.

An *extreme* case is studied in order to better understand an unusual situation, which, however, does not contradict the main findings. If we assume, for instance, that social democracy and the welfare state are somehow related, then Scandinavian countries are extreme cases in this example, since they show both elements especially clearly (but still confirm our finding). It seems to be more debatable, though, whether

---

[1] One could argue that there are no $N = 1$ studies at all, and that every case study is "comparative". The rationale for such an opinion is that it is hard to imagine a case study which is conducted without any reference to other cases, including theoretically possible (but factually nonexisting) ideal cases, paradigmatic cases, counterfactual cases, etc.

the U.S. would qualify as an extreme case with regard to this example. Following Seawright and Gerring (2008), one could argue that they are extreme in the sense of representing the other extreme (i.e. negative) end of the scale, both of social democracy and of the welfare state. In this logic, they could indeed be considered extreme cases. However, there is a discussion in case-study methods as to whether these "0.0" cases (in the sense that they neither show the assumed *explanans* nor the *explanandum*) can make any causal contribution at all (see the notion of "irrelevant cases" in Rohlfing and Schneider 2013; Schneider and Rohlfing 2016; see also Beach and Pedersen 2019; Goertz 2017). For instance, if a researcher is interested in the (causal) connection between democracy and peace, it is not obvious what we can learn about this relationship from cases that are autocracies which are at war with each other. In other words, researchers should be aware of questions pertaining to the asymmetric nature of (causal) structures.

A strict difference has to be made between extreme cases and *deviant* cases. The latter are those cases where the assumed relationship does *not* hold. Their study makes sense if a researcher is interested in how these deviances came about. Note that the observed deviance is not an artifact of methodological choices, but a consequence of a chaotic and complex social world which surrounds us. As for the extreme cases, the analysis of deviant cases makes most sense if the set of cases is large enough in order to justify the qualification of cases as "extreme" or "deviant". In the end, the analysis of deviant cases might result (and this is actually the goal of such an analysis) in the elaboration of the reason for *the deviance, which then subsequently represents an additional explanatory factor.*

Seawright and Gerring (2008, p. 298) also discuss two further options for case selection which go back to the literature on comparative research designs (Berg-Schlosser and De Meur 2009; Przeworski and Teune 1970), namely the "*most similar* cases design" and the "*most different* cases design". Strictly speaking, these logics represent not only guidelines for case selection, but also already indicate a decision in favor of certain research designs and questions. If cases are most similar, then the researcher is interested in finding the reason why they differ with regard to the *explanandum* under study. The strategy is to identify those factors which are dissimilar between the otherwise similar cases in order to consider them accountable for the difference in the dependent variable. Inversely, if cases are most different, then they share a surprising similarity in the outcome under study. This similarity is then traced back to the (few) similarities in the independent variables.[2] As can be seen, the idea of which conclusions can be drawn or ought to be drawn from these designs is more critical for these selection strategies than for the previously mentioned modes of theoretical case selection.

Finally, there is also the notion of the *crucial* (or also critical) case (Eckstein 1975; Rueschemeyer 2003). These cases exist in two variants: the *most likely* case scenario

---

[2] This exposition might suggest that only the combinations of "most independent variables vary and the outcome is similar between cases" and "most independent variables are similar and the outcome differs between cases" are possible. Ragin's (1987, 2000, 2008) proposal of QCA (see also Schneider and Wagemann 2012) however shows that diversity (Ragin 2008, p. 19) can also lie on both sides. Only those designs in which nothing varies, i.e. where the cases are similar and also have similar outcomes, do not seem to be very analytically interesting.

looks into those situations where a case is expected to be most likely to produce a given outcome but fails to do so and hence shows a surprising puzzle. One example is the national party system in the United Kingdom, where the electoral system, according to many theories, would be most likely to result in a two-party system, but does not do so (myriad of regionalist parties, UKIP, the Liberal Democrats who have obtained quite a solid position as a strong third player). Such most-likely-case reasoning is usually used to revise a theory or to show the limitations of a theory, such as in our example the hypothesis that majoritarian electoral systems inevitably imply a two-party system.

The *least likely* case follows the opposite logic. We do not expect a given pattern, but it occurs. The prime example of this is Robert Michels' (1962) study on the oligarchic nature of Social Democratic/Socialist political parties. He expected Social Democratic political parties to be the perfect case where, given the respective understanding of democratic party structures in social-democratic thinking, it should *not* be possible to observe a strong oligarchy. In other words, the political parties which he analyzed were least likely to manifest the phenomenon in question. However, he detected such an oligarchy in the least likely case, and this was a further confirmation of his theory on the "Iron Law of Oligarchy" in that even an ideologically egalitarian organization changed to placing considerable power in the hands of the few. Such reasoning on the basis of most and least likely cases is only possible given two conditions: first, there has to be quite a reasonable number of cases in the reference population among which the cases under study are most or least likely, respectively, and, second, a well-established and agreed upon theory is needed, which indicates the likelihood of the phenomenon to occur.

As this exposition might have demonstrated, there are some rules and indications with regard to theoretical case selection in comparative research, but they are far from competing with the sound rules on drawing random samples in large-*N* statistical research. This, of course, has to do with the nature of comparative research where the actual case, i.e. the country with its historical pathways, its societal and cultural context, and its political *momentum*, is of utmost importance. A selection strategy in such a scenario cannot be blind (as it is one of the main features, but also of the most important strengths of random selection) but has to respect the characteristics of individual cases. And as there are many individual cases, and just as many comparative research questions, case selection strategies have to be adjusted continuously.

4 Limits of Comparative Research

We have already repeatedly pointed above to criticisms with which comparative designs are confronted. Here, we elaborate on them in a more systematic way, concentrating on six major issues which regularly form the center of the criticism: (i) selection bias, (ii) data-driven bias, (iii) causal homogeneity, (iv) links between different levels in the data structure, (v) too intimate case knowledge, and (vi) practical barriers.

4.1 Selection Bias

The strategies described for case selection regularly trigger criticism since advocates of alternative designs continue to underline the advantages of randomization. Against this background, comparative case studies are frequently accused of selection bias which, in a worst-case scenario, even alters the substantive results (for the most prominent examples, see Geddes 1990; King et al. 1994, p. 128 et seqq). Above all, the problem of "selecting on the dependent variable" is discussed. Such a selection strategy, which takes the values of the dependent variable as a reason to include or exclude cases, is quite frequent in comparative studies, since research interests at the country level are often inspired by the outcome of interest and less by an interest in the effect of the presumed causes (see the discussion on causes of effects and effects of causes (Goertz and Mahoney 2012, p. 41 et seqq.; see also Ganghof 2005)). While strong arguments have been made that such a strategy can strongly bias results (most prominently in King et al. 1994, p. 129; but also see Dion 2003, p. 128), others, while admitting that the "criticisms drawn from the quantitative perspective are well reasoned" (Ragin 2004, p. 129), also emphasize that the suspicion of bias might be "based on a very serious misunderstanding of case-oriented research." (Ragin 2004, p. 129). It is reasonable that, in a y-centered research design which looks for causes of effects, i. e. tries to explain variation in a given outcome, researchers consider it fundamental to tackle as much of that variation as possible and therefore apply those selection principles which guarantee broad coverage of the values of the dependent variable.

4.2 Data-driven Bias

From a more practical point of view, comparative studies might suffer from a data-driven bias. Again, this mainly concerns the question of case selection, since we frequently do not have enough (or have less) data about some cases, while the data situation is better for others. Just think about the fact that we certainly have easier access to information about the health insurance system in a Western European country than in some developing countries. Sometimes data are unavailable, and sometimes they would be available, but there are certain obstacles when it comes to obtaining them. This becomes even more relevant when we think about survey research where some countries tend to be over-researched, while only limited numbers of surveys exist for other countries.

The OECD world is certainly much better documented as to social science statistics than other geographical areas are. This is already problematic from an analytical point of view, since we often want to describe or infer our findings to as many parts of the world as possible. However, it also becomes a normative problem, considering that this biases our insights towards the prosperous parts of the world which enjoy a high quality of life. We will always have broader knowledge of social (and individual) life in OECD countries and will therefore always have more indications of how to improve life there even further, while other countries continue to be left out, both from our knowledge and, as a consequence, from the (political) effects of such

increased knowledge. As a consequence, we should always clarify our geographical reference population and what its composition means for the results.

4.3 Causal Homogeneity Assumption

Comparative designs might also face problems concerning the assumptions of unit or causal homogeneity which has been claimed to "[lie] at the base of all scientific research" (King et al. 1994, p. 93; for the difference between the stricter term unit homogeneity and causal homogeneity Collier et al. 2010, p. 41 et seqq.). According to King, Keohane and Verba, "[t]wo units are homogeneous when the expected values of the dependent variables from each unit are the same when our explanatory variable takes on a particular value" (King et al. 1994, p. 91). In other words: some—or many, or even all—advocates of King, Keohane and Verba would call research without causal homogeneity invalid, or at least unscientific.

This is, of course, problematic. As is frequently acknowledged, there are (at least) two obstacles to this in case-oriented comparative research: first, the complexity of the world which we observe and, second, the uniqueness of social phenomena (King et al. 1994, p. 93). Even more, these two aspects are interconnected: social processes—for example, riots—are so complex that they might even be claimed to be unique or idiosyncratic by definition. Causal homogeneity can only be controlled in an experimental setting when, in a laboratory situation, potential alternative independent variables can be held constant. Since the causal homogeneity claim is made *ceteris paribus*—i.e. with everything else being equal—its working also requires a *ceteris paribus* setting. If an assumed cause shows an effect in one case, while it does not in another, we do not know whether we have had the wrong assumptions about the cause, or whether other factors, which are present in one case but not in the other, might have influenced the effect of the cause.

Complexity and uniqueness are therefore two parts of the same story, as social processes (usually) cannot be directly manipulated, and the social world around us is too complex and too manifold to expect an assumption such as causal homogeneity to be realistic. The question then becomes how to circumvent this problem. The proposal of "simplifying reality for the purpose of making causal inferences" (King et al. 1994, p. 93) might not satisfy many researchers in the comparative world, since it is sometimes precisely the complexity of cases which attracts them. So, while a certain acceptance of the correctness of Lijphart's (1971) critical perception of comparative methods is certainly justified from the point of view of questions of inference, it can also be doubted whether inference understood in this sense is really the only goal of comparative research.

4.4 Links Between Different Levels of Aggregation in the Data Structure

This brings us to yet another point of criticism, namely the connection between various levels of analysis. Imagine that researchers are interested in understanding the causal importance of the country level. The working hypothesis could be that patterns found at a lower level, such as between individuals, exist irrespectively of what is going on at the country level. In such a research context, establishing the

robustness of findings can be achieved by diversifying the contexts in which the patterns are researched at the individual level. If we find no effect in a comparison at the country level, this would confirm the stability of the patterns at the individual level. Whatever has been found exists irrespectively of what is going on at the country level. With every country added to the analysis—even more so if the additional country is dissimilar to the ones already forming part of the study—the robustness of the findings increases.

4.5 Some Practical Obstacles

Finally, we would like to point out some more practical pitfalls when engaging in comparative research which are inspired more by research experience than by the literature. First, comparative research requires certain practical skills, without which its quality may suffer. Probably the most important aspect here is a command of the relevant languages. For instance, studying Japanese business structures without sufficient command of Japanese is certainly a constraint. Indeed, there is a language-driven bias towards the comparison of English-speaking countries. For example, Ireland is certainly over-researched, compared to a big country whose language is considered difficult, such as Russia. However, other resources are also necessary, such as data access, among other things. It is highly difficult to organize interviews on political minority rights in a country where there is oppression of the opposition than it is in a democratic country. Possessing the necessary contacts which nevertheless make data access possible is therefore more than just a virtue.

Having said that, the opposite, namely having a too intimate knowledge of a given case, might also become a hindrance. Indeed, comparative research involving new settings is often influenced by facts which were already known before. For example, a German scholar who compares the education systems of Germany and Finland will most probably look at Finland through German analytical lenses, that is, (s)he will most probably focus on those aspects of the Finnish educational system which (s)he finds important in and for a comparison within the German system. In other words, researchers tend to view other countries through the eyes of our own national identity. A famous historical example of this perspective is Tocqueville's "De La Démocratie en Amérique", a contemporary analysis of 1830s early democracy in the United States from a Frenchman's perspective, published in 1835 (first volume) and again in 1840 (second volume). Tocqueville placed considerable emphasis on the problem of tyranny of the majority, and this can be attributed to his own experience in postrevolutionary France.

There is no strategy for avoiding this form of bias altogether. However, researchers can pay close attention to the issue in two ways: a first important point is to be aware of this phenomenon. Again, it can also be considered an issue of badly executed casing if the peculiarities of a well-known case are used in order to derive more general properties of case configurations. A second important point is, however, to dig deep within the cases being studied. Writing about a country which the scholar has never visited is certainly possible and might even bear valuable and exciting results. However, case contact and case intimacy cannot be replaced by other forms of sources. Comparativists are thus encouraged to travel, not only in their minds,

but also in person. To be sure: this does not fully eliminate the danger of focusing too closely on the cases we know best, but we reduce the risk by becoming familiar with more cases.

5 Conclusion

In this article, we pointed out some pitfalls and challenges that comparative research designs at the international level have to face, and which sometimes make it hard to meet the standards of mainly statistically based mainstream social science methodology. In fact, it might even seem from the above that Lijphart's (1971) pessimistic opinion about the "comparative method", as he puts it, is reasonable.

If researchers include the nation-state level in their analysis, it is often inevitable that compromises have to be made with regard to large-N statistics which have mainly been developed for individual level data (for an interesting argument as to why statistical methods underperform at the macro level, see Kittel 2006). Alternative procedures and techniques which can be identified are thus not just simply "*lesser*" variants of statistical methods but correspond to and provide answers for the necessities of an alternative research situation. To put it more bluntly: countries are not equal to individuals—which is why different methodological approaches are needed.

Therefore, the understanding of cases as configurations of their properties, which is typical for comparative research at the macro-level, can be very helpful for finding paths and strategies for comparison. This is ultimately once more linked to the idea of "casing", i.e. the composition of a case which makes it (or some of its properties) comparable to other cases. It might also be useful not to try and imitate statistical methods and modes of inference which are typical of large data sets. (Causal) inference is one goal of the social sciences, but not the only one (although different opinions might certainly exist on this). As was already worked out at a very early date (e.g. Merton 1957), "theories of the middle range" also represent progress towards attaining the goal of knowledge accumulation. Even the fact of having greater knowledge of a single or a very limited number of cases might be seen as a success.

Comparative research methods have seen major innovations in recent decades, thus offering several new avenues (see for instance Bennett and Elman 2006; Mahoney 2010). We can hence observe a two-fold process of consolidation and systematization of macro-comparative research, thus providing important contributions on how to conduct comparisons on the cross-case level. Among these proposals, we find process tracing and causal-process observations as well as advances in comparative case study designs (e.g. Beach and Pedersen 2016a, 2019; Bennett and Checkel 2015; Blatter and Haverland 2012; George and Bennett 2005; Rohlfing 2012), systematic comparative approaches using set theory and formal logic (Ragin 2000, 2008; Schneider and Wagemann 2012), and various proposals on how to combine qualitative and quantitative approaches in multimethod research (Berg-Schlosser 2012; Maggetti et al. 2013), or how to nest cross-case and within-case analysis in integrated research designs (seminally, Lieberman 2005; following titles

offer different perspectives: Beach and Pedersen 2016b; Nielsen 2016; Rohlfing and Schneider 2016; Weller and Barnes 2014). In any case, the discussion on how to (best) engage in international comparative research is far from being over.

References

Beach, Derek, and Rasmus Brun Pedersen. 2016a. *Causal case study methods: foundations and guidelines for comparing, matching, and tracing.* Ann Arbor, MI: University of Michigan Press.

Beach, Derek, and Rasmus Brun Pedersen. 2016b. "electing appropriate cases when tracing causal mechanisms. *Sociological Methods & Research,* online first (January). https://doi.org/10.1177/0049124115622510.

Beach, Derek, and Rasmus Brun Pedersen. 2019. *Process-tracing methods: Foundations and guidelines.* 2. ed. Ann Arbor: University of Michigan Press.

Behnke, Joachim. 2005. Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen. (Can significance tests be applied to fully-fledged surveys? A few essayist remarks) *Politische Vierteljahresschrift* 46:1–15. https://doi.org/10.1007/s11615-005-0240-y.

Bennett, Andrew, and Jeffrey T. Checkel. 2015. Process tracing: From philosophical roots to best practices. In *Process tracing. From metaphor to analytic tool,* eds. Andrew Bennett and Jeffrey T. Checkel, 3–37. Cambridge: Cambridge University Press.

Bennett, Andrew, and Colin Elman. 2006. Qualitative research: Recent developments in case study methods. *Annual Review of Political Science* 9:455–76. https://doi.org/10.1146/annurev.polisci.8.082103.104918.

Berg-Schlosser, Dirk. 2012. *Mixed methods in comparative politics: Principles and applications.* Basingstoke: Palgrave Macmillan.

Berg-Schlosser, Dirk, and Gisèle De Meur. 2009. Comparative research design: Case and variable selection. In *Configurational comparative methods: Qualitative comparative analysis,* 19–32. Thousand Oaks: SAGE Publications, Inc.

Berk, Richard A., Bruce Western and Robert E. Weiss. 1995. Statistical inference for apparent populations. *Sociological Methodology* 25:421–458.

Blatter, Joachim, and Markus Haverland. 2012. *Designing case studies: Explanatory approaches in small-n research.* Basingstoke: Palgrave Macmillan.

Brady, Henry E., and David Collier. Eds. 2004. *Rethinking social inquiry: Diverse tools, shared standards.* 1st ed. Lanham, Md: Rowman & Littlefield Publishers.

Brady, Henry E., and David Collier. Eds. 2010. *Rethinking social inquiry: Diverse tools, shared standards.* 2nd ed. Lanham, Md: Rowman & Littlefield Publishers.

Broscheid, Andreas, and Thomas Gschwend. 2005. Zur statistischen Analyse von Vollerhebungen. (On the statistical analysis of fully-fledged surveys) *Politische Vierteljahresschrift* 46:16–26. https://doi.org/10.1007/s11615-005-0241-x.

Caporaso, James A., and Alan L. Pelowski. 1971. Economic and Political Integration in Europe: A Time-Series Quasi-Experimental Analysis. *American Political Science Review* 65(2):418–433.

Coleman, James S. 1990. *Foundations of social theory.* Cambridge: The Belknap Press of Harvard University Press.

Collier, David. 2014. Symposium: The set-theoretic comparative method—critical assessment and the search for alternatives. SSRN Scholarly Paper ID 2463329. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2463329.

Collier, David, and Robert Adcock. 1999. Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science* 2:537–565.

Collier, David, and James Mahoney. 1996. Insights and pitfalls: Selection bias in qualitative research. *World Politics* 49:56–91. https://doi.org/10.1353/wp.1996.0023.

Collier, David, Jason Seawright and Gerardo L. Munck. 2010. The quest for standards: King, Keohane, and Verba's designing social inquiry. In *Rethinking social inquiry. Diverse tools, shared standards,* eds. Henry E. Brady and David Collier, 2nd edition, 33–64. Lanham: Rowman & Littlefield Publishers.

Dahl, Robert A. Ed. 1966. *Political opposition in western democracies.* Yale: Yale University Press.

Dion, Douglas. 2003. Evidence and inference in the comparative case study. In *Necessary conditions: Theory, methodology, and applications*, ed. Gary Goertz and Harvey Starr, 127–45. Lanham, Md: Rowman & Littlefield Publishers.

Eckstein, Harry. 1975. Case study and theory in political science. In *Handbook of political science,* eds. Fred I. Greenstein and Nelson W. Polsby, 79–137. Reading: Addison-Wesley.

Eijk, Cees van der, and Mark N. Franklin. 1996. *Choosing Europe? The European electorate and national politics in the face of union. Ann* Arbor: The University of Michigan Press.

Fearon, James D., and David D. Laitin. 2008. Integrating qualitative and quantitative methods. In *The Oxford handbook of political methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. Oxford; New York: Oxford University Press.

Franklin, James C. 2008. Shame on you: The impact of human rights criticism on political repression in Latin America. *International Studies Quarterly* 52:187–211. https://doi.org/10.1111/j.1468-2478.2007.00496.x.

Galiani, Sebastian, Stephen Knack, Lixin Colin Xu and Ben Zou. 2017. The effect of aid on growth: Evidence from a quasi-experiment. *Journal of Economic Growth* 22:1–33. https://doi.org/10.1007/s10887-016-9137-4.

Ganghof, Steffen. 2005. Vergleichen in Qualitativer und Quantitativer Politikwissenschaft: X-Zentrierte Versus Y-Zentrierte Forschungsstrategien. (Comparison in qualitative and quantitative political science. X-centered v. Y-centered research strategies) In Vergleichen in Der Politikwissenschaft, eds. Sabine Kropp and Michael Minkenberg, 76–93. Wiesbaden: VS Verlag.

Geddes, Barbara. 1990. How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis* 2:131–150.

George, Alexander L., and Andrew Bennett. 2005. *Case studies and theory development in the social sciences.* Cambridge, Mass: The MIT Press.

Gerring, John. 2007. *Case study research: Principles and practices.* Cambridge; New York: Cambridge University Press.

Goerres, Achim, and Markus Tepe. 2010. Age-based self-interest, intergenerational solidarity and the welfare state: A comparative analysis of older people's attitudes towards public childcare in 12 OECD countries. *European Journal of Political Research* 49:818–51. https://doi.org/10.1111/j.1475-6765.2010.01920.x.

Goertz, Gary. 2006. *Social science concepts: A user's guide.* Princeton; Oxford: Princeton University Press.

Goertz, Gary. 2017. *Multimethod research, causal mechanisms, and case studies: An integrated approach.* Princeton, NJ: Princeton University Press.

Goertz, Gary, and James Mahoney. 2012. *A tale of two cultures: Qualitative and quantitative research in the social sciences.* Princeton, N.J: Princeton University Press.

Goldthorpe, John H. 1997. Current issues in comparative macrosociology: A debate on methodological issues. *Comparative Social Research* 16:1–26.

Jahn, Detlef. 2006. Globalization as "Galton's problem": The missing link in the analysis of diffusion patterns in welfare state development. *International Organization* 60. https://doi.org/10.1017/S0020818306060127.

King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research.* Princeton, NJ: Princeton University Press.

Kittel, Bernhard. 2006. A crazy methodology?: On the limits of macro-quantitative social science research. *International Sociology* 21:647–77. https://doi.org/10.1177/0268580906067835.

Lazarsfeld, Paul. 1937. Some remarks on typological procedures in social research. *Zeitschrift für Sozialforschung* 6:119–39.

Lieberman, Evan S. 2005. Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review* 99:435–52. https://doi.org/10.1017/S0003055405051762.

Lijphart, Arend. 1971. Comparative politics and the comparative method. *American Political Science Review* 65:682–93. https://doi.org/10.2307/1955513.

Lundsgaarde, Erik, Christian Breunig and Aseem Prakash. 2010. Instrumental philanthropy: Trade and the allocation of foreign aid. *Canadian Journal of Political Science* 43:733–61.

Maggetti, Martino, Claudio Radaelli and Fabrizio Gilardi. 2013. *Designing research in the social sciences.* Thousand Oaks: SAGE.

Mahoney, James. 2003. Strategies of causal assessment in comparative historical analysis. In *Comparative historical analysis in the social sciences*, eds. Dietrich Rueschemeyer and James Mahoney, 337–72. Cambridge; New York: Cambridge University Press.

Mahoney, James. 2010. After KKV: The new methodology of qualitative research. *World Politics* 62:120–47. https://doi.org/10.1017/S0043887109990220.

Mahoney, James, and Gary Goertz. 2004. The possibility principle: Choosing negative cases in comparative research. *The American Political Science Review* 98:653–69.

Mahoney, James, and Gary Goertz. 2006. A tale of two cultures: Contrasting quantitative and qualitative research. *Political Analysis* 14:227–49. https://doi.org/10.1093/pan/mpj017.

Marks, Gary, Liesbet Hooghe, Moira Nelson and Erica Edwards. 2006. Party competition and European integration in the east and west. *Comparative Political Studies* 39:155–75. https://doi.org/10.1177/0010414005281932.

Merton, Robert. 1957. *Social theory and social structure.* New York: Free Press.

Merz, Nicolas, Sven Regel and Jirka Lewandowski. 2016. The manifesto corpus: A new resource for research on political parties and quantitative text analysis. *Research & Politics* 3:205316801664334. https://doi.org/10.1177/2053168016643346.

Michels, Robert. 1962. *Political parties: A sociological study of the oligarchical tendencies of modern democracy*. New York: Collier Books.

Nielsen, Richard A. 2016. Case selection via matching. *Sociological Methods & Research* 45:569–97. https://doi.org/10.1177/0049124114547054.

Porta, Donatella della, and Michael Keating. 2008. How many approaches in the social sciences? An epistemological introduction. In *Approaches and methodologies in the social sciences. A pluralist perspective,* eds. Donatella della Porta and Michael Keating, 19–39. Cambridge; New York: Cambridge University Press.

Powell, G. Bingham, Russell J. Dalton and Kaare Strom. 2014. *Comparative politics today: A world view.* 11th ed. Boston: Pearson Educ.

Przeworski, Adam, and Henry J. Teune. 1970. *The logic of comparative social inquiry*. New York: John Wiley & Sons Inc.

Ragin, Charles C. 1987. *The comparative method: Moving beyond qualitative and quantitative strategies.* Berkley: University of California Press.

Ragin, Charles C. 2000. *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, Charles C. 2004. Turning the tables: How case-oriented research challenges variable-oriented research. In *Rethinking social inquiry: Diverse tools, shared standards*, eds. Henry E. Brady and David Collier, 123–38. Lanham, Md: Rowman & Littlefield Publishers.

Ragin, Charles C. 2008. *Redesigning social inquiry: Fuzzy sets and beyond.* Chicago: University of Chicago Press.

Ragin, Charles C., and Howard S. Becker. 1992. *What is a case?: Exploring the foundations of social inquiry.* Cambridge University Press.

Rohlfing, Ingo. 2012. *Case studies and causal inference: An integrative framework*. Basingstokes: Palgrave Macmillan.

Rohlfing, Ingo, and Carsten Q. Schneider. 2013. Improving research on necessary conditions: Formalized case selection for process tracing after QCA. *Political Research Quarterly* 66:220–35.

Rohlfing, Ingo, and Carsten Q. Schneider. 2016. A unifying framework for causal analysis in set-theoretic multimethod research. *Sociological Methods & Research,* online first (March). https://doi.org/10.1177/0049124115626170.

Rueschemeyer, Dietrich. 2003. Can one or a few cases yield theoretical gains? In *Comparative historical analysis in the social sciences*, eds. Dietrich Rueschemeyer and James Mahoney, 305–36. Cambridge; New York: Cambridge University Press.

Sartori, Giovanni. 1970. Concept misformation in comparative politics. *American Political Science Review* 64:1033–53. https://doi.org/10.2307/1958356.

Schmitter, Philippe C. 2008. The design of social and political research. *Chinese Political Science Review*. https://doi.org/10.1007/s41111-016-0044-9.

Schneider, Carsten Q., and Ingo Rohlfing. 2016. Case studies nested in fuzzy-set QCA on sufficiency: Formalizing case selection and causal inference. *Sociological Methods & Research* 45:526–68. https://doi.org/10.1177/0049124114532446.

Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis.* Cambridge: Cambridge University Press.

Seawright, Jason, and David Collier. 2010ra. Glossary."In *Rethinking social inquiry. Diverse tools, shared standards*, eds. Henry E. Brady and David Collier, 2nd ed., 313–60. Lanham, Md: Rowman & Littlefield Publishers.

Seawright, Jason, and John Gerring. 2008. Case selection techniques in case study research, a menu of qualitative and quantitative options. *Political Research Quarterly* 61:294–308.

Shapiro, Ian. 2002. Problems, methods, and theories in the study of politics, or what's wrong with political science and what to do about it. *Political Theory* 30:588–611.

Simmons, Beth A., and Zachary Elkins. 2004. The globalization of liberalization: Policy diffusion in the international political economy. *American Political Science Review* 98:171–89. https://doi.org/10.1017/S0003055404001078.

Skocpol, Theda, and Margaret Somers. 1980. The uses of comparative history in macrosocial inquriy. *Comparative Studies in Society and History* 22:174–97.

Snyder, Richard. 2001. Scaling down: The subnational comparative method. *Studies in Comparative International Development* 36:93–110. https://doi.org/10.1007/BF02687586.

Steenbergen, Marco, and Bradford S. Jones. 2002. Modeling multilevel data structures. *American Journal of Political Science* 46:218–37.

Wagemann, Claudius, Achim Goerres and Markus Siewert. Eds. 2019. Handbuch Methoden der Politikwissenschaft, Wiesbaden: Springer, online available at https://link.springer.com/referencework/10.1007/978-3-658-16937-4

Weisskopf, Thomas E. 1975. China and India: Contrasting Experiences in Economic Development. *The American Economic Review* 65:356–364.

Weller, Nicholas, and Jeb Barnes. 2014. *Finding pathways: Mixed-method research for studying causal mechanisms*. Cambridge: Cambridge University Press.

Wright Mills, C. 1959. *The sociological imagination*. Oxford: Oxford University Press.

A. Goerres, Ph.D. (LSE), Habil. (Cologne), Professor of Empirical Political Science and Head of the Political Science Department, University of Duisburg-Essen. Areas of Research: survey research, comparative political behavior, comparative political sociology and social policy research with a particular focus on ageing and immigration. Publications: Handbuch Methoden Politikwissenschaft. Wiesbaden 2019 (ed. with C. Wagemann und M. Siewert); Ageing Populations in Post-industrial Democracies: Comparative Studies of Politics and Policies. London 2014 (ed. with P. Vanhuysse); What Makes People Worry about the Welfare State? A Three-Country Experiment. British Journal of Political Science, forthcoming (with R. Karlsen and S. Kumlin).

M. Siewert, Dr. phil., post-doctoral researcher at the Department for Social Sciences, Goethe University Frankfurt. Areas of research: social science methodology with a focus on qualitative methods for causal inference, including Qualitative Comparative Analysis, process tracing, and multimethod research; party politics, executive politics and policy performance. Publications: Handbuch Methoden Politikwissenschaft. Wiesbaden 2019 (mit A. Goerres und C. Wagemann); Case Selection and Nesting of Process Tracing Case Studies. In: Process Tracing Methods. Foundations and Guidelines. Ann Arbor 2019 (with D. Beach and R. B. Pedersen); Qualitative Comparative Analysis and Process Tracing. In: Qualitative Comparative Analysis in der Soziologie. Perspektiven, Potentiale und Anwendungsbereiche. Zeitschrift für Soziologie 44, 2015 (with J. Buche).

C. Wagemann 1972, Prof., Ph.D., chair in qualitative empirical research methods of political science at the Department for Social Sciences, Goethe University Frankfurt. Areas of research: set-theoretic and case-oriented methods, most prominently Qualitative Comparative Analysis (QCA) and fuzzy sets, interest groups, the Quality of Democracy, and right-wing extremists. Publications: Set-Theoretic Methods for the social sciences. Cambridge 2012 (mit C. Schneider); Qualitative Methoden in der Politikwissenschaft. Wiesbaden 2018 (with J. Blatter and P. Langer); Handbuch Methoden Politikwissenschaft. Wiesbaden 2019 (with A. Goerres and M. Siewert); Mobilizing on the Extreme Right. Oxford 2009 (with M. Caiani and D. della Porta); Breakdown and change of Private Interest Governments. London 2011.

Families and Their Institutional Contexts: the Role of Family Policies and Legal Regulations

Karsten Hank · Anja Steinbach

Abstract This article provides an overview of families and their institutional contexts in Western societies, focusing on the role of family policies and legal regulations in union dynamics, fertility, children's wellbeing, and intergenerational relations. We argue that family dynamics are driven by changing institutional opportunities and constraints, whereas at the same time, welfare state institutions constantly need to adapt to the changing needs of "new" family forms. The empirical studies covered here provide ample evidence of multiple institutional effects on family-related behaviors and outcomes in a variety of domains. Family policy regimes supporting greater gender equality are those under which favorable outcomes are most likely to occur. Importantly, though, specific effects are not always as large, sustainable, or robust as might have been intended or expected beforehand. Methodologically rigorous evaluations of the effectiveness and efficiency of family policy measures and legal regulations thus appear an important task for future research.

Keywords Family policies · Family law · Social contexts · Welfare state policies · Cross-national comparison

K. Hank (✉)
Institut für Soziologie und Sozialpsychologie, Universität zu Köln
Albertus-Magnus-Platz, 50923 Köln, Germany
E-Mail: hank@wiso.uni-koeln.de

A. Steinbach
Institut für Soziologie, Universität Duisburg-Essen
Lotharstr. 65, 47057 Duisburg, Germany
E-Mail: anja.steinbach@uni-due.de

🙂 Springer

Familien und ihre institutionellen Kontexte: die Bedeutung von Familienpolitik und rechtlichen Regulierungen

Zusammenfassung Der Artikel gibt einen Überblick über die institutionellen Kontexte von Familien in westlichen Gesellschaften. Der Fokus liegt auf der Bedeutung von Familienpolitik und rechtlichen Regulierungen in Bezug auf Beziehungsdynamiken, Fertilität, das Kindeswohl und intergenerationale Beziehungen. Die Autoren zeigen, dass familiale Dynamiken durch sich verändernde institutionelle Rahmenbedingungen beeinflusst werden, während gleichzeitig wohlfahrtsstaatliche Institutionen permanent an die Bedürfnisse „neuer" Familienformen angepasst werden müssen. In den hier berücksichtigten empirischen Studien findet sich vielfältige Evidenz für institutionelle Effekte auf familienbezogenes Verhalten und dessen Folgen in unterschiedlichen Bereichen. Kontexte, in denen familienpolitische Rahmenbedingungen die Gleichberechtigung zwischen den Geschlechtern fördern, erweisen sich für Familien am vorteilhaftesten. Spezifische Effekte zeigen sich jedoch nicht immer so stark, nachhaltig oder robust, wie es a priori möglicherweise beabsichtigt oder erwartet worden war. Methodisch fundierte Evaluationen der Effektivität und Effizienz familienpolitischer Maßnahmen und rechtlicher Regulierungen bleiben daher eine wichtige Aufgabe für zukünftige Untersuchungen.

Schlüsselwörter Familienpolitik · Familienrecht · Soziale Kontexte · Wohlfahrtsstaatpolitik · Internationaler Vergleich

1 Introduction

This review addresses the role of institutional contexts in family-related processes and outcomes, taking a cross-national comparative perspective with a focus on "Western"—that is, demographically advanced—societies. We will concentrate on institutions manifested in family policies or family laws, which are embedded in more general configurations of policies, ideologies, and institutions, often referred to as *family regimes* (Cooke and Baxter 2010, p. 516).

Family policies are shaped by social norms and expectations (e. g., regarding gender roles and responsibilities in the family), but they usually do not directly regulate family life. Rather, family policies support specific types of families or partnerships (e. g., marriage), whilst placing others at a disadvantage (e. g., unmarried cohabitation, which is not illegal but—in the German case, for instance—does not benefit from income tax splitting). This sets incentives for certain behaviors but does not actually prescribe them. *Family law*, on the other hand, is a more direct expression of norms, consisting of "enforceable [...] rules that draw [for example] the boundaries between licit and illicit sex, lay down the grounds for the establishment of maternity and/or paternity and for the membership of kin groups, and define the socially-sanctioned obligations and legitimate expectations of household members and kin" (Willekens 2003, p. 73).

In this article, we will focus on outcomes in the pivotal domains of partnership dynamics (Sect. 2), fertility (Sect. 3), children's wellbeing (Sect. 4), and intergener-

ational relations (Sect. 5), as well as on variations therein by institutional context. Institutional effects on the gendered division of labor will not be considered here in their own right[1], but only insofar as they are linked to the four family-related domains along which we organize our review.

2 Partnership Dynamics

Both union *formation* and union *dissolution* have been highly institutionalized throughout human history, mostly by regulations concerning marriage and divorce (e. g., Goody 1983; Rosenbaum 2014). These regulations reflected economic benefits and constraints as well as social and religious norms affecting, for example, mate selection ("Who marries who?"), age at marriage, as well as individuals' chances of marrying at all. An important and longstanding geographical pattern that emerged from variations in such regulations was described by Hajnal (1965), who observed that late and non-universal marriage had prevailed in Northwestern Europe for centuries, whereas marriage had remained early and near universal in South and Eastern European countries. Only marriage legitimized a heterosexual couple's intimate relationship, whereas unmarried couples remained outside of legal jurisdiction. The extent to which the cultural and demographic divide along the so-called "Hajnal line"—ranging from Trieste to St. Petersburg—has continued to exist in the late 20th and early 21st century is subject to an ongoing debate (see Steinbach et al. 2016 for a recent contribution).

In the economically prosperous and politically conservative period following World War II, a pattern of early and almost universal *marriage* initially gained dominance in Western Europe and North America. When this "golden age of marriage" came to an end in the late 1960s and early 1970s, the Second Demographic Transition (e. g., Lesthaeghe 2010), and its underlying economic, social, and ideational shifts, brought about significant behavioral changes in many parts of Europe and America which have often been described in terms of a "deinstitutionalization" of marriage (e. g., Cherlin 2004; Lauer and Youdanis 2010): Age at first marriage started to increase steadily and substantially, whereas marriage rates decreased (in tandem with increasing divorce rates), to then stabilize at low levels. These developments occurred first in the Scandinavian countries, whereas the Mediterranean countries and—to some extent—the US (e. g., Raley 2001) clearly lagged behind. Women's total first marriage rate peaked in 1964 in Germany, declined by about half until 1991 (from 111 to 57 per 100 women), and has shown only minor fluctuations since. Women's age at first marriage increased from 23 in 1964 to almost 26 in 1991, and to over 30 in 2015 (Federal Institute for Population Research 2017).

These changes—including the rise in divorce—have been paralleled by an increase in *singlehood* (accumulated over the individual's life course and in the population; e. g., Bellani et al. 2017) and, importantly, by a rise in the prevalence of non-

---

[1] See the contribution by Grunow (2019), as well as the review by Cooke and Baxter (2010) for thorough discussions of this issue.

marital *cohabiting unions*[2], particularly in Western (European) societies. In many cases, cohabitation has become more than just a precursor to marriage but has rather evolved as a long-term alternative to marriage (e.g., Hiekel et al. 2014). Within Europe, Noack et al. (2014, p. 21) identify three distinct geographical clusters in the population aged 18–55 years: The first group, mainly consisting of South-Eastern European countries, exhibits a traditional pattern characterized by about 60% of married people, with only 5% or less of the total population cohabiting. The second cluster of predominantly Western and Central European countries constitutes a middle group, with around 50% married and about 10% of cohabiters. The third group, comprising the Nordic countries and France, is characterized by a high proportion of roughly 20% cohabiters in the population, whereas not more than about 40% are married.

Variations in *gender equality* have been suggested to be a main driver of cross-national differences in the proportions of married, cohabiting, and single individuals. With regard to lifelong singlehood, for example, the multilevel analysis by Bellani et al. (2017) provides evidence that permanently living without a partner is concentrated within countries where traditional gender values have waned, but gender egalitarianism remains poorly diffused. Cooke and Baxter (2010, p. 524) note that there is a macro-correlation between men's and women's aggregate economic equality and union type in the sense that "marriage is more prevalent in male breadwinner family regimes such as Italy, whereas cohabitation is more prevalent in regimes supporting greater gender equality such as Sweden […]." Within more equal gender settings, however, we observe a micro-correlation suggesting that women with greater individual resources tend to opt for marriage rather than for cohabitation.

It is interesting to note that the potential role of *policies and legislations* has so far often been neglected in investigations of cross-national variations in the prevalence of cohabiting unions (Perelli-Harris and Sánchez Gassen 2012). One reason for this might be that it is difficult to establish the extent to which changes in policies and laws are the cause or the consequence of the demographic phenomena to which they refer (e.g., Bradley 2001; Eekelaar 2010). The far-reaching legal recognition of cohabitation in contemporary Western societies has clearly lifted much of the social and economic pressure to marry that previous generations of couples had borne. However, the legal situation of cohabiters still varies widely throughout Europe (for a comprehensive overview see Perelli-Harris and Sánchez Gassen 2012): Even though Norway and Sweden have not formalized cohabiting unions as registered partnerships (unlike France and the Netherlands), they are nonetheless among the most advanced countries in terms of the legal harmonization of cohabitation and marriage. Germany[3] and Switzerland represent the other end of the continuum, as they "have been the most reluctant to equalize cohabitation and marriage, or even to recognize cohabitation" (Perelli-Harris and Sánchez Gassen 2012, p. 463; also see Bradley 2001). Differences pertain to rights and responsibilities both *during* the

---

[2] Note that partners in a steady relationship do not necessarily have to cohabit; see, for example, the analysis of "*living apart together*" relationships by Asendorpf (2008) and Liefbroer et al. (2015).

[3] See Wellenhofer (2016) for a more detailed discussion of the case in Germany.

union (e. g., the right to co-insure a partner in the public health insurance system, or the obligation to support each other financially) and *after* union dissolution (e. g., regarding the division of property, the obligation to pay alimony, or—after the death of a partner—entitlements to inheritance). However, even in countries with high levels of recognition and actual cohabitation (such as France or Sweden), attitudes towards cohabitation are not unambiguously positive, and the value attached to marriage remains high (e. g., Noack et al. 2014; Treas et al. 2014).

Marriage is one precondition for *divorce*; the other is the recognition of divorce as a legal act. By 1950, most European countries permitted spouses to divorce (with Ireland being a noteworthy exception, legalizing divorce as late as 1997), but restrictive divorce requirements and procedures still often made it difficult or costly for married couples to legally separate. This was alleviated by the introduction of "no-fault" grounds for divorce (established in most countries by the middle or in the second half of the 20th century) and, subsequently, by a shift from laws requiring mutual consent to those permitting unilateral divorce (occurring mainly in the 1970s and 1980s; for an overview of legal reforms in a variety of countries, see Perelli-Harris et al. 2017, Appendix). Moreover, legal practice—that is, the de facto divorce regime—has been shown to exhibit a significant influence on divorce rates (e. g., Eekelaar 2010; Kneip and Bauer 2009).

Some legal changes had direct effects on divorce rates.[4] Prior to the introduction of divorce as a legal opportunity to exit marriage, official divorce rates were obviously zero (which does not of course mean that marital breakdown did not take place). Another example is the prescription of a one-year separation period before divorce, which was introduced in West Germany during the late 1970s and resulted in a substantial short-term decline in divorce (see Federal Institute for Population Research 2017). Caution is however necessary in order to avoid confusing the effects of de jure changes in divorce laws with other underlying trends, such as the increase in cohabitation (see Perelli-Harris et al. 2017). Moreover, and finally, governments might have changed divorce laws because many couples had already separated.

A relatively recent and important development is the emergence of *same-sex marriage* as a "new social phenomenon" in a number of Western countries (e. g., Chamie and Mirkin 2011; Festy 2006). Whereas Denmark legalized "registered partnerships" as early as 1989, the Netherlands was the first country to allow gay and lesbian couples to actually marry in 2001. Many US Federal States and European countries followed suit in the years that followed. Germany adopted a so called "*Lebenspartnerschaftsgesetz*"(Life Partnership Act) in 2001, which enabled couples to obtain legal recognition for their union through a registration procedure that was distinct from marriage, but still provided them with benefits very similar to those received by married opposite-sex couples. This law also regulated child-related issues in same-sex partnerships, particularly custody and adoption rights (see Rupp and Haag 2016). Germany eventually legalized same-sex marriages in 2017.

In summary, marriage, cohabitation, and divorce continue to be subject to strong legal regulation (determining, for example, at which age or under which conditions

---

[4] Note that legal reforms (e. g., Kneip et al. 2014) and welfare state policies (e. g., Bitler et al. 2004) might also exhibit *indirect* contextual effects on divorce.

the transition into a specific state is possible). However, *there is little evidence to suggest a* direct *impact of family policies or family law on changing partnership dynamics* in "Western" societies. Marriage and divorce obviously have to be legal opportunities: Gay marriage, for example, was not possible in Germany before 2017, and divorce was not possible in Ireland before 1997. But de jure changes in family laws might be a consequence rather than the cause of changes in legal practices and the demographic phenomena to which they refer. Moreover, whether a couple chooses to live in a marital or non-marital union appears to be influenced (at least) as much by a country's level of *gender equality* as by the extent to which marriage and cohabitation are legally harmonized.

3 Fertility

Against the background of sustained below-replacement fertility in demographically advanced societies, the role of family policies in childbearing behaviors has received considerable attention (for reviews see Bujard 2016; Gauthier 2007). Welfare state institutions may intentionally affect the timing and quantum of fertility (as a consequence of pronatalist family policies), or they may do so unintentionally (as a consequence of, for example, labor market policies affecting fertility through employment decisions). Even though there is a plethora of fertility-related policy measures, the core "family policy package," on which we will focus in this section, has been suggested to consist of three main types of policy instruments, namely: *financial transfers, paid leave*, and *childcare services* (e. g., Luci-Greulich and Thévenon 2013).

Drawing primarily on economic—or, more generally, rational choice—approaches to fertility (e. g., Werding 2013), it is argued that "[f]amily policies potentially contribute to re-increases in fertility as they can reduce the costs of fertility, either in monetary terms or in terms of opportunity costs." (Luci-Greulich and Thévenon 2013, p. 390). Direct compensation for the economic costs of children usually comes in the form of cash benefits and/or fiscal transfers to families. An early macro-level time-series analysis covering 22 Organisation for Economic Co-operation and Development (OECD) countries over the period 1970–1996 finds minor positive effects of cash benefits on the total period fertility rate (Gauthier and Hatzius 1997). This result was corroborated more recently by Luci-Greulich and Thévenon (2013), whose study was based on 18 OECD countries in the period 1982–2007. These effects, however, seem more obvious when the timing of births rather than the quantum of fertility is considered.

Research based on microdata generally confirms these findings, but also indicates a varying effect of cash benefits by birth order (e. g., Aassve and Lappegård 2009, for Norway; Laroque and Salanié 2004, for France; Vikat 2004, for Finland). A noteworthy exception is Kalwij (2010), whose cross-nationally comparative analysis of data from the European Social Survey showed no significant impact of more generous family allowance programs on the timing of births or individuals' completed fertility. Also in the German context, analyses of the role of child benefit (*Kindergeld*) payments tend to provide no or at most weak evidence of overall ef-

fects on fertility (Bujard 2016, p. 627). In 2007, however, the German government introduced the new parental allowance (*Elterngeld*), which replaced means-tested parental leave benefits targeted at lower-income families with payments related to pre-birth earnings. Analyzing administrative microdata, Raute (2018) indeed identified an increase in fertility following this reform, driven mainly—as intended—by women at the middle and upper end of the education and income distributions (also see Bujard and Passet 2013).

A similar policy was introduced earlier in Sweden, where Andersson et al. (2006) did not find any major educational differentials in the reaction to the reform. The authors' primary interest, however, lay not in the role played by parental leave benefits, but in the duration of paid parental leave (specifically the eligibility interval during which benefits may be retained). Confirming results of a previous study by Hoem (1993), their analysis of population register data provides evidence that the extension of the eligibility interval set incentives to have another child while still being on parental leave. The Swedish leave policy reform was thus interpreted as a "speed premium" affecting the timing of births. Similar effects are not only found in other Nordic countries (e.g., Rønsen 2004, for Norway and Finland), but also in two Austrian studies (Hoem et al. 2001; Lalive and Zweimüller 2009).

A more recent innovation in parental leave policies is the introduction of "daddy months," dedicating some share of the total leave duration to fathers. The first countries to establish this policy were the Nordic ones, but others—such as Germany—followed suit (see for example Geisler and Kreyenfeld 2011). Whereas the main aim was to promote gender equality, Duvander et al. (2010) showed—based on an analysis of register data—that fathers' take-up of parental leave is positively associated with continued childbearing in Sweden and, even more so, in Norway.

Despite these findings, it is important to note that parental leave policies are not designed to influence parents' fertility behavior directly, but that they particularly aim at enhancing children's wellbeing (see Sect. 4) and the compatibility of childrearing and female employment (e.g., Ellingsæter 2009). This latter issue is important because Brewster and Rindfuss (2000, p. 271), for example, concluded from their review of the literature that "women's labor force participation lies at the heart of most explanations of fertility and fertility change," and that the frequently observed inverse "association between fertility and women's labor force activity reflects the incompatibility between caring for children and participation in economically productive work that typifies industrialized societies." Even though access to affordable, high-quality childcare has been proposed as one of the most important structural conditions to solve this compatibility problem, empirical studies employing multilevel data provide inconclusive evidence regarding its effect on fertility.[5]

In Southern European lowest-low fertility, familialistic welfare state contexts, Del Boca (2002; for Italy) and Baizán (2009; for Spain) found that more comprehensive availability of formal childcare had a positive effect on fertility. Rindfuss et al. (2010) report similar findings for a somewhat different demographic and welfare

---

[5] See, for example, Kreyenfeld and Hank (2000); Zoch and Hondralis (2017) for investigations of the association between child care availability and maternal *employment*.

state context, namely Norway, where greater childcare availability increases transition rates at every parity, and thus also completed fertility. However, such an effect was neither found in earlier Norwegian research (Kravdal 1996; also see Rønsen 2004), nor in Andersson et al.'s (2004) study of continued childbearing in Sweden. For Germany, Hank et al. (2004) found that the availability of public childcare had a positive impact on Eastern German women's transition to the first child, whereas this was not the case for their Western German counterparts. However, this analysis based on Socio-Economic Panel (SOEP) data covered a rather short window of observation around the turn of the millennium, and was thus based on a relatively small number of events. More recently, Bauernschuster et al. (2016) exploited the temporal and spatial variation in childcare coverage induced by a significant expansion of childcare slots for young children in the mid-2000s. Matching information from birth registration records with county-level data on childcare coverage, their analysis suggests that a ten-percentage-point increase in childcare coverage leads to an increase in birth rates of almost three percent. The authors not only claim that their findings actually reflect a quantum effect, but also that investments in public childcare are more efficient with regard to raising fertility than expansions in child benefit expenditures (Bauernschuster et al. 2016, p. 1002).

This latter finding is consistent with Kalwij (2010, p. 517), whose findings from 16 Western European countries indicate "that increased expenditure on family policy programs aimed at empowering women through opportunities to combine family and employment—thereby reducing the opportunity costs of children—generate positive fertility responses. More specifically, extending maternity and parental leave as well as childcare provision causes women to have children earlier in life, and to have more children." It therefore seems important to acknowledge that it is a combination of policy instruments that is most likely to facilitate the choice to have children, but that not all measures have the same weight (see also Harknett et al.'s (2014) analysis of the role of countries' broader "family support environments" in individuals' childbearing plans and actual childbearing behaviors).

Finally, alongside the abovementioned set of family policy instruments there are important *legal regulations* potentially affecting the number of children parents may have, especially if abortions, adoptions, and the use of assisted reproductive technologies (ART) are considered: *Abortion* has been discussed as a possible substitute to modern contraception in less developed countries, and its legalization has thus been suggested to potentially impact fertility (see Gutierrez Vasquez and Parrado 2016; Miller and Valente 2016 for recent investigations). Considerable variation in both legal restrictions and rates of termination of pregnancies continues to exist in Europe (Gissler et al. 2012; see David 1992 for a historical account). However, countries with unrestricted access to early termination of pregnancy do not exhibit higher rates than countries with more restricted access. Germany, for example, which allows early terminations of pregnancies without legal indication upon women's requests, reported only 6 terminations per 1000 women aged 15–49 in 2008 (compared to an EU average of 10/1000; see Table 1 in Gissler et al. 2012).

The prevalence of *adoptions* varies substantially across countries, being relatively high in the US and comparatively low in Germany, where the number of adoptions has continuously declined—to a total of 3812 in 2015—since the 1980s

(Bovenschen et al. 2017). Whereas some of this decline seems to be attributable to more generous state support for families, advances in birth control and reproductive medicine, as well as more liberal abortion laws, higher adoption rates in other countries also suggest an important role played by a lower level of social acceptance and more complicated legal regulations on adoptions in Germany (for a review of the latter see Reinhardt 2017). The number of live births following *ART* treatment in Germany is substantially higher than the number of adoptions, with a peak of more than 18,000 in 2003, followed by a sharp decline in 2004 and a subsequent recovery to roughly 14,000 in 2012. The decline in the number of women treated, treatment cycles, and—consequently—in live births, was not due to changes in the overall legal framework for ART, but resulted from a significant reduction in the reimbursement of the costs of treatment by statutory health insurance (for a detailed overview see Trappe 2017). Variations in reimbursement levels—rather than legal regulations—have also been suggested to be the main driver of cross-national differences in the use of ART across Europe. Usage is particularly high in Denmark, Slovenia, and Spain, where the cost of treatment is completely covered by national health plans (Präg and Mills 2017). Even though the numbers of both adoptions and successful ART treatments are moderate in absolute terms (compared to, for example, a total of more than 730,000 births in Germany in 2015), they are likely to become increasingly relevant phenomena against the background of further medical advances, a sustained delay in childbearing, and the liberalization of same-sex parenthood (e.g. Waaldijk 2009).

In summary, whereas there is some evidence to suggest an impact of specific policy instruments on the timing (financial transfers, paid leave) and quantum (public childcare services) of childbearing, *combinations of such instruments aiming to empower women* appear to be most effective with regard to the aim of raising fertility. Moreover, legal regulations are important to shape the conditions under which, for example, induced abortions or the use of assisted reproductive technologies may take place, but they do not seem to have a major quantitative impact on the fertility outcomes that are derived from such practices.

4 Children's Wellbeing

Whereas families constitute the most important context for children and their development, they are affected both directly and indirectly by institutional contexts shaping the circumstances under which they grow up. The relevant policies and laws here are often the same ones affecting parents' decision to have children, as well as the consequences resulting from this decision (especially in terms of labor force participation; see Sect. 3). A major concern is the role of such welfare state institutions in children's wellbeing—health, educational opportunities, poverty risks—and how they might buffer, for example, adverse effects of family disruption (for a comprehensive analysis see Engster and Stensöta 2011).

A central question is who cares for children (and under what conditions). *Parental leave* regulations provide opportunities and set incentives for parents—primarily mothers, but increasingly for fathers as well (e.g., Boll et al. 2014; Bünning

2015)—to stay away from the labor market for some time and provide full-time care for their children. Longer leave entitlements[6] may potentially affect a variety of child outcomes. To begin with, there might be *health* effects resulting, for example, from reduced maternal stress or prolonged breastfeeding.[7] Macro-level evidence from a number of OECD countries (e.g., Patton et al. 2017; Tanaka 2005) suggests that longer job-protected, paid parental leave substantially decreases mortality among infants born to eligible mothers (with additional smaller positive effects on birth weight). Whereas Tanaka (2005) did not identify any significant effects if leave was provided without job protection or adequate payment[8], Rossin (2011) found that even the introduction of 12 weeks of unpaid maternity leave mandated by the 1993 Family and Medical Leave Act in the US led to small increases in birth weight and a significant decline in infant mortality. Studies assessing other specific health outcomes (such as infections, chronic conditions, or hospital admissions) using microdata did not systematically find causal effects of the length of parental leave on younger children's wellbeing (e.g., Baker and Milligan 2008 for Canada; Beuchert et al. 2016 for Denmark), but recent evidence from Australia indicates that paid leave entitlements might reduce disadvantaged children's probability of having *multiple* ongoing health conditions (Broadway et al. 2017).

The more general institutional setting in which a leave policy is enacted obviously matters: "a reform expanding paid leave from twelve to fifteen months in a setting with subsidized child care and universal health insurance [...] is dramatically different from one that provides six weeks of paid leave for the first time in a setting where neither child care nor health insurance is guaranteed" (Rossin-Slater 2018, p. 14). This might also, and particularly, be the case, if children's *educational* outcomes are considered, given that countries' educational systems (including the arrangements that they make for preschool public childcare) vary widely. However, recent micro-level evidence from institutional contexts as diverse as, for example, Norway (Dahl et al. 2016) or Austria (Danzer and Lavy 2018), does not suggest any significant effect of parental leave extensions on schooling outcomes (such as test scores or high school dropout rates[9]). In a comprehensive study of several parental leave reforms in Germany, Dustmann and Schönberg (2012) showed: (a) that the expansion in paid leave from 2 to 6 months in 1979 did not increase children's average years of schooling, (b) that the expansion from 6 to 10 months in 1986 did not substantially raise the probability of completing a high-track school (i.e., Gymnasium, a grammar school equivalent), and (c) that the expansion in unpaid leave from 18 to 36 months in 1992 even seems to have lowered children's educational attainment. Finally, in an analysis of macrodata from 20 OECD countries, Engster and Stensöta (2011, p. 84)

---

[6] Even though longer leave *entitlements* (and the associated income replacements) have a positive effect on parents' *actual uptake* of parental leave, they are clearly not the only determinant of the time that parents stay away from work in order to spend time with their children (see Rossin-Slater 2018, pp. 9–10).

[7] Next to affecting children's health, parental leave might also be associated with *maternal* health outcomes (e.g., Guertzgen and Hank 2018).

[8] Note that the generosity of parental leave *benefits* may have a non-negligible impact on family income, thereby eventually affecting children's health (e.g. Kuehnle 2014).

[9] Carneiro et al. (2015), however, observed a two-percentage-point decline in high school dropout rates after an *extension* of parental leave duration and the *introduction* of paid leave in Norway in 1977.

found "little long-term effect of family policy regimes on educational achievement (test score), but a significant correlation between family policy generosity and higher educational attainment (remaining in school longer)."

Importantly, some studies also point to differential effects caused by, for example, parental education: Liu and Skans (2010) identified a positive effect of prolonged parental leave for children of well-educated mothers in Sweden, and Cools et al. (2015) report that Norwegian children's school performance improved if their fathers took paternal leave, especially when they had attained a higher level of education than the mother had. Another important distinction is made by Rossin-Slater (2018, p. 15; italics not in the original), who concludes from her review of the literature that "*extensions* in existing paid leave policies have had little impact on children's well-being, [while] the evidence suggests that the *introduction* of short paid and unpaid leave programs can improve children's short- and long-term outcomes."

Whereas leave programs foster parental childcare at home, many countries have also expanded the provision of *public daycare* for children, and a growing number of studies investigate the effects of center-based early childhood education and care programs with regard to children's school achievements as well as their cognitive and socio-emotional development (for reviews see Anders 2013; Burger 2010). Cross-national comparative studies covering a broad range of economically developed societies point to a generally positive micro-level correlation between attendance of pre-school institutions and subsequent PIRLS or PISA test scores (Cebolla-Boado et al. 2017; Schütz 2009). The strength of this association seems to vary by country, depending on the "structural" quality of preschool education: It tends to be strongest in contexts with higher spending on pre-primary education per pupil, larger shares of children attending privately managed pre-primary institutions, as well as higher relative pay and higher levels of training for pre-primary teachers (Schütz 2009). Evidence from Anglo-Saxon countries suggests that early childcare is positively associated with test scores at school entry (e. g., Hansen and Hawkes 2009; Magnuson et al. 2007), but that this effect tends to dissipate later on (which is consistent with Spieß et al. 2003, who show that there is no significant relationship between kindergarten attendance and children's later school placement in the German tracking system). However, even though long-term effects of early educational interventions may be smaller than initial effects, they can still be substantial—especially for children from disadvantaged social backgrounds (e. g., Cebolla-Boado et al. 2017)—if designed properly (e. g., Barnett 2011).

Reducing *child poverty*, which has been shown to exert substantial adverse short- and long-term effects on a variety of life domains (e. g., Duncan et al. 2012), is another major policy concern. Whereas relative child poverty is as low as 5% in Norway, it exceeds 20% in the US—and is higher than overall poverty in most countries (Smeeding and Thévenot 2016: Fig. 1). Household composition and parents' labor market participation have been suggested to play a crucial role among childhood poverty drivers. Particularly single mothers and their children almost universally experience elevated risks of poverty. These are highest in the US and substantially lower in welfare state contexts providing strong public cash support as well as work support to increase mothers' labor earnings (e. g., Smeeding and Thévenot 2016; see also Brady and Burroway 2012). Moreover, when studying child poverty by

family structure in a set of five liberal welfare states during the 2008 recession, Rothwell and McEwen (2017) found that children in cohabiting families were less well protected against market instability than those whose parents were married. The authors also show that family benefits in the form of income transfers substantially contribute to reducing poverty among non-married—often fragile—families, whose risk of being poor is again highest in the US. Finally, Engster and Stensöta (2011, p. 84) conclude from their study of OECD countries that "dual earner regimes, combining high levels of support for paid parenting leaves and public child care, are significantly associated with low levels of child poverty."

A plethora of studies have shown that separation or divorce are associated with a variety of adverse outcomes for children: Alongside increased poverty risks and educational disadvantages, there is also evidence of greater psychological and behavioral problems, as well as a greater propensity to get divorced themselves in adulthood (for reviews see Amato 2000; Härkönen et al. 2017). Whereas such relationships between family disruption and child outcomes are found almost universally, many studies suggest cross-national variation in the strength of the associations observed. Detrimental effects on children's school achievements, for example, seem to be slighter in family policy contexts that balance out resources between single- and two-parent families (e.g., Hampden-Thompson 2013; Pong et al. 2003).

Moreover, and importantly, child support and custody laws are likely to affect children's wellbeing after their parents' separation or divorce (e.g., Del Boca 2003). *Child support* consists of a regular income transfer from the father to the mother that is often ordered—and legally enforced—because of income disparities between the parents (e.g., Huang et al. 2003; Stirling and Aldrich 2008). With regard to *child custody*, one needs to distinguish between legal (regulating parents' decision-making) and physical (regulating parenting time). Sole physical custody usually results in a situation where the child lives with one parent only (most often the mother), thus substantially losing financial and emotional support that was previously provided by the other parent (most often the father). Even though non-resident fathers may still contribute to children's wellbeing (King and Sobolewski 2006), custody agreements and living arrangements have been shown to have a major impact on fathers' involvement in their children's lives (e.g., Swiss and Le Bourdais 2009). Several Western countries have thus revised their custody laws in the past decade, thereby strengthening joint physical custody arrangements that support shared parenting after separation or divorce. Whereas the consequences of such an arrangement (in which the child is supposed to live 35% or more of the time with each parent) are not yet fully investigated, previous research suggests that the wellbeing of children in joint physical custody is at least as high as in sole physical custody (for recent reviews see Baude et al. 2016; Steinbach 2018).

In summary, whereas some studies identify (direct) policy effects on children's wellbeing, such effects are far from universal. Whether parental leave, for example, affects children's health or education very much depends on which specific dimension of the outcome is considered (e.g., infant mortality vs. chronic conditions; educational achievement vs. attainment). With regard to education, children from disadvantaged social backgrounds appear to be the ones benefiting the most from early educational interventions, whereas the children of more highly educated

parents seem to be the main beneficiaries of parental leave extensions. Moreover, introducing a parental leave program may have a larger impact than extending the eligibility interval, and the long-term effect of early educational interventions, for example, may be smaller than their initial effect. Overall, *laws and policies fostering mothers'* and *fathers' active involvement in both parenting and paid work appear to contribute the most to improving children's wellbeing*.

5 Intergenerational Relations

In contradistinction to the notion of "less family" that has sometimes been used to describe the main trends in marriage and fertility observed during the second half of the 20th century (see Sect. 2 and 3 of this review), the "family decline" hypothesis (Popenoe 1993) has been widely rejected as far as intergenerational relations within families are concerned. However, despite high levels of solidarity between family members overall across two or more generations throughout Europe and the US, we also observe considerable variations across welfare states with regard to both upward and downward assistance or transfers. In (Western) Europe, for example, there is a continuum marked by relatively "weak" family ties in the Nordic countries and relatively "strong" family ties in the Mediterranean ones (e. g., Hank 2009). This geographical pattern reflects longstanding variations in cultural characteristics, social norms, and preferences, which are, inter alia, manifested in different policies and legal obligations to support parents or children in need. In more general terms, these have sometimes been described along a "familialism/de-familialization" continuum (see Saraceno and Keck 2010; see also Dykstra 2018).

Requirements to contribute financially to the costs of *eldercare* for parents (upward intergenerational support) are a prominent and obvious example; see Haberkern and Szydlik (2008, 2010) for a detailed discussion. Consistent with the notion of "de-familialization" (that is, reduced family responsibilities and dependencies), there are no such obligations in the Scandinavian countries.[10] The "familialistic" Mediterranean countries (as well as many conservative welfare states), however, provide publicly funded services only if the person in need or his or her close relatives—children or in some cases siblings—cannot afford to bear the costs of care themselves. Accordingly, eldercare provided by the younger generation of family members is substantially more common in the latter countries than it is in Northern Europe, where professional services are more readily available and their use is widely accepted. Unfortunately, most countries so far offer only very limited (financial) support for informal carers, and policies to assess their needs are still at an early stage, especially in familialistic settings (see Courtin et al. 2014). Moreover, there are important *gender differences* in the provision of informal care to elderly parents, which also vary by welfare state context. Daughters are universally more likely to provide care to the older generation, but this gender inequality has been shown to be highest in countries with a high level of intergenerational care, high

---

[10] It goes without saying that a lack of legal *obligations* to provide support does not rule out high levels of *voluntary* intergenerational support or emotional closeness in parent-child relationships.

public spending on old-age cash benefits, low provision of professional care services, high family obligation norms, and a high level of division of labor across gender lines (Haberkern et al. 2015).

Cross-national differences in the provision of *childcare* by grandparents (downward intergenerational support) have been suggested to result from the interplay between female employment and family policies, specifically the provision of public daycare for children (e.g., Bordone et al. 2017; Hank and Buber 2009). Whereas Scandinavian grandparents are more likely than their Southern European counterparts to provide grandchild care, the latter are more likely than the former to provide intensive (that is, regular) childcare. One explanation for this (seemingly counterintuitive) pattern is that the high level of regularly provided public childcare in Northern European countries creates an opportunity structure that fosters maternal employment, but also requires that grandparents occasionally complement institutional care (e.g., if the grandchild's mother needs to work extra hours). In Mediterranean countries, on the other hand, the lack of public daycare for children inhibits maternal employment, and there is only limited demand for grandparents to step in because mothers tend to be full-time carers. If, however, a Mediterranean mother seeks gainful employment, she has to rely on grandparents' support on a regular basis (Hank and Buber 2009; also see Di Gessa et al. 2016).

Clearly, cross-national differences in the use of close kin as providers of child- or eldercare are not driven by legal and structural conditions alone, but also by cultural factors, especially variations in preferences, attitudes, and norms regarding the use of formal care services (e.g., Haberkern and Szydlik 2010; Jaapens and Van Bavel 2012). Moreover, a simple dichotomy distinguishing societies that are characterized by strong (weak) families and weak (strong) welfare state institutions does not provide an adequate concept to explain the more complex empirical patterns that have been observed in recent studies (e.g., Saraceno and Keck 2010). Models postulating a *joint responsibility of welfare states and families* in the production of social services appear as a powerful alternative to previous simplifications. They allow researchers to transcend (partly ideological) questions such as whether welfare states crowd out families, asking instead how existing needs can be met in the most efficient way and in line with people's own preferences. Motel-Klingebiel et al. (2005, p. 864) thus argue that in a "situation of 'mixed responsibilities', it is possible for formal and informal support systems to be complementary and to take on specialised roles." Along these lines, Igel et al. (2009, p. 220) showed for example that, in more generous European welfare states, "[p]rofessional providers take over the more challenging, demanding and essential care of the elderly, whereas children tend to give voluntary, less intensive, and less onerous help."

The interplay between welfare state institutions and families becomes even more complex if the growing shares of *non-intact families* and *non-biological parent–child relationships* are taken into account. Laws regulating child custody or alimony payments, for example, have been shown to have long-term implications for intergenerational relations in adulthood: Custody arrangements affect children's living arrangements (Cancian et al. 2014) and non-resident fathers' involvement with children (Seltzer 1998), whereas the generosity of alimony payments influences the level of economic distress in non-intact families (Kreyenfeld and Martin 2011). Specifi-

cally, Arránz Becker et al. (2013, p. 1133) suggest that more generously provided welfare state support for children "benefits the generally disadvantaged stepchildren especially, and [...] may make the socioeconomic situation of stepchildren less conditional on their relationship with the stepparent." Such institutional effects may ultimately have long-term direct and indirect implications for a variety of interrelated dimensions of (step-)parent–child relationships (see, for example, Steinbach and Hank 2016).

In summary, the provision of care is an important phenomenon at the intersection between families and welfare states. Eldercare (that is, upward intergenerational support) is clearly more directly affected by legal regulations and policies than, for example, the provision of grandchild care (that is, downward intergenerational support). In both cases, however, we observe a *complementary relationship of specialized roles* that families and welfare states take on in the production of care. Maintaining this balance will be a challenge in a situation characterized by population aging, (partial) welfare state retrenchment, changing gender roles, and increasing family complexities.

6 Conclusion

The aim of this article was to provide an overview of families and their institutional contexts in Western societies, focusing on the role played by family policies and legal regulations in union dynamics, fertility, children's wellbeing, and intergenerational relations. This makes the topic of our review a moving target with closely interrelated parts: Family dynamics are driven by changing institutional opportunities and constraints, whereas welfare state institutions constantly need to adapt to the changing needs of "new" family forms (e. g., Vaskovics and Huinink 2016).

The studies covered here provide ample evidence of manifold direct and indirect institutional effects on family-related behaviors and outcomes in a variety of domains. A general conclusion that we can draw from this research is that family policy regimes supporting greater *gender equality* are those under which favorable outcomes—such as higher fertility or greater child wellbeing—are most likely to occur. Importantly though, the *effects of specific policies are not always as large, sustainable, or robust as might have been intended or expected* beforehand. Evaluating the effectiveness and efficiency of family policy measures and legal regulations thus appears to be an important task for future research (e. g., Bonin et al. 2013; Fichtl et al. 2017).

Whether a pronatalist family policy, for example, has been successful can often not be properly assessed by simply comparing a population's total fertility before and after the introduction of that policy, even when trying to hold other factors fixed. It may be difficult to actually disentangle the intended or unintended impact of a specific reform from direct or indirect effects of a country's general institutional ("family regime") set-up and possible parallel changes therein. Moreover, the same policy might affect individuals' fertility in different ways, depending on whether we consider its timing or quantum, first- or higher-order births, marital or non-marital childbearing. In addition, there might be cross-level interactions between policies

and individual characteristics, such as education, resulting in differential effects for various subpopulations. And, eventually, it may be difficult to establish whether certain legal regulations (e. g., the introduction of "daddy months") are primarily a cause or a consequence of changing sociodemographic behaviors (such as fathers' greater involvement in childrearing).

These *methodological challenges*, amongst others, call for great caution to be applied when interpreting the results reported in empirical studies as causal effects. Some studies investigate, for example, consequences of individuals' *use* of parental leave (and/or the uptake of benefits; e. g., Aassve and Lappegård 2009), whereas other—econometrically more rigorous—studies account for exogenous changes in individuals' *eligibility* to take paid leave (e. g., Dahl et al. 2016; Dustmann and Schönberg 2012; Lalive and Zweimüller 2009), thereby avoiding potential selectivity issues. Many of the latter (quasi-experimental) studies apply a regression discontinuity or differences-in-differences design, exploiting within-country institutional variation over time rather than between-country institutional variation, which is a common identification strategy in multilevel research.[11] Even though multilevel modeling has nowadays become a standard tool in cross-national comparative research, there is also an increasing awareness of its limitations, resulting from the necessity of a sufficiently large number of aggregate-level observations in order to obtain reliable estimates of parameters summarizing country effects (e. g., Bryan and Jenkins 2016; Schmidt-Catran et al. 2019). Because multilevel analysis is thus not a panacea, it seems important to further explore the potentials of alternate research designs for "small n" cross-national studies. "Most similar/most different systems" designs, for example, are well established in political science (see Anckar 2008), but have so far rarely been employed in family research (for an application see Berninger 2013).

Inevitably, our review has several limitations: First, we did not consider any "non-Western" societies (see the contributions in Hill and Kopp, 2015, Section I, for overviews of families in African, Asian, and Latin American contexts). Another "geographical" restriction is that we did not systematically account for potentially relevant social contexts at sub-national levels of spatial aggregation (e. g., Hank and Huinink 2015). Second, we exclusively considered institutions manifested in family policies or family laws. The educational system, however, is an important example of other kinds of institutional contexts that might also play an important role in individuals' demographic behaviors, especially partner choice and family formation (e. g., Blossfeld and Huinink 1991; Blossfeld and Timm 1997). Third, and finally, it was beyond the scope of this review to thoroughly incorporate the recent discussion about the diffusion of gender-egalitarian norms, the ongoing "gender revolution," and their interaction with welfare state institutions in shaping changing family behaviors (see Esping-Andersen and Billari 2015; Goldscheider et al. 2015). This latter issue in particular deserves adequate attention in future investigations.

---

[11] See Table 1 in the Appendix for a technical overview of selected studies cited in this review.

Appendix

Table 1 Technical overview of selected studies covered in this review (by first author in alphabetical order)

Authors (Year)	Data source—micro level	Data source—macro level	Regional level	Main outcome(s)	Main explanatory variable(s)	Method
Baker and Milligan (2008)	NLSCY, 1994/95–2004/05	–	Canada	Multiple maternal and child health outcomes	Expansion of maternity leave mandates, 2000	Differences-in-differences
Bauernschuster et al. (2016)	–	Statistical Offices of the *Länder*, 1998–2009	325 West German counties	Births per 1000 women	Increase in public childcare, 2002–2009	Differences-in-differences
Cebolla-Boado et al. (2017)	PIRLS, 2011	–	28 developed countries	Reading literacy of 4th graders	Attendance of a preschool institution	Multilevel model (w/random slopes)
Dahl et al. (2016)	Multiple sources (register data)	–	Norway	Children's schooling, completed fertility, marriage, divorce	Expansions in paid parental leave, 1987–1992	Regression discontinuity design
Di Gessa et al. (2016)	SHARE, 2004/05	EU-SILC, Eurostat LFS, EVS, 2008 (aggregated)	10 Continental European countries	Provision of (intensive) grandchild care	Mothers out of employment, women in paid work, children in formal childcare, attitude towards working mothers	Multilevel analysis
Dustmann and Schönberg (2012)	Administrative data on school choices and social security records	–	Germany	Children's track choice, highest educational qualification, years of education, wages	Expansions in maternity leave coverage, 1979–1992	Differences-in-differences
Haberkern and Szydlik (2010)	SHARE, 2004/05	Multiple sources	11 Continental European countries	Receipt of (weekly) care by adult children	% Receiving home care, % in residential care, legal obligation to care, % state responsibility for care	Multilevel analysis
Kalwij (2010)	ESS, 2004 (retrospective fertility histories)	OECD Social Expenditure Database, 2007	16 Western European countries	Fertility (first and subsequent births ⇒ simulated life cycle fertility)	Family allowances, maternity- and parental-leave benefits, childcare subsidies	Monte Carlo simulation
Kneip and Bauer (2009)	–	Eurostat, 1960–2003	18 Western European countries	(Crude) divorce rates	Divorce law regime (unilateral/bilateral; de facto/de jure)	Fixed effects regression

Table 1 (Continued)

Authors (Year)	Data source—micro level	Data source—macro level	Regional level	Main outcome(s)	Main explanatory variable(s)	Method
Lalive and Zweimüller (2009)	Austrian Social Security Database	–	Austria	(Higher-order) fertility	Extensions in parental leave, 1990 and 1996	Differences-in-differences
Luci-Greulich and Thévenon (2013)	–	OECD Family, Social Expenditures and Employment Databases, 1982–2007	18 OECD countries	Total fertility rate	Spending per birth, ... on cash benefits per child, ... on childcare services, # of paid leave weeks, childcare enrolment	Two-way fixed effects regression
Patton et al. (2017)	–	OECD data, 1960–2012	19 OECD countries	Infant & post-neonatal mortality rates	Job-protected paid parental leave (weeks)	Generalized least square models (w/ year & country-fixed effects)
Raute (2018)	Vital statistics, pensions registry, microcensus	–	Germany	Fertility (giving birth at t)	Introduction of earnings-dependent maternity leave, 2007	Differences-in-differences
Rindfuss et al. (2010)	Norwegian population register, 1973–98	Norwegian Social Science Data Service, 1973–98	435 Norwegian municipalities	Fertility (first and subsequent births ⇒ simulated total # of children)	Childcare availability	Discrete-time hazard models w/ municipality-level fixed effects

References

Aassve, Arnstein, and Trude Lappegård. 2009. Childcare cash benefits and fertility timing in Norway. *European Journal of Population* 25:67–88.

Amato, Paul R. 2000. The consequences of divorce for adults and children. *Journal of Marriage and Family* 62:1269–1287.

Anckar, Carsten. 2008. On the applicability of the most similar systems design and the most different systems design in comparative research. *International Journal of Social Research Methodology* 11:389–401.

Anders, Yvonne. 2013. Stichwort: Auswirkungen frühkindlicher institutioneller Betreuung und Bildung. [Keyword: Impacts of early-childhood institutional care and education] *Zeitschrift für Erziehungswissenschaft* 16:237–275.

Andersson, Gunnar, Ann-Zofie Duvander and Karsten Hank. 2004. Do child-care characteristics influence continued child bearing in Sweden? An investigation of the quantity, quality, and price dimension. *Journal of European Social Policy* 14:407–418.

Andersson, Gunnar, Jan M. Hoem and Ann-Zofie Duvander. 2006. Social differentials in speed-premium effects in childbearing in Sweden. *Demographic Research* 14:51–70.

Arránz Becker, Oliver, Veronika Salzburger, Nadia Lois and Bernhard Nauck. 2013. What narrows the stepgap? Closeness between parents and adult (step)children in Germany. *Journal of Marriage and Family* 75:1130–1148.

Asendorpf, Jens B. 2008. Living apart together: Alters- und Kohortenabhängigkeit einer heterogenen Lebensform. [Age and cohort dependence of a heterogeneous living arrangement] *Kölner Zeitschrift für Soiologie und Sozialpsychologie* 60:749–764.

Baizán, Pau. 2009. Regional child care availability and fertility decisions in Spain. *Demographic Research* 21:803–842.

Baker, Michael, and Kevin Milligan. 2008. Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates. *Journal of Health Economics* 27:871–887.

Barnett, W. Steven. 2011. Effectiveness of early educational intervention. *Science* 333:975–978.

Baude, Amandine, Jessica Pearson and Sylvie Drapeau. 2016. Child adjustment in joint physical custody versus sole costody: A meta-analytic review. *Journal of Divorce & Remarriage* 57:338–360.

Bauernschuster, Stefan, Timo Hener and Helmut Rainer. 2016. Children of a (policy) revolution: The introduction of universal child care and its effect on fertility. *Journal of the European Economic Association* 14:975–1005.

Bellani, Daniela, Gosta Esping-Andersen and Lesia Nedoluzhko. 2017. Never partnered: A multilevel analysis of lifelong singlehood. *Demographic Research* 37:53–100.

Berninger, Ina. 2013. Women's income and childbearing in context: First births in Denmark and Finland. *Acta Sociologica* 56:97–115.

Beuchert, Louise V., Maria K. Humlum and Rune Vejlin. 2016. The length of maternity leave and family health. *Labour Economics* 43:55–71.

Bitler, Marianne P., Jonah B. Gelbach, Hillary W. Hoynes and Madeline Zavodny. 2004. The impact of welfare reform on marriage and divorce. *Demography* 41:213–236.

Blossfeld, Hans-Peter, and Johannes Huinink. 1991. Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation. *American Journal of Sociology* 97:143–168.

Blossfeld, Hans-Peter, and Andreas Timm. 1997. Der Einfluß des Bildungssystems auf den Heiratsmarkt. [The influence of the educational system on the marriage market] *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49:449–476.

Boll, Christina, Julian Leppin and Nora Reich. 2014. Paternal childcare and parental leave policies: Evidence from industrialized countries. *Review of Economics of the Household* 12: 129–158.

Bonin, Holger, Anita Fichtl, Helmut Rainer, C. Katharina Spieß, Holger Stichnoth, and Katharina Wrohlich. 2013. Zentrale Resultate der Gesamtevaluation familienbezogener Leistungen. [Central outcomes of the overall evaluation of family-related benefits] *DIW Wochenbericht* 40:3–13.

Bordone, Valeria, Bruno Arpino and Arnstein Aassve. 2017. Patterns of grandparental child care across Europe: the role of the policy context and working mothers' need. *Ageing & Society* 37:845–873.

Bovenschen, Ina, Paul Bränzel, Fabienne Dietzsch, Janin Zimmermann and Annabel Zwönitzer. 2017. *Adoptionen in Deutschland.* [Adoptions in Germany] Munich: Deutsches Jugendinstitut.

Bradley, David. 2001. Regulation of unmarried cohabitation in West-European jurisdictions—determinants of legal policy. *International Journal of Law, Policy and the Family* 15:22–50.

Brady, David, and Rebekah Burroway. 2012. Targeting, universalism, and single-mother poverty: A multi-vlevel analysis across 18 affluent democracies. *Demography* 49:719–746.

Brewster, Karin L., and Ronald R. Rindfuss. 2000. Fertility and women's employment in industrialized nations. *Annual Review of Sociology* 26:271–296.

Broadway, Barbara, Guyonne Kalb, Daniel Kuehnle and Miriam Maeder. 2017. Paid parental leave and child health in Australia. *Economic Record* 93:214–237.

Bryan, Mark L., and Stephen P. Jenkins. 2016. Multilevel modelling of country effects: A cautionary tale. *European Sociological Review* 32:3–22.

Bujard, Martin, and Jasmin Passet. 2013. Wirkungen des Elterngeldes auf Einkommen und Fertilität [Impacts of the parental allowance on income and fertility]. *Zeitschrift für Familienforschung* 25:212–237.

Bujard, Martin. 2016. Wirkungen von Familienpolitik auf die Geburtenentwicklung. [Impacts of family policy on birth developments], eds. Yasemin Niephaus et al., pp. 619–646. *Handbuch Bevölkerungssoziologie*. Wiesbaden: Springer VS.

Bünning, Mareike. 2015. What happens after the 'daddy months'? Fathers' involvement in paid work, childcare, and housework after taking parental leave in Germany. *European Sociological Review* 31:738–748.

Burger, Kaspar. 2010. How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly* 25:140–165.

Cancian, Maria, Daniel R. Meyer, Patricia R. Brown and Stephen T. Cook. 2014. Who gets custody now? Dramatic changes in children's living arrangements after divorce. *Demography* 51:1381–1396.

Carneiro, Pedro, Katrine V. Løken and Kjell G. Salvanes. 2015. A flying start? Maternity leave benefits and long-run outcomes of children. *Journal of Political Economy* 123:365–412.

Cebolla-Boado, Héctor, Jonas Radl and Leire Salazar. 2017. Preschool education as the great equalizer? A cross-country study into the sources of inequality in reading competence. *Acta Sociologica* 60:41–60.

Chamie, Joseph, and Barry Mirkin. 2011. Same-sex marriage: a new social phenomenon. *Population and Development Review* 37:529–551.

Cherlin, Andrew J. 2004. The deinstitutionalization of American marriage. *Journal of Marriage and Family* 66:848–861.

Cooke, Lynn Prince, and Janeen Baxter. 2010. "Families" in international context: Comparing institutional effects across western societies. *Journal of Marriage and Family* 72:516–536.

Cools, Sara, Jon H. Fiva, and Lars J. Kirkebøn. 2015. The causal effects of paternity leave on children and parents. *Scandinavian Journal of Economics* 117:801–828.

Courtin, Emilie, Nadia Jemiai and Elias Mossialos. 2014. Mapping support policies for informal carers across the European Union. *Health Policy* 118:84–94.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad and Kari V. Salvanes. 2016. What is the case for paid maternity leave? *Review of Economics and Statistics* 98:655–670.

Danzer, Natalia, and Victor Lavy. 2018. Paid parental leave and children's schooling outcomes. *Economic Journal* 128:81–117.

David, Henry P. 1992. Abortion in Europe, 1920–91: A public health perspective. *Studies in Family Planning* 23:1–22.

Del Boca, Daniela. 2002. The effect of child care and part time opportunities on participation and fertility decisions in Italy. *Journal of Population Economics* 15:549–573.

Del Boca, Daniela. 2003. Mothers, fathers and children after divorce: The role of institutions. *Journal of Population Economics* 16:399–422.

Di Gessa, Giorgio, Karen Glaser, Debora Price, Eloi Ribe and Anthea Tinker. 2016. What drives national differences in intensive grandparental childcare in Europe? *Journal of Gerontology: Social Sciences* 71:141–153.

Duncan, Greg J., Katherine Magnuson, Ariel Kalil and Kathleen Ziol-Guest. 2012. The importance of early childhood poverty. *Social Indicators Research* 108:87–98.

Dustmann, Christian, and Uta Schönberg. 2012. Expansions in maternity leave coverage and children's long-term outcomes. *American Economic Journal: Applied Economics* 4:190–224.

Duvander, Ann-Zofie, Trude Lappegård and Gunnar Andersson. 2010. Family policy and fertility: Fathers' and mothers' use of parental leave and continued childbearing in Norway and Sweden. *Journal of European Social Policy* 20:45–57.

Dykstra, Pearl A. 2018. Cross-national differences in intergenerational family relations: The influence of public policy arrangements. *Innovation in Aging* 2: igx032.

Eekelaar, John. 2010. Evaluating legal regulation of family behaviour. *International Journal of Jurisprudence of the Family* 1:17–34.

Ellingsæter, Anne Lise. 2009. Leave policy in the Nordic welfare states: A 'recipe' for high employment/high fertility? *Community, Work & Family* 12:1–19.

Engster, Daniel, and Helena Olofsdotter Stensöta. 2011. Do family policy regimes matter for children's well-being? *Social Politics* 18:82–124.

Esping-Andersen, Gøsta, and Francesco C. Billari. 2015. Re-theorizing family demographics. *Population and Development Review* 41:1–31.

Federal Institute for Population Research. 2017. *Facts & Figures*, accessed at www.bib-demografie.de (October 21, 2017).

Festy, Patrick. 2006. Legal recognition of same-sex couples in Europe. *Population-E* 61:417–454.

Fichtl, Anita, Timo Hener and Helmut Rainer. 2017. Familienpolitik in Deutschland: Kausale Evaluationsstudien und ausgewählte Ergebnisse. [Family policy in Germany: Causal evaluation studies and selected results] *Perspektiven der Wirtschaftspolitik* 18:117–131.

Gauthier, Anne H. 2007. The impact of family policies on fertility in industrialized countries: A review of the literature. *Population Research and Policy Review* 26:323–346.

Gauthier, Anne H., and Jan Hatzius. 1997. Family benefits and fertility: An econometric analysis. *Population Studies* 51:295–306.

Geisler, Esther, and Michaela Kreyenfeld. 2011. Against all odds: Fathers' use of parental leave in Germany. *Journal of European Social Policy* 21:88–99.

Gissler, M., I. Fronteira, A. Jahn, H. Karro, …, and the REPROSTAT group. 2012. Terminations of pregnancy in the European Union. *BJOG—International Journal of Obstetrics and Gynaecology* 119:324–332.

Goldscheider, Frances, Eva Bernhardt and Trude Lappegård. 2015. The gender revolution: A framework for understanding changing family and demographic behavior. *Population and Development Review* 41:207–239.

Goody, Jack. 1983. *The development of the family and marriage in Europe*. Cambridge: Cambridge University Press.

Grunow, Daniela. 2019. Comparative analyses of housework and its relation to paid work: Institutional contexts and individual agency. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00601-1.

Guertzgen, Nicole, and Karsten Hank. 2018. Maternity leave and mothers' long term sickness absence: Evidence from West Germany. *Demography* 55:587–615.

Gutierrez Vasquez, Edith Y., and Emilio A. Parrado. 2016. Abortion legalization and childbearing in Mexico. *Studies in Family Planning* 42:113–128.

Haberkern, Klaus, and Marc Szydlik. 2008. Pflege der Eltern – ein europäischer Vergleich [Caring for parents—a European comparison]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 60:78–101.

Haberkern, Klaus, and Marc Szydlik. 2010. State care provision, societal opinion on children's care of older parents in 11 European countries. *Ageing & Society* 30:299–323.

Haberkern, Klaus, Tina Schmid and Marc Szydlik. 2015. Gender differences in intergenerational care in European welfare states. *Ageing & Society* 35:298–320.

Hajnal, John. 1965. European marriage patterns in perspective. In *Population in History. Essays in Historical Demography,* eds. David V. Glass and David E.C. Eversley, pp. 101–143. London: Edward Arnold.

Hampden-Thompson, Gillian. 2013. Family policy, family structure, and children's educational achievement. *Social Science Research* 42:804–817.

Hank, Karsten. 2009. Generationenbeziehungen im alternden Europa: Analysepotenziale und Befunde des Survey of Health, Ageing and Retirement in Europe [Intergenerational relations in an aging Europe: research potential and findings from the Survey of Health, Ageing and Retirement in Europe]. *Zeitschrift für Familienforschung* 21:86–97.

Hank, Karsten, and Isabella Buber. 2009. Grandparents caring for their grandchildren: Findings from the 2004 Survey of Health, Ageing and Retirement in Europe. *Journal of Family Issues* 30:53–73.

Hank, Karsten, and Johannes Huinink. 2015. Regional contexts and family formation: Evidence from the German Family Panel. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 67:41–58.

Hank, Karsten, Michaela Kreyenfeld and C. Katharina Spieß. 2004. Kinderbetreuung und Fertilität in Deutschland [Childcare and fertility in Germany]. *Zeitschrift für Soziologie* 33:228–244.

Hansen, Kristine and Denise Hawkes. 2009. Early childcare and child development. *Journal of Social Policy* 38:211–239.

Harknett, Kristen, Francesco C. Billari and Carla Medalia. 2014. Do family support environments influence fertility? Evidence from 20 European countries. *European Journal of Population* 30:1–33.

Härkönen, J., Fabrizio Bernardi and Diederik Boertien. 2017. Family dynamics and child outcomes: An overview of research and open questions. *European Journal of Population* 33:163–184.

Hiekel, Nicole, Aart C. Liefbroer and Anne-Rigt Poortman. 2014. Understanding diversity in the meaning of cohabitation across Europe. *European Journal of Population* 30:391–410.

Hill, Paul, and Johannes Kopp (Eds.). 2015. *Handbuch Familiensoziologie*. Wiesbaden: Springer VS.

Hoem, Jan M. 1993. Public policy as the fuel to fertility: Effects of a policy reform on the pace of child-bearing in Sweden in the 1980s. *Acta Sociologica* 36:19–31.

Hoem, Jan M., Alexia Prskawetz and Gerda Neyer. 2001. Autonomy or conservative adjustment? The effect of public policies and educational attainment on third births in Austria, 1975–96. *Population Studies* 55:249.261.

Huang, Chien-Chung, Wen-Jui Han and Irwin Garfinkel. 2003. Child support enforcement, joint legal custody, and parental involvement. *Social Service Review* 77:255–278.

Igel, Corinne, Martina Brandt, Klaus Haberkern and Marc Szydlik. 2009. Specialization between family and state. Intergenerational time transfers in Western Europe. *Journal of Comparative Family Studies* 40:203–227.

Jappens, Maaike, and Jan Van Bavel. 2012. Regional family cultures and child care by grandparents in Europe. *Demographic Research* 27:85–120.

Kalwij, Adriaan. 2010. The impact of family policy expenditure on fertility in western Europe. *Demography* 47:503–519.

King, Valerie, and Juliana M. Sobolewski. 2006. Nonresident fathers' contribution to adolescent well-being. *Journal of Marriage and Family* 68:537–557.

Kneip, Thorsten, and Gerrit Bauer. 2009. Did unilateral divorce laws raise divorce rates in Western Europe? *Journal of Marriage and Family* 71:592–607.

Kneip, Thorsten, Gerrit Bauer and Steffen Reinhold. 2014. Direct and indirect effects of unilateral divorce law on marital stability. *Demography* 51:2103–2116.

Kravdal, Øystein. 1996. How the local supply of day-care centers influences fertility in Norway: A parity-specific approach. *Population Research and Policy Review* 15:201–218.

Kreyenfeld, Michaela, and Karsten Hank. 2000. Does the availability of child care influence the employment of mothers? Findings from western Germany. *Population Research and Policy Review* 19:317–337.

Kreyenfeld, Michaela, and Valerie Martin. 2011. Economic conditions of stepfamilies from a cross-national perspective. *Zeitschrift für Familienforschung* 23:128–153.

Kuehnle, Daniel. 2014. The causal effect of family income on child health in the UK. *Journal of Health Economics* 36:137–150.

Lalive, Rafael, and Josef Zweimüller. 2009. How does parental leave affect fertility and return to work? Evidence from two natural experiments. *Quarterly Journal of Economics* 124:1363–1402.

Laroque, Guy, and Bernard Salnié. 2004. Fertiltiy and financial incentives in France. *CESifo Economic Studies* 50:423–450.

Lauer, Sean, and Carrie Yodanis. 2010. The deinstitutionalization of marriage revisited: A new institutional approach to marriage. *Journal of Family Theory and Review* 2:58–72.

Lesthaeghe, Ron. 2010. The unfolding story of the second demographic transition. *Population and Development Review* 36:211–251.

Liefbroer, Aart C., Anne-Rigt Poortman and Judith A. Seltzer. 2015. Why do intimate partners live apart? Evidence on LAT relationships across Europe. *Demographic Research* 32:251–286.

Liu, Qian, and Oskar Nordstrom Skans. 2010. The duration of paid parental leave and children's scholastic performance. *The B.E. Journal of Economic Analysis & Policy* 10.

Luci-Greulich, Angela, and Olivier Thévenon. 2013. The impact of family policies on fertility trends in developed countries. *European Journal of Population* 29:387–416.

Magnuson, Katherine A., Christopher Ruhm and Jane Waldfogel. 2007. Does prekindergarten improve school preparation and performance? *Economics of Education Review* 26:33–51.

Miller, Grant and Christine Valente. 2016. Population policy: abortion and modern contraception are substitutes. *Demography* 53:979–1009.

Motel-Klingebiel, Andreas, Clemens Tesch-Römer and Hans-Joachim von Kondratowitz. 2005. Welfare states do not crowd out the family: evidence for mixed responsibility from comparative analyses. *Ageing & Society* 25:863–882.

Noack, Trude, Eva Bernhardt and Kenneth A. Wiik. 2014. Cohabitation or marriage? Preferred living arrangements in the West. In *Contemporary issues in family studies: global perspectives on partnerships, parenting, and support in a changing world,* eds. Angela Abela and Janet Walker, pp. 16–30. Oxford: Wiley-Blackwell.

Patton, Dana, Julia F. Costich and Niklas Lindstromer. 2017. Paid parental leave and infant mortality rates in OECD countries: Policy implications for the United States. *World Medical & Health Policy* 9:6–23.

Perelli-Harris, Brienna, and Nora Sánchez Gassen. 2012. How similar are cohabitation and marriage? Legal approaches to cohabitation across Europe. *Population and Development Review* 38:435–467.

Perelli-Harris, Brienna, Ann Berrington, Nora Sánchez Gassen, Paulina Galezewska and Jennifer A. Holland. 2017. The rise in divorce and cohabitation: Is there a link? *Population and Development Review* 43:303–329.

Pong, Suet-Ling, Jaap Dronkers and Gillian Hampden-Thompson. 2003. Family policies and children's school achievement in single- versus two-parent families. *Journal of Marriage and Family* 65:681–699.

Popenoe, David. 1993. American family decline, 1960–1990: A review and appraisal. *Journal of Marriage and the Family* 55:527–542.

Präg, Patrick, and Melinda C. Mills. 2017. Assisted reproductive technologies in Europe: Usage and regulation in the context of cross-border reproductive care. In *Childlessness in Europe: Contexts, causes, and consequences*, eds. Michaela Kreyenfeld and Dirk Konietzka, pp. 289–309. Dordrecht: Springer.

Raley, R. Kelly. 2001. Increasing fertility in cohabiting unions: Evidence for the second demographic transition in the United States? *Demography* 38:59–66.

Raute, Anna. 2018. Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits. *Journal of Public Economics*: forthcoming.

Reinhardt, Jörg. 2017. *Rechtliche Grundlagen des Adoptionswesens in Deutschland im internationalen Vergleich* [The legal basis for the adoption system in Germany in an international comparison]. Munich: Deutsches Jugendinstitut.

Rindfuss, Ronald R., David Guilkey, S. Philip Morgan and Øystein Kravdal. 2010. Child-care availability and fertility in Norway. *Population and Development Review* 36:725–748.

Rønsen, Marit. 2004. Fertility and public policies—Evidence from Norway and Finland. *Demographic Research* 10:265–286.

Rosenbaum, Heidi. 2014. Familienformen im historischen Wandel [Forms of families as they have changed over time]. In *Familie im Fokus der Wissenschaft,* eds. Anja Steinbach et al., pp. 19–39. Wiesbaden: Springer VS.

Rossin, Maya. 2011. The effects of maternity leave on children's birth and infant health outcomes in the United States. *Journal of Health Economics* 30:221–239.

Rossin-Slater, Maya. 2018. Maternity and family leave policy. In *Oxford handbook on the economics of women*, eds. forthcoming in Susan L. Averett et al. New York: Oxford University Press [advance online access].

Rothwell, David W., and Annie McEwen. 2017. Comparing child poverty by family structure during the 2008 recession. *Journal of Marriage and Family* 79:1224–1240.

Rupp, Marina, and Christian Haag. 2016. Gleichgeschlechtliche Partnerschaften und Familien [Same-sex partnerships and families]. In *Handbuch Bevölkerungssoziologie*, eds. Yasemin Niephaus et al., pp. 327–345. Wiesbaden: Springer VS.

Saraceno, Chiara, and Wolfgang Keck. 2010. Can we identify intergenerational policy regimes in Europe? *European Societies* 12:675–696.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. *International comparative social research*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. https://doi.org/10.1007/s11577-019-00607-9.

Schütz, Gabriela. 2009. Does the quality of pre-primary education pay off in secondary school? An international comparison using PISA 2003. *Ifo Working Paper No. 68*. Ifo Institute for Economic Research, University of Munich.

Seltzer, Judith A. 1998. Father by law: Effects of joint legal custody on nonresident fathers' involvement with children. *Demography* 35:135–146.

Smeeding, Timothy, and Céline Thévenot. 2016. Addressing child poverty: How does the United States compare with other nations? *Academic Pediatrics* 16:S67–S75.

Spieß, C. Katharina, Felix Büchel and Gert G. Wagner. 2003. Children's school placement in Germany: Does Kindergarten attendance matter? *Early Childhood Research Quarterly* 18:255–270.

Steinbach, Anja. 2018. Children's and parents' well-being in joint physical custody: A literature review. *Family Process* [advance online access].

Steinbach, Anja, and Karsten Hank. 2016. Intergenerational relations in older stepfamilies: A comparison of France, Germany, and Russia. *Journal of Gerontology: Social Sciences* 71:880–888.

Steinbach, Anja, Anne-Kristin Kuhnt and Markus Knüll. 2016. The prevalence of single-parent and step-families in Europe: Can the Hajnal line help us to describe regional patterns? *The History of the Family* 21:578–595.

Stirling, Kate and Tom Aldrich. 2008. Child support: Who bears the burden? *Family Relations* 57:376–389.

Swiss, Liam, and Céline Le Bourdais. 2009. Father-child contact after separation. The influence of living arrangement. *Journal of Family Issues* 30:623–652.

Tanaka, Sakiko. 2005. Parental leave and child health across OECD countries. *Economic Journal* 115:F7–F28.

Trappe, Heike. 2017. Assisted reproductive technologies in Germany: A review of the current situation. In *Childlessness in Europe: Contexts, causes, and consequences,* eds. Michaela Kreyenfeld and Dirk Konietzka, pp. 269–288. Dordrecht: Springer.

Treas, Judith, Jonathan Lui and Zoya Gubernskaya. 2014. Attitudes on marriage and new relationships: Cross-national evidence on the deinstitutionalization of marriage. *Demographic Research* 30:1495–1526.

Vaskovics, Lazlo A., and Johannes Huinink. 2016. Werden die Regelungen des Familienrechts heutigen Familien und Kindern gerecht? [Do the provisions of family law do justice to today's families?] *Zeitschrift für Familienforschung* 28:221–244.

Vikat, Andres. 2004. Women's labor force attachment and childbearing in Finland. *Demographic Research* 3:177–212.

Waaldijk, Kees. 2009. Overview of forms of joint legal parenting available to same-sex couples in European countries. *Droits & Société* 72:383–385.

Wellenhofer, Marina. 2016. Regelungsaufgabe Paarbeziehungen: Was kann, was darf, was will der Staat? [Marriage, civil partnership and cohabitation: what can, what should, what does the state do?] *Zeitschrift für Familienforschung* 28:162–177.

Werding, Martin. 2013. Children are costly, but raising them may pay: The economic approach to fertility. *Demographic Research* 30:253–276.

Willekens, Harry. 2003. Is contemporary western family law historically unique? *History of the Family* 28:70–107.

Zoch, Gundula, and Irina Hondralis. 2017. The expansion of low-cost, state-subsidized childcare availability and mothers' return-to-work behaviour in East and West Germany. *European Sociological Review* 33:693–707.

Karsten Hank Professor of Sociology at the University of Cologne and Research Fellow at the German Institute for Economic Research (DIW), Berlin. He studied Social Sciences at Ruhr-University Bochum, was a doctoral student at the Max Planck Institute for Demographic Research in Rostock, and Head of the Research Unit "Survey of Health, Ageing and Retirement in Europe" at the Mannheim Research Institute for the Economics of Aging. Hank was a Visiting Professor at King's College London and an International Research Fellow at VU Amsterdam. He is Co-PI of the German Family Panel (pairfam). Recent publications include articles in *Demography*, *European Sociological Review*, *Journals of Gerontology:Social Sciences* as well as the *Journal of Marriage and Family*.

Anja Steinbach Professor of Sociology at the University of Duisburg-Essen. She studied sociology at the University of Leipzig and obtained a PhD as well as her Habilitation in Sociology from the Technical University of Chemnitz. She was a visiting researcher at Syracuse University's Aging Studies Institute and is a Board Member of the ISA-Committee on Family Research. Recent publications include articles in the *Journal of Family Issues*, *Journals of Gerontology: Social Sciences*, *Journal of Marriage and Family* as well as *Social Science Research*.

Kölner Zeitschrift für Soziologie und Sozialpsychologie

Party Competition and Vote Choice

Dennis Christopher Spies · Simon T. Franzmann

Abstract Analyzing the relationships between political parties and voters is one of the central topics of political science. Parties are expected either to be responsive to the demands of their voters or are presumed to have the power to influence voting decisions by structuring the political discourse and thereby competition regarding political issues. These two aspects are covered in the literature by research on the way parties present themselves and by electoral research, respectively. Focusing on the latter, this state-of-the-art article reviews how recent publications have analyzed the impact of party competition (macro level) on vote choice (individual level). It does so by introducing the most prominent theories of voting and party competition, summarizing the most recent results and pointing to potential problems for international comparisons such as methodological choices and different approaches to the measurement of party positions.

Keywords Manifesto data · Vote choice · Positional theory · Issue salience · Expert surveys

Parteienwettbewerb und Wahlverhalten

Zusammenfassung Die Analyse der Beziehungen zwischen politischen Parteien und Wählern ist eines der zentralen Themen der Politikwissenschaft. Hinsichtlich der Parteien wird davon ausgegangen, dass sie sich entweder responsiv gegenüber den

D. C. Spies (✉)
Cologne Center for Comparative Politics, University of Cologne
Herbert-Lewin-Str. 2, 50931 Köln, Germany
E-Mail: spies@wiso.uni-koeln.de

S. T. Franzmann
Social Science Department, Comparative Politics, Heinrich-Heine University Düsseldorf
Universitätsstr. 1, 40225 Düsseldorf, Germany
E-Mail: simon.franzmann@hhu.de

🙌 Springer

Forderungen ihrer Wähler zeigen oder aber die Macht haben, deren Wahlentscheidung zu beeinflussen, indem sie den politischen Diskurs und damit den Wettbewerb um politische Themen strukturieren. Diese beiden Aspekte werden zum einen in der Parteien-, zum anderen in der Wahlforschung behandelt. Mit Blick auf Letztere wird in diesem State-of-the-Art-Artikel dargestellt, wie neuere Veröffentlichungen die Auswirkungen des Parteienwettbewerbs (Makroebene) auf die Wahlentscheidung (Individualebene) analysieren. Dies geschieht durch eine Einführung in die wichtigsten Theorien zur Wahlentscheidung und zum Parteienwettbewerb, auf deren Grundlage eine Zusammenfassung der Ergebnisse neuerer, quantitativer Studien erfolgt. Potenzielle Probleme für den internationalen Vergleich werden hierbei ebenso erörtert wie wichtige methodische Neuerungen und die verschiedenen verfügbaren Ansätze zur Messung von Parteienpositionen.

Schlüsselwörter Parteiprogrammdaten · Wahlverhalten · Räumliche Politikmodelle · Themensalienz · Expertenumfragen

1 Introduction

The analysis of electoral behavior is undoubtedly one of the core topics of political science, since this is where citizens' political preferences are regularly translated into the selection of political personnel. More than other forms of political participation, elections may therefore be regarded as "instruments of democracy" (Powell 2000). While the individual vote decision depends on a variety of factors, it always takes place in an electoral context that is limited in time and place. This context includes not only the electoral system and the system of government (for both see Schmitt-Beck 2019), but also the party system. Parties are expected either to be responsive to the demands of their voters, or are presumed to have the power to influence voting decisions by structuring the political discourse and thereby competition between parties as to political issues. Focusing on the latter, this state-of-the-art article reviews how recent publications have analyzed the impact on vote choice that is exerted by party competition.

Below, we define party competition as institutionally structured interactions, in which political parties strategically cooperate or battle to gain power (Franzmann 2011, p. 320). As attracting voters is crucial, we will focus on how the two most prominent theories of party competition—*positional* and *salience theory*—are related to the individual vote choice. We point out that the two major theories make quite different arguments about how parties behave and interact with each other in order to attract voters, yet for both, the interaction is based on political issues rather than on candidates or other non-thematic criteria.[1] Most recently, combinations of positional and salience theory have been successfully applied in analyzing the transforming party competition across Europe. We discuss these innovations as mixed approaches.

---

[1] This is not to say that non-issue motivations are not relevant to voters. However, as far as the party competition literature is concerned, parties address voters exclusively via political issues.

Regarding the individual vote decisions, we categorize the encompassing theoretical literature as highly rational theories of *spatial voting*, competence-based theories of *issue voting,* and theories pointing to the role of *cleavages,* in order to explain voting patterns. After introducing each of these theories, we summarize the results of the most recent studies, and point to important aspects for analyzing the effect of party competition on voters' choice in an international perspective, including methodological choices and the measurement of party positions. We have concentrated on the theoretical arguments and empirical results of the most recent studies making international comparisons, using quantitative methods to explain individual vote choice. This might be the decision for a distinct party family (e.g., radical right parties), for an incumbent party (in contrast to an opposition party), or for a party that is closest to the voter's own position (highly relevant for spatial voting theories).

2 Theories of Vote Choice

Scholarly studies of electoral behavior have a long and lively history, and it is not our aim to provide an encompassing overview of the literature here (but see Schmitt-Beck 2019). Rather, we focus on three of the arguably most influential theories of vote choice: spatial, issue, and cleavage voting. These grand theories have been especially important in shaping the current contours of electoral studies and are also the main point of reference for those comparative studies concerning the impact on individual vote choice that is exerted by party competition.

2.1 Spatial Voting

Spatial voting theory is important among theories of vote choice and candidate preferences (Downs 1957; Enelow and Hinich 1984; Rabinowitz and Macdonald 1989). It deals with rational choice in the sense that voters' preferences among parties are assumed to be representable by a utility function, which suggests that voters have a preferred position in a predefined issue space. Parties express their positions on the same issue space, and it is assumed that voters have some information about these positions. In their general form, spatial voting theories can be expressed for each political issue (e.g., welfare spending, values, immigration policy, etc.), but voters' and parties' positions are usually expressed in terms of left and right, providing something like a super dimension (Gabel and Huber 2000) by which political positions can be measured.[2] While these assumptions are shared by all spatial voting theories, they can be further divided into *proximity* and *directional theories*, depending on what the utility function of a voter looks like. The utility of voting for a party defined by proximity theory is the following:

$$u_i(v_i, p_j) = -(v_i - p_j)^2$$

---

where v_i is the position of voter i on the left–right ideological scale, u_i is his utility, and p_j is the position of party j in question on the same scale. It is easy to see that the utility of each voter reaches its maximum when the positions of voter i and party j overlap, i.e., they have the same position in terms of left and right. The neutral position or the middle of the scale has no specific meaning or importance in proximity theory. If a voter is on the left of the scale, but the most proximate party is on the right, the voter will still prefer that party, irrespective of the fact that they are on different sides. In contrast, directional theory (Rabinowitz and Macdonald 1989) builds on this differentiation so that utility of the voter is defined as:

$$u_i(v_i, p_j) = (v_i - n)(p_j - n)$$

with n representing the ideological middle, or the point of neutrality between left and right. In contrast to proximity theory, directional theory uses a two-step rationale (Westholm 1997). Voters first choose a side—such as for or against an issue or left vs. right ideology—and select the party that conforms most closely. The choice of the most extreme party on the same side as the voter will generate the highest utility for the voter, but if there is no party on the same side, he or she will choose the party on the other side that is the least extreme.

While both proximity and directional theories have been tested extensively (see for example Adams et al. 2005; Blais et al. 2001; Kramer and Rattinger 1997; Pierce 1997; Westholm 1997), it is virtually impossible to compare their explanatory power. The reason for this is that in a considerable number of cases, both theories come to the same prediction about which party a voter will prefer (Tomz and Van Houweling 2008). Another potential problem for spatial voting theories is the a priori choice of the dimension on which both voters and parties are located. As electoral competition is not necessarily unidimensional, the left–right ideological continuum might not accurately describe the positions held by the majority of the electorate. Also, left and right might have very different meanings across both time and space, resulting in potential problems, particularly for international comparisons. Finally, there is the risk that voters adopt biased judgments about the positions offered by parties, seeing parties that they like as being closer to their own position rather than parties that they dislike. We will discuss these potential problems in greater detail later.

2.2 Issue Voting

In contrast to spatial theories, *issue voting theories* do not see voters and parties taking up predefined positions. Rather, voters have a certain idea about which issues are most important to them, e.g., seeing rising inflation as being a more pressing problem than wage inequality. After ranking issues in that way, they then decide which party might be most competent in handling these issues and cast their vote accordingly (Ansolabehere and Iyengar 1994; Petrocik 1996). Assigning competence to parties relies on the assumption that parties "own" certain issues, e.g., right-wing parties are seen to be more competent in addressing inflation, while left-wing parties enjoy greater trust when it comes to addressing wage inequality. According to issue ownership theory, parties thus have a strong incentive to emphasize issues with

which they are notionally connected (e.g., Belanger and Meguid 2008; Nadeau et al. 2001; Van der Brug 2004), i.e., to increase their *salience* but to downplay issues owned by competing parties. As with spatial theories, issue voting theories inhibit statements on how political parties should behave in order to address voters.[3]

Also relevant for issue voting theories is the distinction between positional and valence issues. Issue ownership theory (Petrocik 1996) was initially developed and tested for valence issues, i.e., those on which all voters and parties share the same goal, such as reducing unemployment or fighting crime. Depending on the salience of unemployment and crime, voters will then support the party they see as most competent in handling the problem. However, not all issues are valence issues, and more recent research acknowledges positional issues on which both voters and parties can disagree (e.g., Belanger and Meguid 2008; Walgrave et al. 2012). While the relationship between valence and positional issues is still a major gap in the literature, the classical view of them as contradictory (Stokes 1963, 1992) is not shared by many recent studies. Linking valence to positional issues, Pardos-Prado (2012) argues that voters may believe that a party is best equipped to deal with a given issue because they share the parties' position on it and further differentiate between issue goals (which can be consensual) and means to reach the goals (which can be positional).

Issue voting theories have been developed in response to the low explanatory power of spatial theories, and argue that it is much easier for voters to decide questions of issue competence rather than calculating proximities between their own position and those of the parties (Green and Hobolt 2008). In other words, voters are expected to care much more about what parties can deliver in terms of addressing political problems than about calculating proximities. For this reason, we also discuss here *performance voting theories* as a sub-type of issue voting theories. Performance voting studies are based on the idea that voters reward or punish incumbents for past behavior and focus on retrospective evaluations of government economic performance as important determinants of voting behavior (see, e.g., Fiorina 1981; Miller and Shanks 1996). Central to this literature is the intuition that citizens sanction incumbents based on their evaluations of a government's policy record, especially in the area of economics. To date, a large body of work finds evidence in support of the performance voting models, demonstrating a strong relationship between economic performance and incumbent support (Lewis-Beck and Stegmaier 2007). While these *economic voting* theories are surely the most prominent in the field of performance studies (see Schmitt-Beck 2019), more recent research also analyzes how they feature in other policy areas (de Vries and Giger 2014).

2.3 Cleavage Voting

We end our list of the three most important voting theories by discussing the *cleavage theory* of voting. Compared with both spatial and issue voting theories, this one is surely the least demanding in terms of voter rationality and informational level.

---

[3] Recent findings support the hypothesis that reciprocal effects between the core electorate and their preferred party lead to the establishment of issue ownerships (Neundorf and Adams 2016).

Voters are not expected to calculate political distance, nor to judge governments' performances, but are assumed to support a political party because of its traditional link to a certain social group defined by long-term social divisions, known as cleavages. Initially developed as a macro-level theory by Lipset and Rokhan (1967), who differentiated between four traditional cleavages, class and religion are now seen as by far the most important. The literature on class-based (Evans et al. 1999) and religious (Norris and Inglehart 2004; Knutsen 2004) voting can thus be seen as sub-types of cleavage voting theories, and both stress that voters tend to engage in long-term relationships with political parties (Tóka and Gosselin 2010), and thus follow a much more stable voting pattern than the other two theories predict.

However, regarding explanatory power, scholars interested in the effects of both class and religion often report that the ability of these variables to explain voter choice is declining (Dalton 2002; Dogan 1995; Franklin et al. 1992). It is argued that long-term changes in the electorate, such as changing labor markets, deindustrialization, or secularization, have eroded group cohesion formed around class or denomination, thereby reducing their relevance for electoral behavior. However, other studies demonstrate relatively stable associations between class (Evans et al. 1999), religion (Elff 2007), and vote choice. Furthermore, it has been suggested that the prevalence of cleavages depends heavily on the strategies adopted by political parties, which can choose to stick to their traditional electorates or downplay cleavages in order to attract wider electoral groups via a "catch-all" strategy (Achterberg 2006; Evans and Tilley 2012). It is therefore important to note from a theoretical perspective that cleavage theory—as with spatial and issue voting theories—makes strong initial assumptions about the weight that individual voters place on party behavior.

3 Theories of Party Competition

As indicated in the previous section, each of the three dominant theories of vote choice already makes assumptions about how political parties should act in order to address voters' concerns. Thus, parties are seen as crucial to theories of vote choice, as they define the supply side of electoral competition required to address the demands of voters. In doing so, they have to take into account the behavior of rival parties. Theoretically, there are two perspectives, namely *positional* and *salience,* which determine how competition between parties might affect the individual voter's decision. Importantly, both are not only distinct theories of vote choice but also distinct approaches as to how party competition should operate and be measured. Also, these theories of party competition are connected, and more recent studies have made several attempts to unify them. For the moment, we introduce positional and salience theories as distinct theoretical approaches.

3.1 Positional Theory

The idea of the *positional theory* of party competition is that the policy platform of any party can be described by its particular position in a predefined political space.

Parties take distinct positions within this space. Budge (2001) has characterized positional theory as a "confrontational theory of party competition," since parties are assumed to always talk confrontationally on the same issues. In the original version by Downs (1957), the political space was defined by economic concerns of a free market vs. a state-oriented economy. Parties can freely take up any position in this space, but as purely and highly rational vote-seekers, they have a strong incentive to adapt their position to the given voter distribution. This is a consequence of the general assumption of parties as vote-seekers: "parties formulate policies in order to win elections, rather than win elections to formulate policies" (Downs 1957, p. 28).[4]

According to positional theory, in a two-party system, rational parties have a strong incentive to locate themselves close to the median voter, dividing the electorate into two spheres. Competition between parties is then centripetal and restricted to the middle space (Pappi 2000). The optimal position of parties is far less straightforward in multi-party systems, as they also have to compete with more extremist parties. Competition among them might then be centripetal or centrifugal, i.e., taking place at the extremes of the left–right dimension (Sartori 1976). In any case, the positioning of parties relative to each other determines the votes that they can attract (Pappi 2000).

Positional theory is clearly related to spatial voting theory. In fact, both refer to the initial study of Downs's *Economic Theory of Democracy* (1957), in which the behavior of voters and parties is described in spatial terms. Both theories rely on the assumption of a very high level of information: parties have to know where voters are located, and voters have to know which positions parties are taking. Furthermore, voters should be able to discriminate between the positions of parties in order to calculate and compare distances between their own position and those of all competing parties—regardless of whether they apply a proximity or directional calculation. Nevertheless, Downs himself had already discussed the role of uncertainty in party competition that might lead to "irrational" election results and governmental decision-making (Downs 1957, pp. 77–95).[5] The most pressing critique of positional theories focuses on these assumptions and points to two problems with which voters might be faced when comparing parties in spatial terms. Both points of criticism are related to the supply side of electoral competition as provided by parties, and thus directly concern our definition of party competition.

The first concern is related to the distribution of parties on the left–right scale. In some cases, these positions might be clearly separable for voters, e.g., when a communist is facing a conservative party in a two-party system. Such separations are arguably much more difficult in other multi-party settings with many rival parties, and demand a very high degree of political information; think, for example, of the Dutch and Israeli systems with more than ten relevant parties. The proxy usually employed for these supply-side characteristics is party-system *polarization*, understood as a measure of the spread of parties along the left–right ideological continuum

---

[4] Nowadays, the literature on party behavior sees parties as simultaneous policy-, office-, and vote-seekers (Strom 1990).

[5] A recent study of Ezrow et al. (2014) shows that voters abstain from voting for parties where they are uncertain as to their position.

(Dalton 2008). For Sartori (1976), party-system polarization is a measure of ideological differentiation. As this polarization is the central variable for positional theory, we have applied the most commonly used formula, which is basically a variance formula, following Taylor and Herman (1971):

$$\sum_{i=1}^{N} \pi \left(X_i - \overline{X} \right)^2$$

Here, the polarization measure for a party system with N number of parties is defined by the weight attached to party i given by its relative vote share at the time of the election observed, π, the left–right weighted mean of the parties' placement on the left–right scale $\overline{Xe}$, and the left–right position of the party on the same scale, X_i. Previous research shows that in countries with more polarized party systems, spatial voting theories are a better description of the voter's preferences (van der Eijk et al. 2005; Lachat 2008; Pardos-Prado and Dinas 2010; Dalton 2008), because highly polarized systems make it easier for the individual voter to identify differences between the parties. However, many empirical studies regarding polarization lack conceptual clearance. Combining a variance-based polarization measurement with Sartori's proposal of simply taking the range between the two extremes of the relevant parties seems to reveal results of higher validity in empirical studies (Schmitt and Franzmann 2018).

The second concern with regard to positional theory is that the political offerings as indicated by the party system polarization formula might not be one dimensional. Thus, while voters might be able to compare parties on a one-dimensional scale in terms of left and right, it is much harder, if not impossible, for them to assess parties with two, three, or even more dimensions. Recent comparative party research indeed shows that most countries can best be described by a multi-dimensional political space, including, e.g., an economic dimension and a cultural dimension of competition (Enyedi and Deegan-Krause 2010; Kitschelt 1994; Kriesi 2010). This is of limited relevance as long as these single dimensions are parallel, i.e., the position of a party in one dimension closely resembles its position in the other. However, if the distinct dimensions are orthogonal to each other, i.e., parties take very different positions in them, voters might be unable to summarize the positions into only one super dimension by which they can compare parties. The number of relevant dimensions is thus crucial for spatial theories to work; more dimensions mean a potential loss in explanatory power.

3.2 Salience Theory

The second major theory of party competition is *salience theory*, as originally defined by Robertson (1976). The theory relies on parties selecting one issue for emphasis and can be seen as a direct response to positional theory. One of the oldest criticisms of the Downsian approach was its neglect of non-confrontational issues, i.e., issues where almost the whole electorate is in agreement, such as fighting unemployment. Stokes (1963) has labeled these "valence issues," and salience theory argues that parties do not compete on them in spatial terms, but rather selectively emphasize such

issues as their "own," i.e., the ones they are seen as most competent in addressing. Taking inflation as an example, right-wing parties are usually seen as more competent in addressing rising prices and therefore tend to emphasize this issue much more than left-wing parties, who themselves might highlight questions of social inequality and refuse to speak about inflation—and vice versa. In this way, parties signal to voters engaging in issue voting which issues are most important in an election, and have a strong incentive to emphasize issues that they own but to ignore those owned by other parties (Budge and Farlie 1978; Budge 2015).

Salience theory does not engage in locating parties (or voters) in a spatial way, but is interested in measuring the degree to which a party emphasizes a specific issue relative to others, i.e., how salient this issue is for the party. In orthodox salience theory, issue salience is not related to an ideological direction. Distinguishing between distinct issues is at the core of salience theory, an approach standing in sharp contrast to the tendency of summarizing all issues into one single dimension, as followed by positional theory. Separating electoral competition into distinct issues, salience theory also seems more compatible with cleavage voting theories, which rely on the assumption of several underlying political conflicts rather than on the one-dimensionality of electoral competition. However, compared with positional theory, salience theory appears as the less self-contained approach because parties cannot be entirely free to emphasize certain issues, but rather have to react to the wider socioeconomic context. For example, it will be difficult for a left-wing party to ignore the issue of rising prices when inflation is at a historic high. In such a context, voters will expect left-wing parties to say something about inflation, even if this means that they emphasize an issue that is traditionally owned by right-wing parties. In contrast to spatial positioning, emphasizing issues depends much more on contextual effects outside the electoral arena.

4 Measuring Party Competition in Comparative Studies

There is no uncontested gold standard on how to measure parties' policies and the emphasis they chose to place on given issues. Each of the dominant methods—expert surveys, mass surveys, manifestos, computer-assisted text analysis, and media data—certainly has its own advantages and disadvantages, which have been discussed in great detail in the comparative party literature (Benoit and Laver 2007; Budge and Pennings 2007). In the following, we summarize the most important findings of these debates, especially pointing to problems arising for international comparisons. We illustrate these problems by focusing on the left-right dichotomy, which is the standard approach in the literature when it comes to measuring parties' policy positions.

Expert surveys have been one of the commonest datasets applied to party positions since the 1980s because they are relatively cheap. Another major advantage of all expert surveys is the use of standardized and often very detailed questionnaires monitored by a country's scientists. As country experts, they usually bring to light such things as coalition signals, political rhetoric, etc., which are seldom captured by other methods. At the same time, however, several problems arise from the

question of how neutral and objective such experts are, as even scientists may have some sympathy or antipathy for particular political parties which will influence their judgments. Consequently, the expert survey carried out by Laver and Hunt (1992), and its successor by Benoit and Laver (2006), include a sympathy score of each coder for each party, enabling a correction of their judgments. In general, expert surveys also suffer from not providing an extensive time series. Further, if it is not the same coder who gathers the information, this could affect reliability. Fortunately for scholars who are interested in European politics, the Chapel Hill Expert Survey (CHES) started in 1999 has become the most widely used dataset. Its quality regarding validity and reliability is permanently controlled, and it has been updated every four years since 2002. Beginning with 19 countries in 1999, in 2014, CHES provided data for all EU member states plus Turkey, Switzerland, and Norway. As with the datasets of Laver and Hunt and Benoit and Laver, multiple policy dimensions are covered. In addition, not only positions but also saliences of policy domains are estimated.

Another very common source for estimating party positions are mass surveys, for which respondents are asked to indicate their own position and those of the major parties on a predefined scale, usually in terms of left and right. The answers are then summarized in the form of a national mean. While such items can easily be included in surveys, this approach is also not without its problems. Most importantly, individuals might not share a common understanding of the underlying scale, some, for example, having economic or other cultural leanings when indicating a right- or left-wing position. Further, country peculiarities might make it difficult to compare left–right scores between countries and over time—a problem that is also relevant for expert surveys. Finally, mass surveys often lack salience measurements. Nevertheless, surveys on general elections often at least ask about the most important problems and comprise extensive batteries of issues.

Perhaps the most important data source for longitudinal cross-country analysis is the Manifesto Research Group on Political Representation (MARPOR; formerly the Comparative Manifesto Project, CMP). It follows a strict salience-based approach, coding so-called quasi-sentences. A quasi-sentence is the smallest unit of issue emphasis (Budge 2001). Each quasi-sentence is allocated to one of 57 predefined categories belonging to seven main policy dimensions. The data expose what percentage of a manifesto is devoted to an issue represented by one of the 57 categories (including one residual category). Thus, these salience scores vary from 0 to 100%. The main advantage of the MARPOR data is their continuous re-codability, combined with wide coverage in time and space. There is no other dataset going back to 1945 in almost all Organisation for Economic Co-operation and Development (OECD) countries. One of the major disadvantages of MARPOR is that only predefined categories are covered. The more new issues arise, the less suitable the codebook becomes for covering current policies. While the MARPOR data come in the form of issue saliences, there are several ways to generate left-right party positions from them (e.g., Lowe et al. 2011). The most straightforward is to define some categories as left or right and then add up their percentages. This is exactly what the long-established RILE (right–left) scale does. Here, classifying issue categories as

right or left is derived from Marxist theory, partly inductively and partly deductively (Budge and Pennings 2007).

MARPOR relies on classical hand coding, but Wordscores (Laver et al. 2003) and Wordfish[6] (Slapin and Proksch 2008) apply algorithms to judge party positions via computer coding. Wordscores work as a variant of correspondence analysis (Lowe 2008), which is a novel approach to extracting dimensional information from political texts using computerized content analysis, allocating party statements within a dictionary-based space in comparison to a reference text. Thus, the reliability and validity of this approach crucially depends on the chosen reference. Wordfish does not need such a reference text or a dictionary. It relies on a statistical model of word counts. Frequencies are used to place documents on a single dimension. Hence, the reliability and validity of this approach relies on which texts are considered. Data quality is higher when the results are stable despite changes in the text corpus. No wide-ranging inter-country comparable dataset exists for either approach.

Another method for determining left–right positions is the analysis of media content. Notably, inter-country comparisons were conducted by Kriesi et al. (2006). In order to analyze the changing political scene in Europe from the 1970s onwards, the analysis of the first paragraph of relevant newspaper articles was examined, coding sentence by sentence. The aim was to detect relationships between those involved in politics and political issues (Kriesi et al. 2006, p. 931). The main disadvantage of this approach (as well as of MARPOR) is its resource-intensive coding process. Data are available for only limited points in time and for six countries.

While each of the described methods leads to specific problems when applied to cross-national and longitudinal comparisons, one circumstance that all approaches have to take into account is that party competition always takes place within a national context. This means that parties' positioning as well as the emphasis that they place on a single issue ask for a place- and time-specific interpretation. Especially the umbrella terms of "left" and "right" can cover different issues in different countries. Left can be defined as seeking to change the status quo towards greater equality, right towards greater inequality or keeping the current state of inequality (Inglehart 1984). The same issue can therefore have different left–right meanings in two countries, thus making comparisons difficult. For instance, introducing a very low minimum wage within a laissez-faire economy might be seen to be a classical left position, while introducing the same policy within a coordinated economy might be considered a right position. Such problems of measurement pertain to each of the methods presented here, and not all approaches allow country- and time-specific meanings of left and right to be disentangled. The most sophisticated approaches are available for the MARPOR data, whereas the standard RILE approach is widely criticized for not considering the country- and time-specific meaning of left and right (Benoit and Laver 2007). Methods that consider context-specific information reveal much better results than the standard RILE approach (Gabel and Huber 2000; Franzmann 2015). Context sensitivity can be fulfilled by the country- and time-specific

---

[6] A detailed description of the Wordscores approach can be found at: https://www.tcd.ie/Political_Science/wordscores/index.html. Wordfish's project website is: http://www.wordfish.org/. Both websites also provide the program codes ready for application.

determination of the left–right character of each category (Franzmann and Kaiser 2006), by multi-dimensional scaling (Jahn 2011), bridge observations and Bayesian factor analysis (Koenig et al. 2013), or Bayesian latent trait models (Elff 2013). Therefore, because they offer the widest range across time and space, MARPOR data are still the main authoritative sources to judge party policies when making international comparisons.

5 Methodological State-of-the-Art

As our summary of articles in the next section will show, recent comparative studies predominately rely on multilevel models when analyzing the effects of party competition on the individual vote choice. This use of sophisticated statistical tools portrays a general trend in electoral research (see Schmitt-Beck 2019), acknowledging the fact that individual vote choice is entwined in countries, elections, and time. Indeed, nearly all the studies that we summarize make use of multilevel models, an approach that is eased by the availability of international comparative survey data, as well as by the establishment of international expert surveys for measuring party positions.

Multilevel models (see Schmidt-Catran et al. 2019) may focus on the effects of contextual variables on the individual dependent variable beyond individual characteristics (random intercept), or may analyze whether the contextual variables affect associations between individual-level variables (random slope models). Both approaches are frequently used in research on party competition and vote choice, and random intercept models are widely applied, especially when modeling the effect of issue saliences. Besides that, many studies use random slope models to examine cross-level interaction effects in order to test whether the variables of party competition moderate the effects of individual-level variables on the dependent variable, ranging from proximity (spatial theory) to issue preferences and competence perceptions (issue voting), to social group status (cleavage theory). While there is thus a great deal of variation in the independent variables (as well as possible interactions between them), the dependent variable in voting studies also shows a great deal of variation, ranging from party families, incumbent parties, closest parties in terms of proximity or directional theory, etc. In sum, this makes the comparison of results between studies quite demanding.

Regarding data structure, the most common way is to see individual voters nested in elections, which then vary in terms of party competition. However, some studies also see party choices nested in voters, and the latter in turn nested in elections, as voters are expected to compare the utility to be gained from each party by comparing several utility functions. Here, the use of so-called propensity-to-vote questions (PTVs) establishes itself as an alternative to categorical dependent variables—but this approach is still limited, as most international comparative surveys do not include these variables.

6 Recent Findings on the Effects of Party Competition on Vote Choice

The following review is based on the findings of the most recent studies on the effect(s) of party competition on voter choice published in leading academic journals between 2010 and 2017.[7] Our focus is on studies applying an international comparative perspective and using quantitative methods to explain individual choices. We therefore excluded the thematically closely related literature on voter–party representation, in which aggregated voters' positions are the independent variable in order to explain parties' positional reactions. We also excluded studies solely interested in electoral turnout (but see Schmitt-Beck 2019), and here focus exclusively on the choice that voters make between political parties. Our search resulted in 26 articles fulfilling the criteria. In order to structure our summary, we start by separating the studies according to the dominant theory of party competition that they apply, beginning with positional theory (studies listed in Table 1). We then discuss this theory's effects on the different voting theories before we do the same with salience theory (studies listed in Table 2). Finally, we take a look at studies interested in the interplay of positional and salience theory arguments ("mixed approaches," see Table 3). For reasons of space, we were unable to summarize each of the studies listed in the tables also in the text. We have therefore limited ourselves to those studies that best illustrate the current state of research on the three party-competition approaches. These studies are quoted in italics both in the tables and in the text.

6.1 Positional Theory and Vote Choice

Starting with positional theory in the tradition of Downs, we find studies applying this theory of party competition to each of the three major voting theories. Unsurprisingly, most of these studies are interested in the effects of party system polarization on spatial voting. The study by *Pardos-Prado* (2015) represents one of the most sophisticated approaches by analyzing the vote decision for a (niche) radical right party compared with an (established) mainstream right party. The author first calculates voter–party proximities on the issue of immigration, which is seen as the most relevant to the support of radical right parties. He then argues that mainstream right parties in some countries have been able to draw support from anti-immigrant voters, while in others these voters have supported the radical right, and points to the relevance of the dimensionality of party system competition for this pattern. *Pardos-Prado* argues that mainstream parties have been more successful in addressing anti-immigrant concerns when the national issue space is one-dimensional, i.e., when immigration concerns are very closely correlated with existing economic and cultural dimensions of party competition. Using three-level models of vote choice (party-specific choices are nested within individuals, who are nested within elections), covering 40 elections in 18 countries, and expert surveys to locate parties' positions on

---

[7] These are: *American Journal of Political Science, American Political Science Review, American Journal of Sociology, American Sociological Review, Electoral Studies, Comparative Political Studies, European Journal of Political Research, European Union Politics, Journal of Politics, Kölner Zeitschrift für Soziologie und Sozialpsychologie, Party Politics*, and *West European Politics*.

Table 1 Recent studies on the effects of positional theory on vote choice. Author's own work

Author/year	Dependent variable	Theory of vote choice	Main party competition variables	Method	Countries covered	Time covered	Data on party competition	Main results
De Vries et al. (2011)	Incumbent vote	Performance (EU) voting	Effective number of parties	MLA (two levels: voters, countries)	19 European countries	2004	None	Due to their higher accountability, voters engage more in EU performance voting in countries with a low number of effective parties
De Vries and Giger (2014)	Incumbent vote	Performance (economic) voting	None	MLA (two levels: voters, elections)	25 countries	2002–2006	None	Highly sophisticated voters attaching greater salience to an issue will reward/punish the incumbent for its performance on this issue
Duch et al. (2010)	Vote choice	Spatial voting	Left–right position of (expected) coalition government	Bayesian conditional logit model	31 countries	1981–2006	MARPOR	Voters base their choice much more on the expected coalition governments than on utility gained from a single party
Evans and Tilley (2012)	Vote choice	Cleavage (class)	Party system polarization on left–right dimension	MLA (two levels: voters, countries)	UK	1986–2010	MARPOR	The more parties are polarized, the more important class and income are for vote choice
Fazekas and Méder (2013)	Vote choice	Spatial (proximity vs. directional) voting	Party system polarization on left–right dimension	MLA (two levels: voters, countries)	27 European countries	2009	Mass survey	Party system polarization increases the explanatory power of directional theory, but leaves proximity voting unaffected

Table 1 (Continued)

Author/year	Dependent variable	Theory of vote choice	Main party competition variables	Method	Countries covered	Time covered	Data on party competition	Main results
Gómez-Reino and Llamazares (2013)	Vote choice for ERP	Issue voting	Position of ERPs on EU issue; deviation of ERP position on EU compared with mean of all other parties	OLS and Logit models (separate models comparing countries)	11 European countries	2008	Expert survey	The more extreme the position of ERPs regarding European integration, the more important Euroscepticism is as a predictor of ERP voting
Han (2016)	Vote choice for radical right party	Cleavage voting (and others)	Party positions on nationalism	MLA (three levels: individuals, countries, years)	16 Western European countries	1990–2012	Mass survey	Rising economic inequality has different effects on individuals depending on their socioeconomic status. The poor tend to vote for the radical right, while the rich are disinclined to vote for them
Harteveld (2016)	Vote choice	Cleavage and issue voting	Party positions on economy	TSCS models	10 European countries	1999, 2004, 2009	Mass survey	The class gap in voting is larger for socioeconomic left parties. Radical right parties reveal a process of proletarianization
Hernández and Kriesi (2016)	Vote for Eurosceptic parties	Spatial and issue voting	Party positions on EU and left–right dimension	MLA (two levels: voters, countries)	24 European countries	2014	Expert survey	In contexts where there is a Eurosceptic party (on the left or the right), Eurosceptic voters will vote on the basis of their EU preferences, given that the Eurosceptic party also shares their domestic left–right position
Milazzo et al. (2012)	Vote choice	Cleavage (class)	Party system polarization on left–right dimension	SEM	UK	1987–2001	Mass survey	The more Conservative and Labour converge, the less relevant class-based considerations are for voter choice

Table 1 (Continued)

Author/year	Dependent variable	Theory of vote choice	Main party competition variables	Method	Countries covered	Time covered	Data on party competition	Main results
Pardos-Prado (2015)	*Vote for mainstream right or ERP*	*Proximity (on immigration issue)*	*Correlations of party positions an economic, cultural and, social dimension*	*MLA (three levels: party choice, individuals, elections)*	*18 countries (with 40 elections)*	*2002–2009*	*Expert survey*	*Center-right parties can compete more successfully on the immigration issue with extreme right parties when the issue space is one-dimensional*
Singh (2010)	*Vote choice*	*Proximity voting*	*Party system polarization and dimensionality of political space*	*MLA (two levels: voters, elections)*	*34 countries*	*1996–2006*	*Expert and mass survey*	*Voters tend to follow proximity voting logic more when the party system is polarized but the number of political dimensions is low*
Tóka and Gosselin (2010)	*Volatility (EP compared with national elections)*	*Cleavage (several)*	*Party system polarization (control)*	*MLA (two levels: voters, countries)*	*20 European countries*	*2004*	*Expert survey*	Insignificant effect of party system polarization on volatility

Table 2 Recent studies on the effects of salience theory on vote choice. Author's own work

Author	Dependent variable	Theory of vote choice	Main party competition variables	Method	Countries covered	Time covered	Data on party competition	Main results
Burschner et al. (2015)	*Vote choice for ERP*	*Issue voting*	*Salience of immigration and crime in news media*	*Rare events logistic regression, (one per party/country)*	*11 Western European countries*	*2009*	*News content analysis*	*High salience of crime and immigration issue increases the likelihood of a vote for ERPs. Effect is stronger for voters already sympathetic to ERPs*
Jansen et al. (2012)	*Vote choice*	*Cleavage (religion)*	*Salience of traditional morality issue*	*Conditional logistic regression*	*Netherlands*	*1971–2006*	*MARPOR*	*The higher the salience of "traditional morality" in parties' manifestos, the lower the vote intention of secular voters, but the higher the vote intention of religious voters for religious parties*
Williams (2015)	*Vote choice*	*Issue voting*	*Salience of positive defense spending statements*	*Meta-analysis of (defense issue voting) coefficients received from multinomial logits*	*26 countries*	*1985, 1990, 1996, 2006*	*MARPOR*	*The more parties emphasize defense spending in their manifestos, the more voters base their choice on that issue*

Table 3 Recent studies on the effects of mixed approaches on vote choice. Author's own work

Author	Dependent variable	Theory of vote choice	Main party competition variables	Method	Countries covered	Time covered	Data on party competition	Main results
De Sio and Franklin (2012)	Vote choice	Proximity voting	Salience of different issues	MLA (three levels: party choice, individuals, elections)	27 European countries	2009	MARPOR and expert survey	Issue yield is more effective than manifesto emphasis for party support
De Sio and Weber (2014)	Voters' issue preferences and party's policy emphasis	Issue voting	Salience of different issues	MLA (three levels: parties, party families, countries)	27 European countries	2009	Euromanifesto Project	Parties are able to exploit the multidimensionality of the political issue space for their own purposes
De Sio et al. (2016)	Vote choice	Proximity voting	Salience of different issues	MLA (three levels: party choice, individuals, elections)	28 (2009) and 29 (2013) European political systems	2009 and 2014	Chapel Hill Expert Survey	Parties avoid discussing issues (such as European integration) that would split their electorate
De Vries et al. (2011)	Vote choice in EP elections	Spatial voting	Party system polarization; news coverage	MLA (two levels: voters, countries)	27 European countries	2009	Mass survey	Issue voting increases with more party conflict over EU issues and with greater media attention being paid to EU issues
De Vries and Hobolt (2012)	Vote for mainstream government, mainstream opposition or challenger party	Spatial and issue voting	Position and salience of support for EU; challenger parties' deviation from party system mean and issue salience	MLA (two levels: voters, countries)	21 Western and Eastern European countries	2004 (individual level); 1984–2006 (macro-level)	Expert survey	Parties that employ an entrepreneurial issue strategy are more successful electorally, and therefore try to establish new issue dimensions

Table 3 (Continued)

Author	Dependent variable	Theory of vote choice	Main party competition variables	Method	Countries covered	Time covered	Data on party competition	Main results
Gomez et al. (2016)	Vote choice for radical left party	Issue voting	None	MLA (two levels: voters, countries)	13 Western European countries	1989–2009	MARPOR and expert survey	Radical left-wing parties are (largely) representative in terms of their voters' policy positions
Hobolt and Spoon (2012)	Vote choice in EP compared with national elections (stay the same, switch to another party, abstain)	Spatial voting (proximity on left–right and EU dimension) vs. protest voting	Party system polarization on EU. Salience of EU in media	MLA (two levels: voters, countries)	27 countries	2009	News content analysis (salience). Mass survey and expert survey for party positions	The degree of politicization of the EU in the domestic debate (polarization and news coverage) shapes the extent to which voters rely on EU, rather than national, considerations when voting in European Parliament elections
Hong (2015)	Vote choice for niche parties in European Parliament elections	Spatial and issue voting	None	MLA (two levels: voters, countries)	15 Western European countries	2009	Mass survey	Voters tend to base their choice more on the European dimension (proximity) and on niche issues (salience) in European Parliament elections
Pardos-Prado (2012)	Vote choice	Spatial and issue voting	Party system polarization on left–right dimension	MLA (two levels: voters, countries)	21 countries	2004	Mass survey	Party competence perceptions (issue voting) are stronger electoral determinants in countries with higher levels of party system polarization

EP European Parliament, ERP Extreme Right Party, EU European Union, MARPOR Manifesto Research Group on Political Representation, MLA Multi-level Analysis, OLS Ordinary Least Squares, SEM Structural Equation Modeling, TSCS Time-series Cross-section, UK United Kingdom

the distinct dimensions, the author models cross-level interaction effects between the strength of correlation of party positions on three dimensions and immigration-proximity vote. These models show that mainstream right parties can compete more successfully with radical right parties when immigration can be assimilated into existing patterns of competition.

Other studies applying positional theory are more interested in its relevance for the predictive power of spatial voting theories or their proximity and directional subtypes. Addressing spatial theories demanding assumptions about voters' ability to calculate distances between their own position and those of parties, *Singh (2010)* argues that this ability is determined by both party system polarization and the number of issue dimensions: voters tend to follow a proximity voting logic more when the party system is polarized, i.e., when the positions of parties are clearly separable (see also Pardos-Prado and Dinas 2010; Lachat 2008). However, high party system polarization loses this effect if the number of political dimensions increases, i.e., when parties hold different positions on distinct dimensions of competition. Singh shows that issue dimensionality shows very high variation across countries, the UK being the closest example of a one-dimensional space, but countries such as Spain and New Zealand following a multi-dimensional logic of party competition. Calculating multilevel models for 34 countries between 1996 and 2006, and using mass survey as well as expert survey data to identify party positions, he shows that if the complexity of an election increases, it becomes more difficult for voters to decide between parties and, ultimately, to follow a proximity logic (see also Wessels and Schmitt 2008). This joint effect of the number of parties, their polarization, and the number of dimensions is highly relevant: "all else [being] equal, the odds of a person living in New Zealand in 2002, in which political variation was not unidimensional, voting proximately are only 50% of those of an individual residing in Australia in 2004, in which politics conformed well to a single dimension" (Singh 2010, p. 433).

Also pointing to the role of party system polarization, Fazekas and Méder (2013) examine whether this contextual variable impacts on spatial voting in general, and they look at the relative explanatory power of proximity compared with directional theory. The authors start from the same assumption as Pardos-Prado (2015), namely that voters will find it easier to distinguish between the positions of parties in more polarized systems, increasing the general explanatory power of spatial theories. However, increased polarization should also lead to an increase in the explanatory power of directional compared with proximity theories, but only for those cases that really allow for a discriminatory distinction of the rival theories—this being about 25% in their sample of 27 countries. The reason for this, so the authors argue, lies in the polarization formula (see above), as this takes not only party positions into account but also parties' voting share. If more important parties—usually located in the middle of the left–right scale—become more extreme, this sharply increases polarization, leading to very clear choices for voters who follow a directional calculation. At the same time, the distance between the central median voter and each of the major parties is increased in such a scenario, thus making proximity theory less predictive. The authors hence test for the explanatory power of increased centripetal party competition, which they assume to drive party system polarization

more than increased centrifugal competition does. Their multilevel models confirm these theoretical expectations, i.e., increased party polarization leads to an increase in the number of voters engaged in spatial voting from 20% (least polarized) to 60% (most polarized context). While polarization also increases the explanatory power of directional theory—for those cases that are comparable—directional voting never outperforms proximity theory, even in highly polarized party systems.

The studies by Gómez-Reino and Llamazares (2013) and Hernández and Kriesi (2016) apply a positional logic to issue voting theories. Both studies are interested in the vote choice for niche parties as being *issue entrepreneurs*, that is parties that try to introduce a certain issue in order to divide established parties and their voters over it. Comparing the vote decision for a radical right party in eleven European countries, Gómez-Reino and Llamazares point to the influence of Euroscepticism on this decision. They argue that the radical right, alongside its pronounced anti-immigrant platform, tries to activate Eurosceptic attitudes in order to increase its electoral support. The authors argue that radical right parties are more successful in activating Euroscepticism the more they deviate from the mean party position on European integration. This assumption is tested and supported with mass and expert survey data on parties' policy positions using a series of separate OLS and logit models comparing countries. Hernández and Kriesi (2016) also see that niche parties are more successful in European parliamentary elections the more they diverge from the mean of the national party system regarding European integration. The effect of this positioning is, however, subordinated to the domestic left–right dimension, as Eurosceptic voters do not support an anti-EU party that ignores their domestic preferences.

Finally, arguments derived from positional theory of party competition have also been applied to cleavage voting. Investigating the effects of party system polarization in the UK, the studies of Evans and Tilley (2012) and Milazzo et al. (2012) report that the electoral information offered by Labour and the Conservatives is highly correlated with the level of class voting over time. Their findings—based on multilevel (Evans and Tilley 2012) and structural equation models (Milazzo et al. 2012)—suggest that the less polarized the parties' positions are on the class dimension, the less we can expect voters' positions on the class dimension to determine their vote (see also Achterberg 2006; Elff 2007; Spies 2013 for similar findings in a cross-national perspective). Both studies operationalize party system polarization on a left–right dimension, arguing that this is the dominant dimension in the UK party system, highly related to class conflict. While the decline of class-based voting has usually been considered a consequence of changes on the demand side of electoral competition (less unionization, increased individualization, less demand for redistributive policies, etc.), both studies come to the conclusion that the electoral strategies of the two major parties are equally relevant to the decline in class-based voting. Milazzo et al. identify two reasons for this. First, when the distance between Labour and the Conservatives declines, citizens are less likely to perceive class differences between the parties, which they see as a necessary condition for class status to impact voter choice. Secondly, party elites have, in general, fewer incentives to campaign on issues that do not distinguish the party from its opponent(s), since such dimensions may be of little relevance even to those voters who still perceive party differences.

6.2 Salience Theory and Vote Choice

We now turn to studies analyzing the impact of salience theory in a cross-national perspective. We were not able to find any study exclusively applying salience arguments to spatial voting. This is hardly surprising, as spatial voting theories are closely related to positional theory in the tradition of Downs and therefore measure party statements in positions rather than issue saliences. In contrast, salience theory is closely related to issue ownership theories, and we found several studies following this tradition. One of the clearest examples is the study by Williams (2015), who analyzes defense issue voting from a cross-national perspective. Defense spending is not usually considered an area that is relevant to vote choice outside the US. However, relying on survey data from 26 countries between 1985 and 2006, Williams is able to show that the more parties emphasize defense spending in their manifestos, the more voters base their choice on that issue. The rationale behind this is that parties' first reaction to international crises is to emphasize the need to increase defense spending. Aware of the international problems, voters listen to these statements and choose the party most competent to handle them.

The study by Burschner et al. (2015) analyzes the vote choice for a radical right party in eleven Western European countries, and points to the role of the salience of immigration and crime in national news media. It is argued on the basis of issue ownership theory that anti-immigrant parties own both issues, link them, and closely focus their electoral campaigns on them. Voters are therefore seen to react to increased issue salience in the core domain of the radical right when national media also support the view that immigration and crime are important problems that need to be addressed. Using elaborate and time-intensive hand coding of more than 20,000 news stories in the countries being analyzed, matching these with survey data from 2009, and applying rare-events logistic regressions, the authors conclude that the salience of crime and immigration sharply increases the likelihood of voters choosing a radical right party. This effect is stronger for voters who already show some sympathy for such parties.

That purely salience-motivated arguments can also be applied to cleavage theories is shown by Jansen et al. (2012). The authors analyze the influence of religion on voter choice in the Netherlands from 1971 to 2006. As with class, religious denomination and religiosity are usually seen as a cleavage of declining importance for individual voters. Demand-side factors such as secularization and individualization are considered to be the main drivers here. However, the authors argue that party policy also plays an important role in the decline of the influence of religion. Applying this top-down perspective to the Netherlands, and borrowing their theoretical argument from the class voting literature, Jansen et al. make the point that parties have to activate voters' religious orientations by emphasizing the salience of religiosity in politics. In particular, this can be done by stressing issues of traditional morality in manifestos (MARPOR data), which the authors then find to be strongly correlated with vote choice and denomination. This effect of issue salience is also much stronger for churchgoers than for other voters, leading the authors to conclude that religious parties face a tradeoff between binding their traditional supporters to them and maximizing vote share by de-emphasizing traditional moral values. This

tradeoff resembles that faced by social democratic parties in terms of class, but the important point here is that parties can decide between the two options and adapt their electoral strategies accordingly.

6.3 Mixed Approaches and Vote Choice

We have introduced positional and salience theory as distinct and partly contradictory theoretical perspectives on party competition. While this is reasonable because of their distinct assumptions about parties' behavior, issue dimensionality, and of the level of information on the side of both parties and voters, there have been several attempts in the literature during the last decade to unify the approaches. These studies focus on the interplay of party system polarization and issue salience, arguing, e.g., that in order to be regarded as a salient issue by voters, parties should provide voters with clearly distinguishable positions on a given issue. Rational voters would otherwise have little reason to cast their vote in a dimension in which parties do not show any difference, and rational parties would show little interest in emphasizing issues which do not distinguish them from rival parties. Most recently, the interplay of positional and salience theory has also been used to analyze the rise of so-called niche parties (Meguid 2008; Wagner 2012; Zons 2016) or "challenger" (De Vries and Hobolt 2012), e.g., radical right or radical left parties. Also, the concept of "issue yield" (De Sio and Franklin 2012) contributes to the further development of modeling party behavior beyond clear predefined issue dimensions.

Analyzing the vote choice for challenger parties, De Vries and Hobolt (2012) argue that these new competitors have a strong incentive to manipulate existing patterns of party competition by introducing a new issue dimension, i.e., by increasing its salience. Differentiating parties between mainstream government, mainstream opposition, and challenger parties, the authors argue that in particular the latter will try to engage in this strategy, as they are the losers in the existing dimension(s) of competition. Faced with the objective of increasing issue salience, challenger parties then adopt a highly extreme position in the issue dimension that they want to establish, which is assumed to be a more successful strategy when the other parties adopt centrist positions. As established parties will react to the new competitor—some will move from the middle towards the challenger's position and some away from it—they heighten the public awareness of party differences on new issues and thereby increase their salience. If voters care about this newly-evolving issue—along the lines of issue voting theories—both rising salience and polarization may lead to changes in mass identification on the basis of the new issue dimension, eventually leading to changing voting behavior. Combining these salience and positional arguments, *De Vries and Hobolt* then empirically address them for European integration as a newly emerging issue dimension. Based on multilevel multinomial models for 21 Western and Eastern European countries in 2004, they first show that Eurosceptic voters are more likely to vote for challenger parties. Using time series cross-section models with challenger parties' voter shares as the dependent variable, they then apply expert survey data on both issue salience and the positions of parties to show that the product of both variables in the dimension of European integration is significantly correlated with better electoral results for challenger parties. These

models include 14 Western European countries from 1984 to 2006, and also control for parties' positions in the left–right dimension, thereby also addressing the question of issue-space dimensionality.

Also combining issue salience and polarization arguments, Hobolt and Spoon (2012) analyze vote switching (compared with national elections) in European Parliament elections. Based on multilevel models for 27 countries in 2009, they show that the degree of politicization of the EU in the domestic arena shapes the extent to which voters engage in EU-specific proximity voting at the European level. More specifically, their indicators of the level of politicization are the degree of party system polarization on the issue of European integration (measured by expert survey data) and the contentiousness of European integration in the campaign coverage (based on a coding of more than 50,000 television and newspaper stories in the weeks prior to the European Parliament election). *Hobolt and Spoon* argue that party polarization is a central determinant of the politicization of European Parliament elections, as only a polarized national party system offers voters real choices on the issue of European integration; it also increases the salience of European issues to them. While this should lead voters to cast their vote in European Parliament elections more on the EU issue, this is further amplified by the national news content—a salience theory argument. Thus, the authors expect that disagreements between voters and parties over EU issues will play a greater role in voters' European Parliament choices when the problems associated with European integration are highlighted in the media during the campaign. Their empirical investigation supports these theoretical assumptions, but only the interplay of neutral or negative news campaigning and a high level of party system polarization shows an increase in EU proximity voting.

The study carried out by Pardos-Prado (2012) addresses the relationship between valence judgments—i.e., voters' perceptions of party competence in dealing with salient issues—and polarization. While it is a common assumption in the literature that a high salience of valence issues should be associated with a low salience of positional issues—and therefore low levels of party system polarization—the author challenges this assumption. Based on multilevel models of 21 countries in 2004, he finds highly significant cross-level interaction effects between left–right party system polarization and issue voting based on party competence perceptions in the areas of unemployment, immigration, and environment protection, areas traditionally seen as being valence based or positional. These findings question the prevalent view of the relationship between polarization and valence issue voting, and point to the need to distinguish conceptually between valence and positional issues, rather than seeing them as simply contrasting terms. Also, the study makes a strong claim to relate positional theory to voting theories other than spatial voting.

The perhaps most remarkable theoretical innovation that bridges the divide between salience and position theory is the concept of the "issue yield" (De Sio and Franklin 2012). The central idea of issue yield is that parties modulate their issue emphases according to their issue-specific, individual risk–opportunity profile. Hence, according to this model, concentrating on issues with the highest salience in a party system is not the most promising strategy for many parties when it comes to attracting voters. Instead, parties will focus on those issues which unite their core

voters and, at the same time, are widely supported by the overall electorate. Issue yield can therefore be defined "as the degree to which an issue allows a party to overcome the conflict between protection and expansion of electoral support" (De Sio and Weber 2014, p. 871). In general, the issue-yield model applies the Downsian rationale in explaining party strategies. However, it does not necessitate the existence of a predefined left–right policy space. The issue-yield concept considers valence and position issues as two poles of a continuum in which the parties can define the character of each issue. Parties are free to choose those bundles of issues that they think deliver the best risk–opportunity structure (De Sio et al. 2016, p. 485). Analytically, this model is suitable for application in analyzing party systems without stable ideological dimensions.

Applying multi-level modeling, De Sio and Franklin (2012) demonstrate how the issue-yield model is able to better explain the reasons why voters choose a specific party. Studying the European Parliament elections of 2009 and 2013, the issue-yield concept enables us to explain why the issue of European integration has been kept out of party competition by and large, while it was the most important topic among the electorate (De Sio et al. 2016). Using data from the European Election Study (EES) and the Chapel Hill Expert Survey (CHES), the authors apply a three-level model, which sees voter–party combinations on propensities to vote nested in individuals, and these to be nested in party systems. They show that mainstream parties avoided talking about European integration since their core electorate was split on this question, tending to punish the party elite in the case of it taking any position on this issue.

7 Conclusion

Summarizing our review of the international comparative literature on party competition and vote choice, we conclude that recent studies still mainly rely on positional or salience theory when theoretically addressing the role of parties. Regarding the individual vote decisions, both highly rational theories of spatial voting, competence-based theories of issue voting, and theories pointing to the role of cleavages are applied. While positional and salience theory are distinct and partly contradictory, theoretical perspectives on party competition with distinct assumptions about parties' behavior, issue dimensionality, and level of information on the side of both parties and voters, there have been several attempts in the literature during the past decade to unify the approaches. These studies frequently focus on the interplay between party system polarization and issue salience—a promising road for further research.

Regarding methodological issues, the literature shows a highly sophisticated level of methodological complexity when it comes to modeling voter–party relationships. However, more serious but rarely discussed potential problems arise from the objective of measuring party competition. While expert surveys and data provided by the hand coding of party manifestos are by far the most prominent approaches here, each presents its own advantages and disadvantages. While we do not want to state that one approach is superior to the other, we end our review by calling to mind

that expert surveys and MARPOR data are strongly affiliated to either positional or salience theory. Therefore, they also entail distinct problems when addressing the effects of party competition on vote choice in a cross-national perspective, and especially so where different periods are concerned.

The major challenge for the contemporary study of European party competition is how to integrate the dynamic of the transforming political landscape. The "old" catch-all parties are vanishing, or are at least losing their dominant role across Europe, while populist and left-libertarian cosmopolite parties have become stronger. The traditional socioeconomic cleavages are in decline, and cultural issues dominate the agenda. However, this has not yet led to a stable ideological space as it is assumed in Downsian theory. Rather, especially in Southern Europe, parties select issues regardless of the ideological camp to which these issues belong. Parties thus freely combine issues without generating a new common ideological space. Combining both positional and salience arguments, the issue-yield concept offers a framework that enables researchers to analyze individual party calculus without making strong assumptions on the ideological space of a given party system. Hence, it is suitable to be applied in explaining patterns of dynamically changing party systems, both in single- (see, e.g., Franzmann et al. 2018) and cross-country studies. Future studies of party competition and vote choice will have to do even more in order to overcome the old divide between positional and salience theory. It is only then that those studies will be able to analyze the causal mechanisms behind the dramatic change in the European party system landscape.

References

Achterberg, Peter. 2006. Class voting in the new political culture. Economic, cultural and environmental voting in 20 western countries. *International Sociology* 21:237–261.

Adams, James, Samuel Merrill III, and Bernard Grofman. 2005. *A unified theory of party competition: A cross-national analysis integrating spatial and behavioral factors.* Cambridge University Press, Cambridge.

Ansolabehere, Stephen, and Shanto Iyengar. 1994. Riding the wave and claiming ownership over issues: The joint effects of advertising and news coverage in campaigns. *The Public Opinion Quarterly* 58:335–357.

Belanger, Éric, and Bonnie M. Meguid. 2008. Issue salience, issue ownership, and issue-based vote choice. *Electoral Studies* 27:477–491.

Benoit, Kenneth, and Michael Laver. 2006. *Party policy in modern democracies*. London: Routledge.

Benoit, Kenneth, and Michael Laver. 2007. Estimating party policy positions: Comparing expert surveys and hand coded content analysis. *Electoral Studies* 26:90–107.

Blais, André, Nadeau, Richard, Elisabeth Gidengil, and Neil Nevitte. 2001. The formation of party preferences: Testing the proximity and directional models. *European Journal of Political Research* 40:81–91.

Van der Brug, Wouter. 2004. Issue ownership and party choice. *Electoral Studies* 23:209–233.

Budge, Ian. 2001. Theory and measurement of party policy positions. In *Mapping policy preferences. Estimates for parties, electors, and governments 1945–1998*, eds. Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tanenbaum, 75–92. Oxford: Oxford University Press.

Budge, Ian. 2015. Issue emphases, saliency theory and issue ownership: A historical and conceptual analysis. *West European Politics* 38:761–777.

Budge, Ian, and Dennis Farlie. 1978. The potentiality of dimensional analyses for explaining voting and party competition. *European Journal of Political Research* 6:203–232.

Budge, Ian, and Paul Pennings. 2007. Missing the message and shooting the messenger: Benoit and Laver's 'response'. *Electoral Studies* 26:136–141.

Burschner, Bjorn, Joost van Spanje, and Claes H. de Vreese. 2015. Owning the issues of crime and immigration: The relation between immigration and crime news and anti-immigrant voting in 11 countries. *Electoral Studies* 38:59–69.

Dalton, Russell J. 2002. Political cleavages, issues, and electoral change. In *Comparing democracies 2. New Challenges in the Study of Elections and Voting,* eds. Lawrence LeDuc, Richard G. Niemi, and Pippa Norris, 189–209. Thousand Oaks, CA: Sage Publications.

Dalton, Russell J. 2008. The quantity and the quality of party systems. *Comparative Political Studies* 41:899–920.

Dogan, Mattei. 1995. Erosion of class voting and of the religious vote in Western Europe. *International Social Science Journal* 47:525–538.

Downs, Anthony. 1957. *An economic theory of democracy.* New York: Harper and Row.

Duch, Raymond M., Jeff May, and David A. Armstrong. 2010. Coalition-directed voting in multiparty democracies. *American Political Science Review* 104:698–719.

Elff, Martin. 2007. Social structure and electoral behavior in comparative perspective: The decline of social cleavages in Western Europe revisited. *Perspectives on Politics* 5:277–294.

Elff, Martin. 2013. A dynamic state-space model of coded political texts. *Political Analysis* 21:217–232.

Enelow, James M., and Melvin J. Hinich. 1984. *The spatial theory of voting.* Cambridge; University Press.

Enyedi, Zsolt, and Kevin Deegan-Krause. 2010. Introduction: The structure of political competition in Western Europe. *West European Politics* 33:415–418.

Evans, Geoffrey, and James Tilley. 2012. The depoliticization of inequality and redistribution: Explaining the decline of class voting. *The Journal of Politics* 74:963–976.

Evans, Geoffrey, Anthony Heath, and C. Payne. 1999. Class: Labour as a catch-all party? In *Critical elections: British parties and voters in long-term perspective,* eds. Geoffrey Evans and Pippa Norris, 87–101. London: Sage.

Ezrow, Lawrence, Jonathan Homola, and Margit Tavits. 2014. When extremism pays: Policy positions, voter certainty, and party support in postcommunist Europe. *The Journal of Politics* 76:535–547.

Fazekas, Zoltán, and Zsombor Zoltán Méder. 2013. Proximity and directional theory compared: Taking discriminant positions seriously in multi-party systems. *Electoral Studies* 32:693707.

Fiorina, Morris. P. 1981. *Retrospective voting in American national elections.* New Haven, CT: Yale University Press.

Franklin, Mark. N., Thomas T. Mackie, and Henry Valen. 1992. *Electoral change: Responses to evolving social and attitudinal structures in western countries.* Cambridge: Cambridge University Press.

Franzmann, Simon T. 2011. Competition, contest, and cooperation. The analytical framework of the issue-market. *Journal of Theoretical Politics* 23:317–343.

Franzmann, Simon T. 2015. Towards a real comparison of left-right indices: A comment on Jahn. *Party Politics* 21:821–828.

Franzmann, Simon, and André Kaiser. 2006. Locating political parties in policy space. A reanalysis of party manifesto data. *Party Politics* 12:163–188.

Franzmann, Simon T., Heiko Giebler, and Thomas Poguntke. 2018. It's not only the economy stupid! Issue Yield at the 2017 German Federal Election. Special Issue in *West European Politics* 1/2019, ed. by Lorenzo De Sio, and Romain Lachat (under review).

Gabel, Matthew J., and John D. Huber. 2000. Putting parties in their place: Inferring party left-right ideological positions from party manifesto data. *American Journal of Political Science* 44:94–103.

Gomez, Raul, Laura Morales, and Luis Ramiro. 2016. Varieties of radicalism: Examining the diversity of radical left parties and voters in Western Europe. *West European Politics* 39:351–379.

Green, Jane, and Sara B. Hobolt. 2008. Owning the issue agenda: Party strategies and vote choices in British elections. *Electoral Studies* 27:460–476.

Gómez-Reino, Margarita, and Iván Llamazares. 2013. The populist radical right and European integration: A comparative analysis of party-voter links. *West European Politics* 36:789–816.

Han, Kyung Joon. 2016. Income inequality and voting for radical right-wing parties. *Electoral Studies* 42:54–64.

Harteveld, Eelco. 2016. Winning the 'losers' but losing the 'winners'? The electoral consequences of the radical right moving to the economic left. *Electoral Studies* 44:225–234.

Hernández, Enrique, and Hanspeter Kriesi. 2016. Turning your back on the EU. The role of Eurosceptic parties in the 2014 European Parliament elections. *Electoral Studies* 44:515–524.

Hobolt, Sara B., and Jae-Jae Spoon. 2012. Motivating the European voter: Parties, issues and campaigns in European Parliament elections. *European Journal of Political Research* 51:701–727.

Hong, Geeyoung. 2015. Explaining vote switching to niche parties in the 2009 European Parliament elections. *European Union Politics* 16:514–535.

Inglehart, Ronald. 1984. The changing structure of political cleavages in western society. In *Electoral change in advanced industrial democracies: Realignment or dealignment?* Eds. Russell J. Dalton, Scott C. Flanagan, and Paul A. Beck, 25–69. Princeton, NJ: Princeton University Press.

Jahn, Detlef. 2011. Conceptualizing left and right in comparative politics: Towards a deductive approach. *Party Politics* 17:745–765.

Jansen, Giedo, Nan Dirk de Graaf, and Ariana Need. 2012. Explaining the breakdown of the religion-vote relationship in The Netherlands. 1971–2006. *West European Politics* 35:756–783.

Kitschelt, Herbert. 1994. *The transformation of European social democracy*. Cambridge: Cambridge University Press.

Knutsen, Oddbjørn. 2004. Religious denomination and party choice in Western Europe: A comparative longitudinal study from eight countries, 1970–97. *International Political Science Review* 25:97–128.

Koenig, Thomas, Moritz Marbach, and Moritz Osanbrügge. 2013. Estimating party positions across countries and time—a dynamic latent variable model for manifesto data. *Political Analysis* 21:468–491.

Kramer, Jürgen, and Hans Rattinger. 1997. The proximity and the directional theories of issue voting: comparative results for the US and Germany. *European Journal of Political Research* 32:1–29.

Kriesi, Hanspeter. 2010. Restructuration of partisan politics and the emergence of a new cleavage based on values. *West European Politics Volume* 33:673–685.

Kriesi, Hanspeter, Edgar Grande, Romain Lachat, Martin Dolezal, Simon Bornschier, and Timotheos Frey. 2006. Globalization and the transformation of the national political space: Six European countries compared. *European Journal of Political Research* 45:921–956.

Lachat, Romain. 2008. The impact of party polarization on ideological voting. *Electoral Studies* 27:687–698.

Laver, Michael, and W. Ben Hunt. 1992. *Policy and party competition*. London: Routledge.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. Estimating the policy positions of political actors using words as data. *American Political Science Review* 97:311–331.

Lewis-Beck, Michael, and Mary Stegmaier. 2007. Economic models of voting. In *The Oxford handbook of political behavior*, eds. Russell J. Dalton, and Hans-Dieter Klingemann, pp. 518–537, Oxford: Oxford University Press.

Lipset, Seymour Martin, and Stein Rokkan. Eds. 1967. *Party systems and voter alignments: Cross-national perspectives*. Toronto: Free Press.

Lowe, Will. 2008. Understanding wordscores. *Political Analysis* 16:356–371.

Lowe, Will, Kenneth Benoit, Slava Mikhailov, and Michael Laver. 2011. Scaling policy positions from coded units of political texts. *Legislative Studies Quarterly* 36:123–155.

Meguid, Bonnie M. 2008. *Party competition between unequals: Strategies and electoral fortunes in Western Europe*. Cambridge: Cambridge University Press.

Milazzo, Caitlin, James Adams, and Jane Green. 2012. Are voter decision rules endogenous to parties' policy strategies? A model with applications to elite depolarization in Post-Thatcher Britain. *The Journal of Politics* 74:262–276.

Miller, Warren E, and J. Merrill Shanks. 1996. *The new American voter*. Cambridge, MA: Harvard University Press.

Nadeau, Richard, André Blais, Elisabeth Gidengil, and Neil Nevitte. 2001. Perceptions of party competence in the 1997 election. In *Party politics in Canada,* eds. Hugh G. Thorburn, and Alan Whitehorn, 413–430. Prentence Hill, Toronto.

Neundorf, Anja, and James Adams. 2016. The micro-foundations of party competition and issue ownership: The reciprocal effects of citizens' issue salience and party attachments. *British Journal of Political Science* 48:385–406.

Norris, Pippa, and Ronald Inglehart 2004. *Sacred and secular: Religion and politics worldwide.* Cambridge: Cambridge University Press.

Pappi, Franz Urban. 2000. Zur Theorie des Parteienwettbewerbs. [On the theory of party competition] In *50 Jahre empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten,* eds. Markus Klein, Wolfgang Jagodzinski, Dieter Ohr, Ekkehard Mochmann, 85–105. Wiesbaden: Westdeutscher Verlag.

Pardos-Prado, Sergi. 2012. Valence beyond consensus: Party competence and policy dispersion from a comparative perspective. *Electoral Studies* 31:342–352.

Pardos-Prado, Sergi. 2015. How can mainstream parties prevent niche party success? Center-right parties and the immigration issue. *The Journal of Politics* 77:352–367.

Pardos-Prado, Sergi, and Elias Dinas. 2010. Systemic polarisation and spatial voting. *European Journal of Political Research* 49:759–786.

Petrocik, John R. 1996. Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science* 40:825–850.

Pierce, Roy. 1997. The directional theory of issue voting: directional versus proximity models: Verisimilitude as the criterion. *Journal of Theoretical Politics* 9:61–74.

Powell, G. Bingham. 2000. *Elections as instruments of democracy: Majoritarian and proportional visions.* New Haven and London: Yale University Press.

Rabinowitz, George, and Stuart Elaine Macdonald. 1989. A directional theory of issue voting. *The American Political Science Review* 83:93–121.

Robertson, David. 1976. *A theory of party competition.* London: John Wiley & Sons.

Sartori, Giovanni. 1976. *Parties and party systems: A framework for analysis.* New York: Cambridge University Press.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schmitt-Beck, Rüdiger. 2019. Political systems and electoral behavior: A review of internationally comparative multilevel research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Schmitt, Johannes, and Simon Franzmann. 2018. *Measuring ideological polarization in party systems: An evaluation of the concept and the indicators.* Manuscript, Düsseldorf: Düsseldorf University.

Singh, Shane P. 2010. Contextual influences on the decision calculus: A cross-national examination of proximity voting. *Electoral Studies* 29:425–434.

De Sio, Lorenzo, and Mark N. Franklin. 2012. Strategic Incentives, Issue Proximity and Party Support in Europe. *West European Politics* 35:1363–1385.

De Sio, Lorenzo de, and Till Weber. 2014. Issue yield: A model of party strategy in multidimensional space. *American Political Science Review* 108:870–85.

De Sio, Lorenzo, Mark N. Franklin, and Till Weber. 2016. The risks and opportunities of Europe: How issue yield explains (non-)reactions to the financial crisis. *Electoral Studies* 44:483–91.

Slapin, Jonathan, and Sven-Oliver Proksch. 2008. A scaling model for estimating time series party positions from texts. *American Journal of Political Science* 52:705–722.

Spies, Dennis. 2013. Explaining working-class support for extreme right parties: A party competition approach. *Acta Politica* 48:296–325.

Stokes, Donald E. 1963. Spatial models of party competition. *American Political Science Review* 57:368–377.

Stokes, Donald E. 1992. Valence politics. In *Electoral politics,* ed. Dennis Kavanagh, 141–164. Oxford: Clarendon Press.

Strom, Kaare. 1990. A behavioral theory of competitive political parties. *American Political Science Review* 34:565–598.

Taylor, Michael, and V.M. Herman. 1971. Party systems and government stability. *American Political Science Review* 65:28–37.

Tomz, Michael, and Robert P. Van Houweling. 2008. Candidate positioning and voter choice. *American Political Science Review* 102:303–318.

Tóka, Gábor, and Tania Gosselin. 2010. Persistent political divides, electoral volatility and citizen involvement: The freezing hypothesis in the 2004 European election. *West European Politics* 33:608–633.

Van der Eijk, Cees, Hermann Schmitt, and Tanja Binder. 2005. Left–right orientations and party choice. In *The European voter,* ed. Jacques Thomassen, 167–191. Oxford: Oxford University Press.

De Vries, Catherine E., and Nathalie Giger. 2014. Holding governments accountable? Individual heterogeneity in performance voting. *European Journal of Political Research* 53:345–362.

De Vries, Catherine E., and Sara B. Hobolt. 2012. When dimensions collide: The electoral success of issue entrepreneurs. *European Union Politics* 13:246–268.

De Vries, Catherine E., Erica E. Edwards, and Erik R. Tillman. 2011. Clarity of responsibility beyond the pocketbook: How political institutions condition EU issue voting. *Comparative Political Studies* 44:339–363.

Wagner, Magnus. 2012. Defining and measuring niche parties. *Party Politics* 18:845–864.

Walgrave, Stefaan, Jonas Lefevere, and Anke Daniela Tresch, 2012. The associative dimension of issue ownership. *Public Opinion Quarterly* 76:771–782.

Wessels, Bernhard, and Hermann Schmitt. 2008. Meaningful choices, political supply, and institutional effectiveness. *Electoral Studies* 27:19–30.

Westholm, Anders. 1997. Distance versus Direction: The Illusory Defeat of the Proximity Theory of Electoral Choice. *American Political Science Review* 91:865–883.

Williams, Laron K. 2015. Estimating the defense spending vote. *Electoral Studies* 39:243–255.

Zons, Gregor. 2016. How programmatic profiles of niche parties affect their electoral performance. *West European Politics* 39:1205–1229.

Dennis Christopher Spies 1981, PD Dr., University of Cologne, Cologne Center for Comparative Politics. Areas of Research: Comparative Politics, Political Economy, Voting Behavior. Publications: Immigration and welfare state retrenchment: Why the US experience is not reflected in Western Europe. Oxford 2018; Do parties "Playing the Race Card" undermine natives' support for welfare? Comparative Political Studies 47, 2014.

Simon T. Franzmann 1977, PD Dr., University of Düsseldorf, Heinrich-Heine University Düsseldorf, Social Science Department, Comparative Politics. Areas of Research: Comparative Politics, Party Competition, Opposition, Populism. Publications: Locating political parties in policy space. A reanalysis of party manifesto data. Party Politics 12, 2006; Calling the ghost of populism: The AfD's strategic and tactical agendas until the EP election 2014. German Politics 25, 2016.

**Kölner Zeitschrift
für Soziologie und
Sozialpsychologie**

ABHANDLUNGEN

Multilevel Structural Equation Modeling for Cross-National Comparative Research

Bart Meuleman

Abstract This contribution focuses on a model that is gaining currency in cross-national research, namely multilevel structural equation modelling (MSEM). Similarly to standard multilevel modelling (MLM), this model distinguishes between various levels of analysis (e.g., individuals nested within countries) and, in doing so, takes the hierarchical structure of cross-national data into account. However, MSEM incorporates a latent-variable approach into the multilevel framework, making it possible to assess the measurement quality of latent constructs. As such, MSEM is a synthesis of structural equation modeling (SEM) and MLM that combines the best of both worlds. The MSEM approach makes it possible to model multilevel mediations and group-level outcomes, and therefore provides a more complete representation of Coleman's bathtub model. This contribution presents the statistical and conceptual background of MSEM in a formal but accessible manner. The paper discusses applications of MSEM that are particularly useful for cross-national comparative research (CNCR), namely two-level confirmatory factor analysis (CFA), multilevel mediation models, and models for group-level outcomes. A practical step-by-step strategy on how MSEM can be used for applied research is provided and illustrated by means of a didactical example.

Keywords Multilevel confirmatory factor analysis · Multilevel mediation · Coleman's bathtub · Group-level outcomes · Measurement isomorphism

---

B. Meuleman (✉)
Centre for Sociological Research
Parkstraat 45, box 3601, 3000 Leuven, Belgium
E-Mail: bart.meuleman@kuleuven.be

🌀 Springer

Mehrebenen-Strukturgleichungsmodelle für ländervergleichende Forschung

Zusammenfassung Dieser Beitrag konzentriert sich auf ein Modell, das in der ländervergleichenden Forschung immer mehr an Bedeutung gewinnt, nämlich die Mehrebenen-Strukturgleichungsmodellierung (MSEM, „multilevel structural equation modelling"). Ähnlich wie Standard-Mehrebenenmodelle (MLM, „multi-level models") unterscheidet dieses Modell zwischen verschiedenen Analyseebenen (z. B. Personen in Ländern) und berücksichtigt die hierarchische Struktur der ländervergleichenden Daten. Bei MSEM werden jedoch latente Variablen in das Mehrebenenmodell integriert, um die Messqualität nicht direkt beobachteter Konstrukte beurteilen zu können. So ist MSEM eine Synthese aus Strukturgleichungsmodellen (SEM „structural equation models") und MLM, die das Beste aus beiden Welten vereint. Der MSEM-Ansatz ermöglicht Mediationsanalysen in Mehrebenendaten und die Modellierung von Outcomes auf der Makroebene und kann daher sehr viel besser das Badewannenmodell von Coleman abbilden. In diesem Beitrag wird der statistische und konzeptionelle Hintergrund von MSEM auf formale, aber leicht zugängliche Weise dargestellt. Es werden Anwendungen von MSEM erörtert, die für ländervergleichende Forschung besonders nützlich sind, nämlich die konfirmatorische Mehrebenen-Faktorenanalyse, Mehrebenen-Mediationsmodelle und Modelle für Outcomes auf der Makroebene. Schritt für Schritt wird anhand eines didaktischen Beispiels erläutert und veranschaulicht, wie MSEM für angewandte Forschung eingesetzt werden kann.

Schlüsselwörter Konfirmatorische Mehrebenen-Faktorenanalyse · Mehrebenenmediation · Colemans Badewanne · Outcomes auf der Makroebene · Isomorphismus der Messung

1 Introduction

Cross-national comparative research (CNCR) is flourishing in various domains of social science thanks to increasing data availability. This is hardly surprising, as comparative approaches have proven to be very effective for investigating the interplay between macro and micro phenomena. Nowadays, quantitative comparativists have a wide range of statistical tools at their disposal (Van de Vijver and Leung 1997; Davidov et al. 2018). Multilevel regression (MLM) has become the method of choice since the 2000s to analyze the hierarchical nature of cross-national datasets and to investigate how contextual factors shape individuals' values, opinions, or behavior (Schmidt-Catran et al. 2019). An alternative approach in CNCR takes structural equation models (SEM) as a point of departure, and places a stronger emphasis on evaluating the validity, reliability, and comparability of measurement instruments (Cieciuch et al. 2019; Davidov et al. 2014).

The rising popularity of CNCR, coupled with the increasing availability of cross-national datasets, has spurred a search for new analytical techniques that are suited to analyze cross-national data in all its complexity. This contribution focuses on

a model that is gaining currency in cross-national research, namely multilevel structural equation modeling (MSEM; Muthén 1989, 1994; Mehta and Neale 2005). MSEM is an attractive strategy because it incorporates a latent-variable approach into the multilevel framework (Skrondal and Rabe-Hesketh 2004). As such, MSEM is a synthesis of SEM and MLM that combines the best of both worlds. On the one hand, the multilevel nature of MSEM makes it possible to distinguish between various levels of analysis, and to study the connection between macro- and micro-level variables. On the other hand, the SEM aspect facilitates testing assumptions about the measurement of constructs and specifying more complex causal chains.

This contribution discusses applications of MSEM for cross-national research. First, the paper reviews the opportunities this approach offers to answer comparative research questions. Next, the statistical background of the MSEM is presented in a formal but accessible manner. Subsequently, the paper zooms in on specific types of MSEM that are particularly useful for CNCR, namely two-level confirmatory factor analysis (CFA), multilevel mediation models, and models for group-level outcomes. Finally, the paper presents a practical step-by-step strategy on how MSEM can be used for applied research, and illustrates this procedure by means of a didactical example.

2 Advantages of MSEM for CNCR

2.1 A More Accurate Representation of the Bathtub Model

Cross-national survey research can be used to address several types of research questions (for a typology, see Billiet and Meuleman 2014; see also Andreß et al. 2019). Probably the most basic application of CNCR consists of testing the extent to which relationships between individual-level variables can be generalized across countries. The most interesting applications of CNCR, however, exploit the fact that cross-national survey data is characterized by various levels of analysis, as it contains information on individuals (the micro-level) that are clustered within particular territorial units (e. g., countries, the macro-level). Relationships between micro and macro level variables are often understood in terms of James Coleman's (1990) "bathtub" or "boat" model (see Fig. 1). This model stipulates that macro-level phenomena are not directly linked to one another, but instead presuppose individuals' agency (Mills et al. 2006). To understand macro-level relationships, it is paramount to uncover the underlying theoretical mechanisms at the micro level (Hedström and Swedberg 1996; see also Kroneberg 2019). In concrete terms, this requires the specification of (1) a situational mechanism detailing how contexts constrain individuals' perceptions and preferences (macro-micro link); (2) an action formation mechanism explaining how these dispositions lead to individual behavior (micro-micro link); and (3) a transformational mechanism specifying how individual behavior is aggregated to macro-level change (Hedström and Ylikoski 2010).

Take for example the oft-studied impact of economic context on the occurrence of revolutions (i. e., two macro-level phenomena). According to Coleman's methodological individualism, poor economic conditions cannot set off a revolution au-
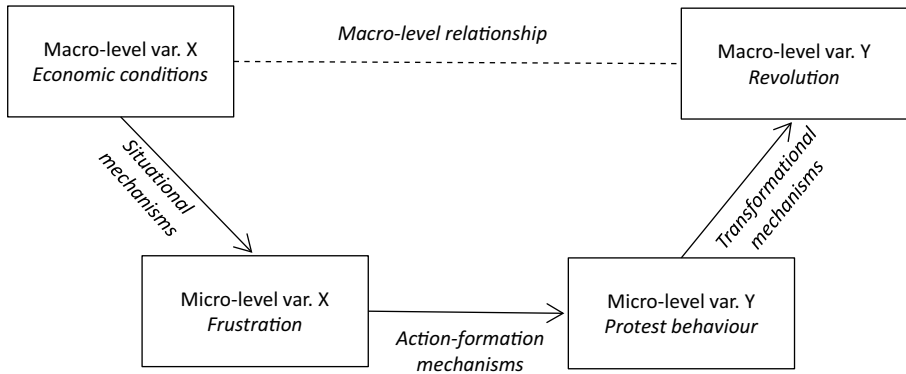
Fig. 1 Illustration of Coleman's bathtub model for the relationship between economic context and revolution. Author's own work. *var.* variable

tomatically, but require intentions, choices, and behavior on the part of individual agents. Demarest et al. (n.d.) argue that unfavorable economic conditions (macro-level X) frustrate the aspirations of individual citizens (situational mechanism), which in turn stimulates protest behavior (action formation mechanism). Under the right conditions, individual protest behavior can be mobilized and aggregated to overthrow the current political power (transformational mechanism).

Understanding social change and macro-level phenomena thus requires a careful investigation of the interplay between individual and contextual variables. In theory, cross-national data—with its various levels of analysis—offers good opportunities to do so. In practice, however, tests of the full bathtub model are extremely rare. CNCR typically uses MLM to investigate how a combination of country-level and individual-level variables affects a particular individual-level dependent variable. This reduces the ingenuity of Coleman's bathtub in two important ways at least. First, the focus in the field of CNCR has been almost exclusively on individual-level dependent variables. While macro-level outcomes constitute the endpoint of the bathtub model, they have been modeled only rarely beyond simple aggregate analysis (Kittel 2006; Andreß et al. 2019). This bias towards individual-level outcomes is not so much driven by a lack of interest in country-level variables—the genesis of specific policies and legislation, institutional arrangements, or collective action, are of prime concern for social scientists. Rather, this pattern is due to the fact that the standard MLM—the standard statistical tool in CNCR—is not equipped to deal with the technical difficulties of modeling macro-level outcomes (Croon and Van Veldhoven 2007). Second, current research only rarely estimates the mediation relationships that are implied by the bathtub model. Assessing the mediation model in a multilevel setting can yield highly relevant information, but again involves additional statistical issues that are not taken into account by the standard MLM approach.

In summary, while Coleman's bathtub is often used to underpin the causal relationships in CNCR, in practice this model is only tested very partially. A crucial advantage of the MSEM approach is that, in contrast to the more popular MLM,

it does permit the full set of situational, action formation, and transformational mechanisms to be tested. By specifying latent variables, MSEM is able to model macro-level outcomes without inducing the biases that are associated with simple aggregation (Becker et al. 2018). Furthermore, MSEM offers a flexible framework that makes it possible to assess multilevel mediation (Preacher et al. 2010). MSEM offers interesting opportunities to estimate direct as well as indirect effects between combinations of micro and macro variables.

2.2 Taking Measurement Error Seriously

In his 1979 address as president of the American Sociological Association, Blalock (1979) identified measurement problems as one of the major obstacles to progress in the social sciences. This statement still rings true four decades on. The validity and reliability of measurements is an issue of crucial importance for any survey. However, the sheer complexity of collecting cross-national survey data (Harkness et al. 2003, 2010) brings along additional challenges for measurement. Comparative survey research hinges crucially on the assumption that measurements are comparable or equivalent (Davidov et al. 2014). Respondents in international surveys were socialized in different cultural backgrounds, speak different languages, and have culture-specific understandings of certain ideas and concepts. There is no guarantee that survey measurements travel successfully across national and cultural borders (Jowell et al. 2007). As a result, the validity and reliability of comparisons of survey measurements is of great potential concern. Comparative researchers have increasingly acknowledged the importance of the comparability of measurements in recent years (Davidov et al. 2014; Cieciuch et al. 2019).

Yet the popular MLM approach does not include a measurement paradigm. MLM is exclusively focused on structural relations between variables, irrespective of how these variables are measured. Operationalization is seen as a separate process detached from substantive analysis, and the quality of measurement is seen as an independent step preceding the actual model estimation. This separation between measurement and structural models is unfortunate because it makes it difficult to take the presence of measurement errors into account, and this can introduce substantial bias into the results (Saris 1998).

The latent-variable approach of MSEM, by contrast, takes as a point of departure the fact that certain abstract and complex concepts cannot be measured in a direct and error-free manner. Instead, multiple fallible indicators are used to represent such concepts. As is the case in regular SEMs, the variance of the indicators is split up into unique variance (measurement error) and the common variance shared between the indicators (capturing the underlying concept). MSEM facilitates modeling the measurement relationships between indicators and latent variables at the same time as modeling structural relationships. As such, MSEM makes it possible to estimate relationships between concepts while controlling for measurement error. Recent studies (Cieciuch et al. 2019; Davidov et al. 2012; Jak et al. 2013) have furthermore illustrated how the MSEM framework can be used to assess the cross-national comparability of measurements.

3 MSEM: Statistical Background and Conceptual Logic

3.1 Point of Departure: SEM

As was mentioned above, MSEM extends the SEM approach by adding a multilevel component. The SEM framework is a good point of departure from which to explain the logic of MSEM (Bollen 1989). SEM basically attempts to model the covariance and mean structures of multivariate data as well as possible. SEM identifies a set of parameters that reproduces the observed data structures in the most accurate way possible. A typical feature of SEM is that besides the observed or manifest variables, latent or unobserved variables can also be defined. As a result, an SEM can include measurement models estimating the relationships between manifest indicators and latent concepts—also known as CFA—as well as the structural relationships of interest. Take for example the model depicted in Fig. 2. This model contains two latent constructs (the η's) that are measured by three manifest indicators each (the Y's). The measurement model can be formally written as:

$$Y_i = \nu + \Lambda \eta_i + \varepsilon_i \tag{1}$$

where subscript i refers to the individual cases. In Eq. 1, Y_i is a vector of observed responses on the indicators. These responses are modeled as a function of latent variables η_i, with matrix Λ containing the regression weights or factor loadings. In this measurement model, vector ν contains the intercepts, while ε_i refers to the residuals or measurement errors.

The appealing feature of SEM is that the model can simultaneously estimate explanation models for the latent concepts. Figure 2 specifies structural relationships between exogenous variables (the X's) and the endogenous latent variables. In formal notation, this structural model equals:

$$\eta_i = \alpha + B \eta_i + \Gamma X_i + \zeta_i \tag{2}$$

where B is a matrix containing the effects among latent variables, and Γ the direct effects of the X's on the latent variables. In this regression equation, α refers to the intercepts and ζ_i to the residuals (unexplained part) of the endogenous variables.

Fig. 2 Example of a single-level Structural Euqation Model (SEM). Author's own work
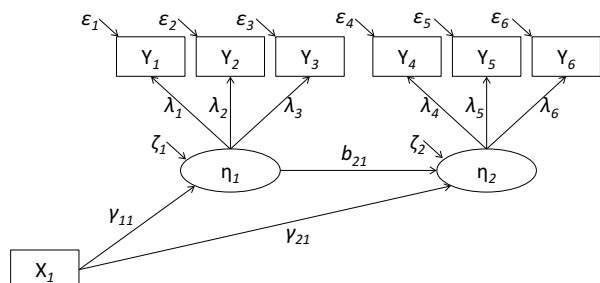
Every set of estimates for the measurement and structural parameters implies a particular covariance matrix (Σ) and mean structure. The most suitable set of parameters is determined by minimizing the statistical distance between the observed means and covariances on one hand, and the model-implied means and covariances on the other. An attractive feature of SEM is that several indices exist to evaluate how successful the model is in replicating the observed data structure, such as the root mean squared error of approximation (RMSEA), the comparative fit index (CFI), the Tucker–Lewis index (TLI), or the standardized root mean square residual (SRMR; Hu and Bentler 1999).

3.2 Decomposition of Covariance Structures

Conceptually, the addition of a multilevel component to the SEM framework can be approached from two different perspectives, each of which draws on different strands of statistical literature (Muthén 1994, pp. 376–377). A first perspective, rooted in sampling theory, refers to the clustered nature of cross-national samples. The standard SEM assumes that the individual errors are distributed independently and identically. This implies that the expected covariance between the residuals for two individual observations should equal zero. Cross-national data clearly violates this assumption of independence, as individuals are clustered within countries (see also Schmidt-Catran et al. 2019). Two individuals living within the same country are expected to be more similar than persons coming from different countries, and the model should reflect this clustered structure. Second, the perspective of random coefficients stresses that certain parameters of interest—such as intercepts (random intercepts) or slopes (random slopes)—might vary across subgroups in the total sample.

Statistically speaking, the point of departure of MSEM is a population of i individuals that are hierarchically nested within g groups (in the case of CNCR: countries). The observed scores can be decomposed orthogonally[1] into a group or between component (i. e., the group average), and an individual or within component (i. e., the deviation from that group average):

$$y_{ig} = \overline{y_g} + (y_{ig} - \overline{y_g}) \tag{3}$$

Based on this decomposition, the total covariance structure can be split into two covariance matrices:

$$\Sigma_T = \Sigma_B + \Sigma_W \tag{4}$$

The within covariance structure, Σ_W summarizes how the individual components are related. The between covariance structure Σ_B describes how the group-level components of the variables covary. Note that decomposition is analogous to how MLM decomposes the dependent variable into individual and country level

---

[1] Orthogonally means here that the variance components at both levels are independent, i. e., they are not correlated. As a result, a clear-cut separation between the two levels is possible.

variation. In MSEM, however, all individual-level predictors involved in the analysis are decomposed as well, which boils down to the group-mean centering of these individual-level predictors (Schmidt-Catran et al. 2019). By default, MSEM thus estimates separate effects for the within- and between-level components of individual variables. This has important repercussions for interpretation: The within-level effects should always be interpreted in terms of differences between individuals *within groups*.

3.3 Adding a Multilevel Component

Rather than analyzing a single covariance structure, MSEM aims to simultaneously model separate covariance structures for the two levels. MSEM formulates within and between models to reproduce the within and between data structures, respectively (Muthén 1994; Mehta and Neale 2005; Rabe-Hesketh et al. 2012). Figure 3 represents a multilevel extension of the single-level SEM depicted in Fig. 2. At the within level, the measurement model is:

$$Y_{ig} = \nu_g + \Lambda_W \eta_{ig} + \varepsilon_{\text{W}ig} \tag{5}$$
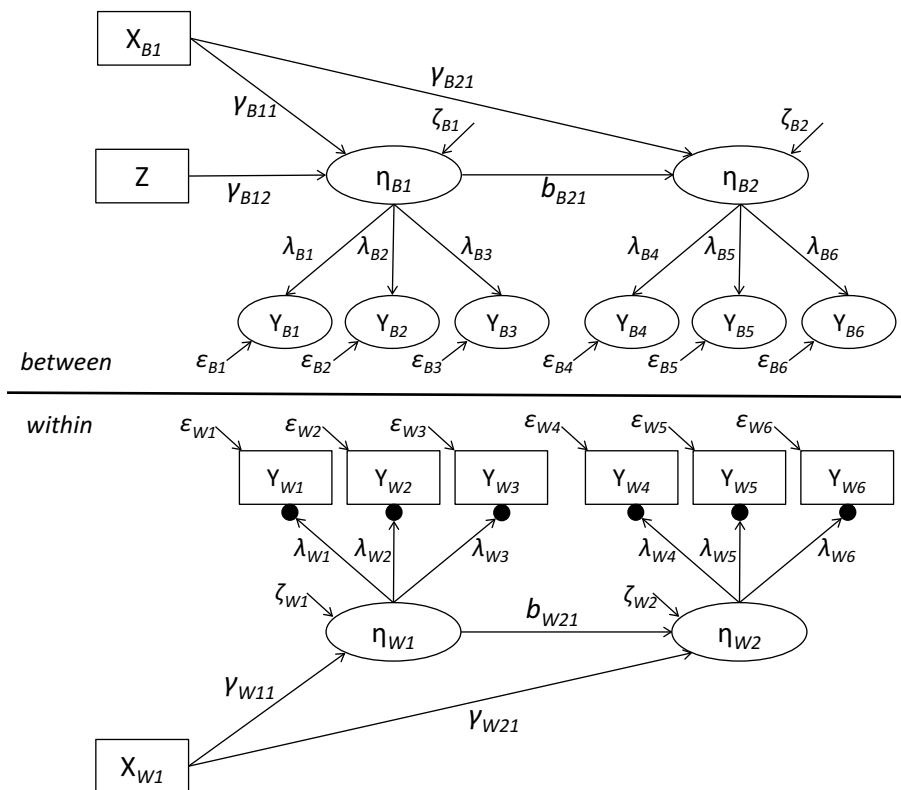
Fig. 3 Example of a two-level Structural Equation Model (SEM). Author's own work

Note that, contrary to Eq. 1, the subscript for response vector Y contains the letters i and g, referring to individuals clustered within groups. These responses are a function of within-level latent variables η_{ig}, and the within factor loadings Λ_W capture the strength of the within-level measurement relations. The multilevel aspect is further incorporated by the fact that the item intercepts ν_g are not constant but vary across groups (hence the subscript g). In other words, this MSEM contains six random intercepts—one for every observed indicator. Consistent with the graphical notation used in the Mplus users' guide (Muthén and Muthén 1998–2017), Fig. 3 depicts random coefficients by means of a black dot. The random intercept variances constitute latent variables in the between model (and for that reason, the indicators are depicted as circles rather than rectangles in the between model). At the between level, the random intercepts are modeled as a function of between-level latent variables η_g:

$$\nu_g = \upsilon + \Lambda_B \eta_g + \varepsilon_{Bg} \tag{6}$$

where υ is a vector with fixed (grand) intercepts, Λ_B contains the between factor loadings, and ε_{Bg} refers to the group-level residuals of the indicators.

Regarding the structural part of the model, at the within level as well as at the between level, the latent variables are regressed on predictors—similarly to those in Eq. 2. The within structural model equals:

$$\eta_{ig} = \alpha_W + B_W \eta_{ig} + \Gamma_W X_{ig} + \zeta_{\text{W}ig} \tag{7}$$

For the between level, the structural model is:

$$\eta_g = \alpha_B + B_B \eta_g + \Gamma_B X_g + \zeta_{Bg} \tag{8}$$

In these equations, B_W and B_B refer to the effects among the endogenous variables at the within and between levels. Γ_W and Γ_B contain the direct effects of exogenous within and between predictors. In the model in Fig. 3, the only exogenous predictor at the within level is the group mean-centered component of variable X_1, that is X_{W1}. At the between level, the group component X_{B1}, as well as contextual variable Z, serve as predictors.

This example makes it clear that MSEMs contain within-level parameters that describe the structure of within-group variables (thus, differences between individuals within groups), while the between-level parameters describe relationships among between variables (that is, group averages of individual variables as well as country-level variables). The within parameters reproduce within covariance structure Σ_W, while the between parameters imply covariance structure Σ_B. To determine the optimal set of parameters, the discrepancy between, on the one hand, model-implied Σ_B and the observed between covariance structure S_B, and on the other hand model-implied Σ_W and the so-called observed pooled-within covariance matrix S_W is minimized. This pooled-within covariance matrix constitutes a weighted average of the within structures that are observed in each of the groups. By default, maximum likelihood-based estimation procedures are used to estimate the parameters

(Muthén 1994). Similar to the case of traditional SEM, a wide range of fit indices is available to evaluate how closely the model-implied covariances fit the observed data structures.

Figure 3 obviously shows only one MSEM out of a wide range of possibilities (for a more general presentation of the MSEM framework, see du Toit and du Toit 2008; Rabe-Hesketh et al. 2004; Mehta and Neale 2005). The model can be further extended, for example, by including additional random components. In principle, each of the parameters in the within model can be allowed to vary across groups, and can be included as a random coefficient. Every additional random component results in a new latent variable at the between level, representing the country-level variation of a within parameter.

4 Useful Applications of MSEM for CNCR

Sect. 3 made it clear that the MSEM approach consists of a wide variety of possible applications. Some of these models are of particular relevance for cross-national research. This section presents three such models—namely multilevel CFA (MCFA), multilevel mediation models, and models for group-level outcomes—in greater detail. In doing so, the general logic of the MSEM approach will be further clarified.

4.1 Multilevel Confirmatory Factor Analysis (MCFA)

4.1.1 The MCFA Model

Multilevel CFA (MCFA) is essentially the part of the MSEM focusing on how latent constructs are measured by multiple observed indicators. Conventional single-level CFA aims to explain the covariances among observed indicators Y by assuming that the indicators reflect a smaller number of latent variables η (Bollen 1989; see Eq. 1). Following the general logic set out in the previous section, this model can be extended to a two-level CFA (or even a three-level CFA, see Jak 2014). MCFA takes into account that individual observations (i) are not independent, but instead are clustered in groups (g). MCFA makes a distinction between the individual and country levels, and formulates a separate measurement model for both levels (see Eqs. 5 and 6). By substituting Eq. 6 into 5, we obtain the full two-level CFA model in reduced form:

$$y_{ig} = \upsilon + \Lambda_W \eta_{ig} + \Lambda_B \eta_g + \varepsilon_{\mathrm{Wig}} + \varepsilon_{Bj} \tag{9}$$

The variables and parameters at both levels should be distinguished conceptually because they tell different stories about the data structure. The within latent variables η_{ig} capture how individual features vary within countries, and therefore refer to the scores of individuals on the latent variables. The within factor loadings Λ_W are parameters of prime interest as they reflect measurement quality at the individual level. These factor loadings reflect how the within components of the indicators change when a particular latent variable at the within level increases by one unit.

The within residual variances indicate how much random measurement error is present in the within level indicators. The between level, conversely, is completely geared towards explaining aggregate data patterns at the country level. The between latent variables η_g explain the covariances between the country averages of the indicators—that is, the random intercepts of the indicators. The between-level latent variables therefore reveal the position of countries rather than individuals on the latent constructs. Note that the factor structure is identical across levels in this example. As Sect. 4.1.3 will demonstrate, however, this is not necessarily the case. The between factor loadings and residual variances are indicative of the quality of the measurement process at the between level. Importantly, the measurement quality can diverge across levels (see below). Most of the random measurement errors in cross-national surveys take place at the individual level (after all, it is individual respondents who complete questionnaires), while estimates at the country level are quite reliable. As a result, the amount of within residual variance generally exceeds the between error variances by far. Finally, the vector of grand item intercepts—υ in Eq. 9—refers to the expected value for the observed indicators, conditional on the within and between latent constructs being equal to zero. If the latent constructs are standardized (and thus have mean zero), these grand intercepts equal the expected response for the average respondent in the average country, and thus reproduce the observed item means.

The two-level CFA reproduces the within covariance matrix Σ_W and the between covariance matrix Σ_B as shown below (Muthén 1994, p. 382):

$$\Sigma_B = \Lambda_B \Psi_B \Lambda'_B + \Theta_B \tag{10}$$
$$\Sigma_W = \Lambda_W \Psi_W \Lambda'_W + \Theta_W \tag{11}$$

where Ψ_B and Ψ_W denote variance–covariance matrices of the latent factors at the between and within levels, respectively, Θ_B and Θ_W are diagonal matrices with between- and within-level residual variances. These equations make the variance decomposition into a within and between component very explicit. Thus, MCFA not only decomposes the observed items into a within and a between component, but also makes it possible to contrast the variance components of the latent variables across levels. Muthén (1991, p. 345) shows how for each latent construct the "*true intraclass correlation coefficient*" (ICC) can be determined by dividing the between variance of a latent variable by its total variance (i.e., the sum of between and within variance). This latent ICC estimates the amount of variation of a concept located at the country level, controlling for random measurement error. Given that most of the error is generally located at the individual level, the latent ICCs are usually considerably higher than the ICCs of the separate items.

4.1.2 Cross-National Measurement Equivalence from a Multilevel Perspective

More recently, various studies have shown how MCFA can be exploited to gain additional insights into the comparability of measurements, either across countries or across levels. In the field of CNCR, the issue of comparability of measurements across different countries has become increasingly acknowledged (Davidov et al.

2014). Because comparative surveys are fielded in a variety of countries with different linguistic and cultural contexts, the comparability of measurements is not guaranteed, but should be tested. Multigroup CFA (MGCFA) has become the technique of choice to do so (Steenkamp and Baumgartner 1998; Vandenberg and Lance 2000). Basically, the MGCFA approach estimates a separate measurement model per group (in the case of CNCR: country), and subsequently evaluates the similarity of the measurement models across groups. Three levels of measurement equivalence are usually distinguished within the MGCFA framework: (1) Configural equivalence implies that factor structures are equal across groups and guarantees that construct bias is absent (van de Vijver 1998); (2) Metric equivalence requires that factor loadings be equal, and guarantees the equality of scale intervals of the latent variable; (3) Scalar equivalence additionally assumes equality of item intercepts and is a precondition for comparing latent means (see Cieciuch et al. 2019).

These levels of cross-national measurement equivalence can also be approached from a multilevel perspective (Jak et al. 2013, 2014; Davidov et al. 2012). To make this point clear, it is necessary to highlight some differences and similarities between the two approaches. In the multi-group model, a separate variance–covariance matrix is calculated per group. By consequence, MGCFA models within-group covariance structures, and discards all between-group relationships. MCFA models differences between individuals within groups in the within part of the model, but takes a different approach to within-group covariances. Rather than modeling a separate within-matrix per group, a single pooled-within variance–covariance matrix (Σ_W) is analyzed. Importantly, this pooling operation assumes that the within-level covariance structures are similar across groups. The assumption of equal covariance structures implies that within-level factor structures are also equal and, by consequence, that measurements are configurally equivalent. The assumption of configural equivalence is thus a blind spot for the MCFA model: Even if cross-group differences in a within-level factor structure were present in the data, they would become intractable by pooling the within structures.

A similar argument can be made for metric equivalence. MCFA estimates a single set of within-group factor loadings that hold for all countries in order to reproduce the pooled-within covariances. As a result, the model assumes that within-level factor loadings are equal across groups (i.e., metric equivalence). However, a random coefficient can be added to the standard MCFA model to discard the hypothesis of metric equivalence. By adding a random slope for one or more factor loadings, the strength of indicators can vary across groups. As a result, the model incorporates deviations from metric equivalence (Asparouhov and Muthén 2012; Meuleman and Schlüter 2018).

Scalar equivalence is evaluated in MCFA by assessing the between-level residual variances (Jak et al. 2013; Davidov et al. 2012). These residuals contain the cross-national deviations in item intercepts that are not captured by the between-level latent variable, and are thus unique for one particular item. Non-zero between-level errors indicate that the group average deviates from what can be expected from the mean of the between-level latent variable (Davidov et al. 2012, p. 563), and that a particular item functions differently across groups. In other words, MCFA can

detect deviations from scalar equivalence by testing whether between-level residuals differ from zero.

Probably the most appealing feature of the multilevel approach is that MCFA cannot only be used to detect deviations from equivalence, but also offers opportunities to explain measurement equivalence by means of contextual variables. This feature derives directly from the fact that the multilevel approach models inequivalence as random coefficients (random factor loadings for metric equivalence and random intercepts in the case of scalar equivalence). These random coefficients figure as unexplained variance in the between model. This conceptualization opens the door to attempts to gain a deeper insight into the processes leading to measurement inequivalence. Meuleman and Schlüter (2018), for example, use MCFA to assess the metric equivalence of the ISSP scale measuring citizenship conceptions. Using a random loading model, they find substantial cross-national variance for the factor loading of the item stating that "*To have been born in [COUNTRY]*" is an important precondition to be considered as a "true citizen". Interestingly, this random loading variance is partly explained by the percentage of immigrants in the country, indicating that the item becomes more discriminatory in high-immigration contexts. Analyzing universalist values, Davidov et al. (2012) similarly showed how the human development index (HDI) can explain cross-country differences in the random intercept of an item measuring the importance of the environment. People in less developed countries react more sensitively to this item, and consequently the item's origin (intercept) is higher than in countries with higher levels of HDI. Both examples illustrate how MCFA can yield useful information to interpret the source of measurement bias in CNCR.

4.1.3 Using MCFA to Test Isomorphism

Besides comparisons of measurements across countries, the multilevel structure of cross-national data also facilitates comparisons across levels (individual vs. country level). Yet, while the importance of the comparability of measures is widely acknowledged, this is less the case for cross-level comparability (also referred to as isomorphism; Fontaine 2008; Fontaine and Fischer 2010; Ruelens et al. 2018). Cross-group and cross-level equivalence are conceptually and operationally two very different issues, with diverging consequences for comparability. While cross-group equivalence refers to the structural equivalence of a construct at the individual level in each of the cultural groups, isomorphism relates to the similarity of the within and between measurement models. Multilevel studies often aggregate constructs measured at the individual level to draw inferences at the country level, but rarely discuss whether such cross-level generalizations are justified. Nevertheless, isomorphism is a necessary condition for drawing valid conclusions that involve generalizations across levels (Ruelens et al. 2018), as is done in the classical MLM approach.

Isomorphism implies in concrete terms that the associations between indicators and latent constructs at the individual level (thus within countries) are similar to the associations between indicators and latent constructs at the country level (i.e., between countries). For obvious reasons, MCFA offers good opportunities to test isomorphism. Similar to the hierarchically ordered levels of cross-group measure-
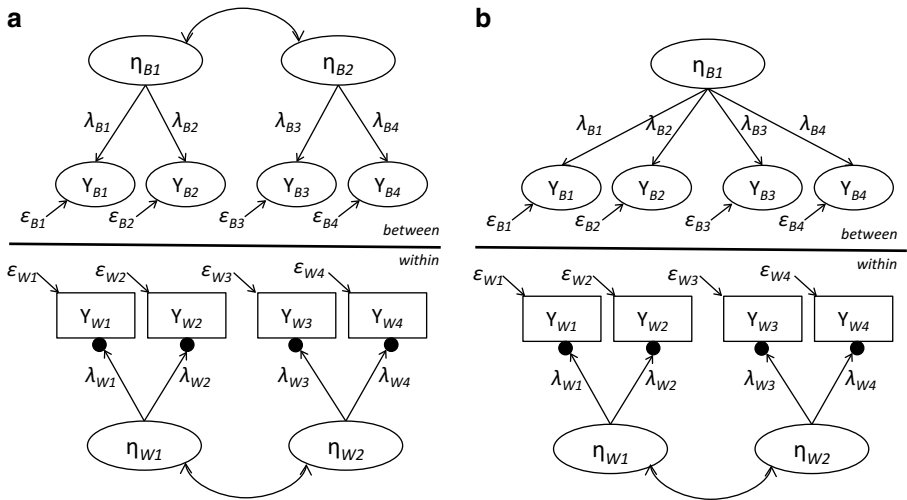
Fig. 4 Isomorphism in two-level CFA models. **a** configural isomorphism, **b** configural non-isomorphism. Author's own work

ment equivalence, the literature distinguishes between various levels of isomorphism (Tay et al. 2014; Ruelens et al. 2018) – see Fig. 4. First, when the dimensionality of factor structures is identical across levels, one can conclude that *configural isomorphism* holds. This implies that the items exhibit the same configuration of factor loadings between the latent construct and the indicators at the individual and at the country level of analysis. In formal terms, configural isomorphism is tested by specifying identical within and between measurement models in an MCFA:

$$H_0 : pattern\,(\Lambda_W) = pattern\,(\Lambda_B) \tag{12}$$

As a consequence, the constructs under study have, qualitatively speaking, the same meaning at the individual and country levels. Quantitative comparisons, however, require that the latent constructs be measured on an identical scale, which demands factor loadings to be identical across levels (Dyer et al. 2005). This level of isomorphism is termed *metric isomorphism* (Tay et al. 2014; Ruelens et al. 2018):

$$H_0 : \Lambda_W = \Lambda_B \tag{13}$$

When metric isomorphism holds, an increase of one unit on the measurement scale has the same meaning at the individual level as at the group level (Mehta and Neale 2005). This is a precondition to decompose the variance of latent variables. Note that multilevel regression models decompose the variance of dependent variables, and therefore in fact assume that this dependent variable is measured in an isomorphic manner.

Tests for isomorphism are still rare, which has inspired Byrne and Van de Vijver (2014, p. 170) to call the study of measurement isomorphism "*probably the most*

underrated topic in cross-cultural research methods." One exception is the study by Ruelens et al. (2018), who use MCFA to show that the ESS scale for political trust is—to a large extent—metrically isomorphic, which legitimizes the practice of transposing the concept of political trust to the country level and comparing countries on their average trust. An interesting example of non-isomorphism can be found in cross-cultural research on value patterns. Schwartz (1994) observes that fewer value dimensions can be distinguished at the country level than among individuals (Fischer et al. 2010; Fischer 2012). Clearly, individuals have more fine-tuned value preferences, while a crude classification of values is sufficient to describe culture-level differences. In such a situation of non-isomorphism, it could be clearly misleading to simply aggregate individuals' value priorities to the country level, and draw conclusions on that basis.

4.2 Multilevel Mediation Models

The idea of multilevel mediation—with direct as well as indirect relationships between variables at various levels—is inherent to the bathtub model (Coleman 1990) that guides many cross-national studies. A multitude of different mediation constellations are possible. Take the simplest situation with three variables, namely an independent variable X, a dependent variable Y, and an intermediary variable M that partially channels the relation between X and Y (thus: $X{\rightarrow}M{\rightarrow}Y$). Each of these three variables can be either individual characteristics (level 1) or country-level variables (level 2). As a result, eight different mediation designs are possible (1-1-1; 2-1-1; 1-2-1; 2-2-1; 1-1-2; 2-1-2; 1-2-2; 2-2-2; Preacher et al. 2010). This section focuses on models with individual-level dependent variables (group-level outcomes are dealt with in Sect. 4.3). Among the four remaining designs, the 2-1-1 model—with a country level-independent variable and an individual-level mediator and dependent variable—is most useful for CNCR, as the 2-1-1 logic fits perfectly with the first three steps of Coleman's bathtub: A macro-level phenomenon conditions individual preferences (macro-micro link), which can subsequently instigate individual behavior (micro-micro link; see Fig. 1).

Various authors have proposed to analyze multilevel mediation models by means of MLM techniques (Krull and MacKinnon 1999; Zhang et al. 2009). Similar to the single-level case (Judd and Kenny 1981), this approach estimates several models—with and without the mediator M—and the difference between parameter estimates for X indicates the extent to which the relationship between X and Y is mediated by M. The indirect effect is calculated as the product of two effects ($X{\rightarrow}M$ and $M{\rightarrow}Y$) that stem from different models. In the multilevel case, however, this approach contends with a serious problem, namely the conflation of within and between effects (Preacher et al. 2010). To illustrate this problem, take the example of the 2-1-1 design. By definition, the effect of country level X on individual level Y can only be an effect taking place at the between level. After all, X only varies between groups (no within-group variation), and can thus only affect the between-group component of Y. As a consequence, the mediation can only run via the between component of M, and only the between effect of M on Y is relevant to assess the mediation (Preacher et al. 2010, p. 210; for a similar argument see Schlüter and

Fig. 5 Multilevel mediation for a 2-1-1 design. Author's own work
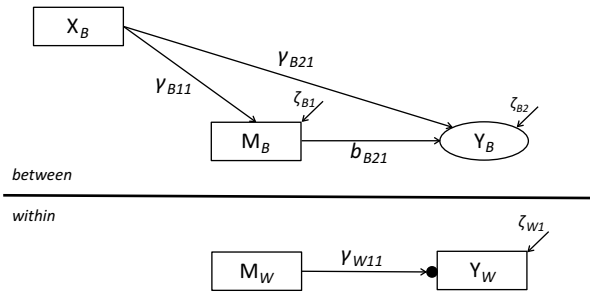
Wagner 2008). Yet the standard MLM approach does not separate the within and between components of the *M-Y* relationship, but instead estimates the so-called total effect, which is a mixture of the within and the between effects (Raudenbush and Bryk 2002). This conflation can introduce bias into the estimation of the indirect effect, and the bias will be stronger as the within and between effects diverge more. If the effect of *M* on *Y* for individuals within countries is weaker than the effect for country averages (which is often the case in CNCR because within relationships tend to be attenuated by larger measurement errors), the MLM approach will underestimate the mediation effects. A similar problem occurs in all mediation designs that contain a 1-1 component (that is, a micro-micro link).

The conflation bias can be solved by separating the between from the within effects, and the MSEM framework is particularly well suited to do so (Preacher et al. 2010). Below, we elaborate this model for the case of the 2-1-1 design (see Fig. 5). For reasons of clarity, we assume that all variables in the mediation are manifest (so that no measurement model is necessary). The resulting 2-1-1 model is a special case of the structural models presented in Eqs. 7 and 8. Two variables figure in the within model, namely *M* and *Y*.

$$Y_{ig} = \alpha_W + \gamma_{W11} M_{ig} + \zeta_{Wig} \qquad (14)$$

At the between level, there is mediation between country-level variable *X* and the between components of *M* and *Y* (Cheong and MacKinnon 2012):

$$\begin{bmatrix} M_g \\ Y_g \end{bmatrix} = \begin{bmatrix} \alpha_{B1} \\ \alpha_{B2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \beta_{B21} & 0 \end{bmatrix} \begin{bmatrix} M_g \\ Y_g \end{bmatrix} + \begin{bmatrix} \gamma_{B11} \\ \gamma_{B21} \end{bmatrix} [X_g] + \begin{bmatrix} \zeta_{B1g} \\ \zeta_{B2g} \end{bmatrix} \quad (15)$$

This elaboration makes it clear that the mediation takes place at the between level only—after all, the within level contains two variables only. The between indirect effect can be calculated as the product of the between effects of *X* on *M* and of *M* on *Y* ($\gamma_{B11} \times \beta_{B21}$). The between total effect of *X* on *Y* equals $\gamma_{B21} + \gamma_{B11} \times \beta_{B21}$.

A rare but interesting application of the MSEM approach to mediation can be found in Schlüter and Wagner (2008), who investigate how the size of the immigrant population at the regional level affects intergroup contact and perceived threat, which then, in turn, shape outgroup derogation (thus, a 2-1-1-1 design). The MSEM analysis indeed illustrates that the between components of the two 1-1 links are

considerably stronger than the within parts. Within regions, the impact of intergroup contact on perceived threat is –0.33, while this effect equals –0.89 between regions. Similarly, the effect of threat on outgroup derogation is 0.38 at the within level and 0.68 at the between level. The indirect effect is thus considerably stronger at the between level, and an MLM approach would have underestimated the mediation substantially.

### 4.3 Models for Group-Level Outcomes

Outcome variables at the country level are potentially of high relevance in comparative research. Social scientists are, for example, highly interested in explaining the level of social expenditure in a country (Kittel 2006), the persistence of democratic systems (Becker et al. 2018), or the rise of protest movements (Demarest et al. n.d.). However, quantitative cross-national studies rarely analyze such macro-level phenomena (Andreß et al. 2019). This is surprising, given that the macro level is the endpoint of Coleman's bathtub model.

One possible explanation for this relative neglect might be the unfamiliarity of models to analyze group-level outcomes. The standard MLM approach is clearly geared towards explaining individual-level outcomes. The most common approach to cope with group-level outcome variables has been aggregation or disaggregation of the data (Croon and Van Veldhoven 2007). In the case of disaggregation, the group-level variable scores are assigned to individuals and subsequently analyzed at the individual level. Aggregation involves calculating group averages for the individual-level predictors and estimating effects at the group level. However, Croon and Van Veldhoven (2007) show how both strategies lead to inaccurate estimation. Disaggregation leads to violation of the independence assumption and tends to produce pseudo-significant results. Aggregate analysis fails to take the individual heterogeneity underlying the group averages into account. The group averages of the predictor variables contain sampling errors because they are calculated on the basis of a sample of individual respondents (Marsh et al. 2009). As a result, aggregate data analysis ignores a source of variability in the measurement of the predictors and produces biased estimates. Aggregate estimates are only unbiased when the sampling error of the aggregated predictors equals zero (Croon and Van Veldhoven 2007).

To resolve these statistical issues, multilevel models for group-level outcomes should "*explicitly acknowledge the existence of these different levels and [...] attempt to formulate the interaction between the levels in the production of the outcome variable*" (Croon and Van Veldhoven 2007, p. 46). In other words, appropriate modeling strategies should take into account the process of aggregation of the individual-level predictors. MSEM is highly-suited to this purpose, as it explicitly decomposes the variation of individual predictors into a within and a between component (Bennink et al. 2013; Becker et al. 2018). Marsh et al. (2009) refer to this process as latent aggregation.

A very simple example of this approach is an MSEM explaining a group-level variable Y by means of a country-level variable Z and an individual-level predictor X

Fig. 6 Multilevel Structural Equation Model (MSEM) for a group-level outcome variable
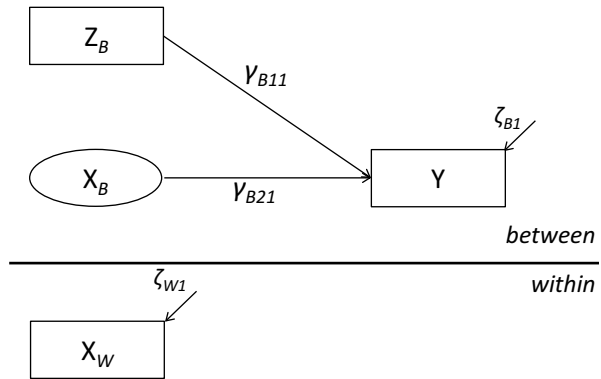
(see Fig. 6). The within model is empty in the sense that it does not contain regression effects, but merely decomposes the total variance of X:

$$X_{ig} = \alpha_g + \zeta_{Wig} \tag{16}$$

In this equation, the random intercept α_g constitutes the group-level component of X. At the between level, the between components of X as well as Z are used to explain Y:

$$Y_g = \alpha_B + \gamma_{11} Z_g + \gamma_{11} \alpha_g + \zeta_{Bg} \tag{17}$$

An illustration of this approach is presented by Becker et al. (2018), who explain the persistence of democracies by means of country variables such as GDP and ethnolinguistic fractionalization, and individual predictors such as support for democratic values, social class, and education.

5 Sample Size Requirements: Bayesian Estimation to the Rescue?

The previous section showed that MSEM has the potential to offer new insights into cross-national data. However, the MSEM approach also has a major disadvantage: The requirements for sample size at the highest level are difficult to meet in comparative research (for a similar argument in the case of MLM, see Schmidt-Catran et al. 2019). A simulation study by Meuleman and Billiet (2009) has shown that maximum likelihood (ML) estimation of a relatively simple MSEM (with 1 latent factor at both levels and 1 contextual predictor) requires 40 groups, which is more than typical cross-national datasets offer. As model complexity increases, so do the sample size requirements. When the number of groups is too small, between parameters tend to be overestimated and standard errors underestimated. This combination of parameter and standard error bias tends to produce poor coverage and pseudo-significant effects.

Several strategies exist to mitigate this problematic situation. One possibility is to reduce the complexity of the between model. One can, for example, decide to

leave out the between components of individual variables when this is theoretically justified and these between effects are not of prime interest (Preacher et al. 2010, p. 215). For measurement models, imposing cross-level constraints on the factor loadings (i.e., metric isomorphism) reduces the number of parameters that need to be estimated, and can thus improve estimation accuracy (besides enhancing comparability across levels).

A more promising avenue consists of switching from ML to Bayesian estimation. Using a similar simulation setup as Meuleman and Billiet (2009), Hox et al. (2012) conclude that the performance of the Bayesian estimation of MSEMs causes considerably smaller amounts of bias for between parameters. Even country-level sample sizes as small as 20 already yield accurate statistical inference (for similar findings in the case of MLM, see Stegmueller 2013). Bayesian estimation combines prior knowledge (i.e., the prior distribution) with observed evidence (in the form of the likelihood of data given a set of parameters) to produce a so-called posterior distribution of parameter estimates. This posterior distribution expresses the amount of uncertainty regarding the parameters that still exists after having observed the data. In the case of complex models such as MSEM, the posterior distribution is not solved analytically, but is simulated using Monte Carlo Markov chains (MCMC). The posterior distributions (and the derived 95% credibility intervals) are obtained by means of an iterative procedure that estimates parameters repeatedly. As such, Bayesian estimation does not make assumptions about the distribution of test statistics, and is not dependent on large-sample theory (Kaplan and Depaoli 2012; van de Schoot et al. 2014). As such, switching from a frequentist to a Bayesian framework might also be helpful in dealing with the issue that countries in CNCR are not a random sample from a large population (for a critical discussion of this issue, see Schmidt-Catran et al. 2019).

6 A Step-by-Step Approach for MSEM Estimation

The multilevel component of MSEM adds an additional layer of complexity to model estimation, and confronts data analysts with a plethora of decisions. To streamline this process, this section provides a step-by-step procedure for the estimation of MSEMs that can be adjusted to the specific needs of the model. This procedure is furthermore illustrated by means of a didactic example based on Baute et al. (2018), who analyze European Values Study data to uncover individual and contextual sources of Euroscepticism. Because Euroscepticism is a latent variable (measured by five indicators), MSEM is very suitable. The syntax to estimate the example models in Mplus is provided in the online appendix.

6.1 Step 1. Formulating the Measurement Model

Prior to estimating the full model, it is advisable to assess the validity and reliability of the latent variables involved. According to the guidelines provided by Muthén (1994, p. 390 et seqq.), this step can be broken down into various substeps (see also Hox et al. 2010; Jak et al. 2014). First, explore the pooled data structure by

Table 1 Multilevel Confirmatory Factor Analysis (MCFA) results for the Euroscepticism scale ($N_{country} = 28$; $N_{individual} = 40,995$)—metric isomorphism

	Within		Between	
	Factor loading	SE	Factor loading	SE
The loss of social security	0.711	0.012	0.766	0.041
The loss of national identity and culture	0.770	0.010	0.946	0.021
Our country paying more to the EU	0.745	0.012	0.834	0.049
A loss of power in the world for (country)	0.764	0.013	0.911	0.027
The loss of jobs	0.761	0.011	0.712	0.065

Chi-square: 231.96; Df: 13; RMSEA: 0.020; CFI: 0.979; TLI: 0.968; SRMR(within): 0.017; SRMR(between): 0.096

An error correlation between loss of national identity and the loss of jobs was specified at the within level.

performing a conventional factor analysis on the total covariance structure, without distinguishing the within from the between level. This step provides an indication of how many factors to expect. Second, evaluate the amount of between variation by calculating the ICCs for the manifest indicators. Only if substantial variation is present at the country level does it make sense to proceed with a multilevel analysis. Third, formulate a suitable factor model for the within part of the model. To make sure that misspecifications at the between level do not distort this, it is best to either remove all between variance from the indicators[2] or saturate the between model (i. e., add between parameters until the number of degrees of freedom is 0). Fourth, after a well-fitting within model is found, estimate the factor model at the between level. A good starting point is frequently the working hypothesis that the factor structure is identical across levels (configural isomorphism), and that the same concepts are measured at both levels. Depending on the results, it might be necessary to change the factor structure at the between level, e. g., by reducing the number of between factors. In that case, configural isomorphism does not hold, so that the meaning of the concept cannot be generalized across levels. Fifth and finally, if configural isomorphism holds, it is advisable to test metric isomorphism (i. e., cross-level equality of factor loadings) as well (Ruelens et al. 2018). Metric isomorphism not only enhances cross-level comparability, but also reduces model complexity, and therefore can improve the accuracy of estimates.

Table 1 presents the final results from a two-level CFA (with metric isomorphism) for the five EVS items measuring Euroscepticism (Baute et al. 2018). These items ask respondents to state the degree to which they fear that EU integration has undesirable consequences, such as the loss of social security, jobs, or national identity. Answers are recorded on a 10-point scale (from 1 = not afraid at all to 10 = very much afraid). The ICCs for these items range between 4.3 and 9.7%. Robust maximum likelihood estimation (MLR) is used, and factor loadings are constrained to be equal across levels (metric isomorphism). All standardized factor loadings are very strong (>0.70), especially at the country level. The latter indicates that the between level is less contaminated by measurement error. Note that even in the case of metric

---

[2] In Mplus, this is done by specifying the indicators at "within" variables in the "variable" statement.

isomorphism, the standardized loadings can differ across levels. After all, the equality constraints are placed on the unstandardized parameters. The fit indices show that this model gives an appropriate description of the multilevel data: RMSEA is well below common cut-off points, and CFI and TLI are sufficiently close to 1. The SRMR at the between level equals 0.096, which is slightly higher than the 0.08 cut off proposed by Hu and Bentler (1999). Removing the cross-level constraints on the factor loadings reduces the between SRMR to 0.053, but does not lead to a real improvement in terms of the other fit indices (RMSEA: 0.022; CFI: 0.983; TLI: 0.968). All in all, we can conclude that the five items are valid and reliable indicators of Euroscepticism, both at individual and country levels. The evidence for metric isomorphism makes it clear that the concept is similarly measured at both levels.

6.2 Step 2: Estimating the Structural Model

Once the quality and comparability of measurements is established, the structural relations can be added to the model. Similar to the estimation of the measurement model, it can be helpful to start with the within variables, and add the between variables only at a later stage.

In the didactic example, the structural model predicts Euroscepticism by a number of individual and contextual predictors (see Table 2; this is a simplified version of Baute et al. 2018; see the online appendix for the syntax). At the within level, ethnocentrism turns out to be the most influential predictor of Euroscepticism. Besides, Euroscepticism is also more widespread among females, the lower educated, in the lowest income groups, and among those favoring state responsibility. Note that these are pure within effects, referring to differences between individuals living within countries (equivalent to country mean-centered effects in multilevel regression). Only one of the between effects is significant: Euroscepticism is more widespread in countries with more social spending. In principle, it would be possible to include the between components of individual variables at the between level. To reduce model complexity, however, this option was not pursued here. At both levels, the predictors explain about 20% of the variation in Euroscepticism. This full MSEM also has a good model fit.

6.3 Step 3: Robustness Checks: Bayesian Estimation and Simulation

The small sample size at the country level is the Achilles heel of MSEM for CNCR. Therefore, it is good practice to assess the accuracy of estimation. A straightforward way to do this is to repeat the standard maximum likelihood-based analysis with a Bayesian estimator.[3] A more elaborate but advisable alternative consists of

---

[3] It is of course possible to start with a Bayesian estimator from the outset; however, the Bayesian estimation procedure implemented in Mplus has the disadvantage that it offers few useful tools to assess model fit.

Table 2 Structural part of a Multilevel Structural Equation Model (MSEM) explaining Euroscepticism ($N_{country}$ = 28; $N_{individual}$ = 39,483)

	Std. Par.	SE	*p*-value
Within model			
Age	–0.010	0.014	0.4750
Gender (ref = male)	0.066	0.009	0.0000
Education			
Lower secondary	0.131	0.014	0.0000
Higher secondary	0.114	0.011	0.0000
Tertiary (ref)	–	–	–
Income			
1st quartile	0.069	0.011	0.0000
2nd quartile	0.065	0.009	0.0000
3rd quartile	0.033	0.007	0.0000
4th quartile (ref)	–	–	–
Missing	0.051	0.010	0.0000
Employment status			
Paid employment (ref)	–	–	–
Retired	–0.022	0.009	0.0110
Student	–0.037	0.005	0.0000
Unemployed/disabled	0.009	0.007	0.1960
Others	–0.010	0.005	0.0570
Pro-state responsibility	0.092	0.015	0.0000
Ethnocentrism	0.384	0.021	0.0000
R-squared	0.199		
Between model			
Spending on social benefits (% GDP)	0.489	0.167	0.0030
Transfers received from EU (% GDP)	0.113	0.165	0.4940
EU immigrants (per 1000 inh.)	0.013	0.118	0.9150
R-squared	0.190		

Chi-square: 871.13; Df: 81; RMSEA: 0.016; CFI: 0.971; TLI: 0.962; SRMR(within): 0.009; SRMR(between): 0.081

performing a Monte Carlo simulation study[4] to determine parameter bias, standard error bias, and coverage of the confidence intervals (Muthén and Muthén 2002).

Table 3 presents relevant indicators of estimation accuracy for a selection of parameters, for both the default maximum likelihood and Bayesian estimation (see the online appendix for the Mplus syntax). The results indicate that within estimates for both estimators are quite accurate. At the within level, parameter bias (that is the percentage deviation between the average parameter estimate over all replications

---

[4] The basic idea of a Monte Carlo simulation study is that a large number of random samples are generated according to a population model. Subsequently, each of the generated random samples (replications) is analyzed, and the results obtained are compared to the true population parameters. This procedure makes it possible to assess how accurate estimation and statistical inference are. See Muthén and Muthén (2002) for more details.

Table 3 Results from a Monte Carlo simulation study for the MSEM explaining Euroscepticism—ML vs. Bayesian estimation (500 replications each)

	Maximum likelihood estimation			Bayesian estimation		
	Parameter bias (%)	Standard error bias (%)	95% coverage (%)	Parameter bias (%)	Standard error bias (%)	95% coverage (%)
Within model						
Factor loading	0.0	–3.0	93.9	–0.1	1.4	94.7
Residual variance	0.0	–0.9	94.5	0.0	0.8	94.6
Structural effect	–0.1	–4.6	93.3	0.4	1.2	95.0
Between model						
Factor loading	0.0	–3.0	93.9	–0.1	1.4	94.7
Residual variance	–4.6	–8.9	86.7	15.5	20.4	94.5
Structural effect	–1.9	–15.3	88.8	–10.8	5.0	95.1

and the population parameter) is less than 1%. Bayesian estimation yields accurate estimates of the within-level standard errors, while maximum likelihood estimation tends to slightly underestimate the standard errors (by 3.0% for the within factor loadings and 4.6% for the within structural effect). Taking everything together, the coverage of 95% confidence intervals at the within level is good for both estimators: The percentage of replications where the confidence interval includes the true population parameter is very close to the nominal 95% for all within parameters. ML estimation is more problematic at the between level. For the between structural effect, for example, the standard errors are underestimated by 15.3%, which is considerable. As a result, the 95% confidence intervals for this parameter only contain the true parameter in 88.8% of the replications, which implies an alpha of 0.112 rather than the nominal 0.05. As a result, there is a risk of finding pseudo-significant effects. The coverage is much more accurate when Bayesian estimation is used (despite some parameter bias for the structural effects).

7 Conclusion

The popularity of cross-national research has spurred a search for new analytical tools that are helpful in answering comparative research questions. This paper contributes to this search by discussing the advantages and pitfalls of a novel approach in the field of CNCR, namely multilevel structural equation modeling (MSEM). MSEM combines the strengths of classical multilevel regression (MLM) and structural equation modeling (SEM). On the one hand, the multilevel component of MSEM discerns between various levels of analysis, and therefore enables researchers to study the interplay between individual and country-level characteristics. On the other hand,

the SEM features make it possible to incorporate latent variables and estimate measurement models, to specify causal chains with direct and indirect effects, and to study group-level outcomes. As a result of this combination of features, MSEM is well equipped *par excellence* to estimate (parts of) Coleman's bathtub model—the theoretical template that underpins the causal relationships in many comparative studies.

Three models that are of particular relevance for CNCR were discussed in greater detail. Multilevel CFA models are used to assess the measurement of latent concepts in multilevel settings, and offer new opportunities to study the comparability of measurements, either across countries or across levels (isomorphism). Multilevel mediation models and models for group-level outcomes focus on the structural relationships between variables, and are indispensable to fully understand macro-micro and micro-macro processes. This contribution also identified what is arguably the most important obstacle to using MSEM in CNCR, namely sample size requirements at the highest level, and discussed possible remedies (such as Bayesian estimation).

To make MSEM accessible to comparative researchers, this contribution proposed a step-by-step strategy to estimate complex MSEMS, and provided examples of Mplus syntax to do so.

References

Andreß, Hans-Jürgen, Detlef Fetchenhauer and Heiner Meulemann. Eds. 2019. In *Cross-national comparative research — analytical strategies, results, and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00594-x.

Asparouhov, Tihomir, and Bengt Muthén. 2012. *General random effect latent variable modeling: Random subjects, items, contexts, and parameters*. Retrieved from https://www.statmodel.com/download/NCME12.pdf.

Baute, Sharon, Bart Meuleman, Koen Abts and Marc Swyngedouw. 2018. European integration as a threat to social security: Another source of Euroscepticism? *European Union Politics* 19:209–232.

Becker, Dominik, Wiebke Breustedt and Christina Isabel Zuber. 2018. Surpassing simple aggregation: Advanced strategies for analyzing contextual-level outcomes in multilevel models. *methods, data, analyses* 12:233–264.

Bennink, Margot, Marcel A. Croon and Jeroen K. Vermunt. 2013. Micro-macro multilevel analysis for discrete data: A latent variable approach and an application on personal network data. *Sociological Methods & Research* 42:431–457.

Billiet, Jaak, and Bart Meuleman. 2014. Some methodological challenges of quantitative cross-national social policy research. In *Invisible social security revisited: essays in honour of Jos Berghman*, ed. Wim van Oorschot W., Hans Peeters and Kees Boos, p. 289–303. Leuven: Lannoo Campus.

Blalock, Herbert M. 1979. The presidential address: Measurement and conceptualization problems: The major obstacle to integrating theory and research. *American Sociological Review* 44:881–894.

Bollen, Kenneth. 1989. *Structural equations with latent variables*. New York: Wiley.

Byrne, Barbare M., and Fons J.R. van de Vijver. 2014. Factorial structure of the family values scale from a multilevel-multicultural perspective. *International Journal of Testing* 14:168–192.

Cheong, JeeWon, and David A. MacKinnon. 2012. Mediation/Indirect effects in structural equation modeling. In *Handbook of Structural Equation Modeling*, ed. Rick H. Hoyle. New York: The Guilford Press.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt and René Algesheimer. 2019. How to obtain comparable measures for cross-national comparisons. In *Cross-national comparative research – analytical strategies,*

results and explanations. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00598-7.

Coleman, James 1990. *Foundations of Social Theory*. Cambridge: Harvard University Press

Croon, Marcel A., and Marc J. van Veldhoven. 2007. Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods* 12:45–57.

Davidov, Eldad, Hermann Dülmer, Elmar Schlüter, Peter Schmidt and Bart Meuleman. 2012. Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology* 43:558–575.

Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt and Jaak Billiet. 2014. Measurement equivalence in cross-national research. *Annual Review of Sociology* 40:55–75.

Davidov, Eldad, Peter Schmidt, Jaak Billiet and Bart Meuleman. 2018. *Cross-cultural Analysis: Methods and applications*. New York: Routledge.

Demarest, Leila, Arnim Langer and Bart Meuleman (forthcoming). *Between fallacy and feasibility? Theory and empirics in quantitative social disorder studies*.

Dyer, Naomi G., Paul J. Hanges and Rosalie J. Hall. 2005. Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The leadership quarterly* 16:149–167.

Fischer, Ronald. 2012. Value isomorphism in the European Social Survey: Exploration of meaning shifts in values across levels. *Journal of Cross-Cultural Psychology* 43:883–898.

Fischer, Ronald, C-Mélanie Vauclair, Johnny R. Fontaine and Shalom H. Schwartz. 2010. Are individual-level and country-level value structures different? Testing Hofstede's legacy with the Schwartz Value Survey. *Journal of cross-cultural psychology* 41:135–151.

Fontaine, Johnny. 2008. Traditional and multilevel approaches in cross-cultural research: An integration of methodological frameworks. *Multilevel analysis of individuals and cultures*, 65–92.

Fontaine, Johnny, and Ronald Fischer. 2010. Data analytic approaches for investigating isomorphism between the individual-level and the cultural-level internal structure. In *Cross-cultural research methods in psychology*. pp. 273–298. Cambridge University Press.

Harkness, Janet A., Michael Braun, M., Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith. 2010. *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken NJ: John Wiley & Sons.

Harkness Janet A., Fons J.R. van de Vijver and Peter Ph. Mohler. 2003. *Cross-Cultural Survey Methods*. New York: John Wiley.

Hedström, Peter, and Richard Swedberg. 1996. Social mechanisms. *Acta sociologica* 39:281–308.

Hedström, Peter, and Petri Ylikoski. 2010. Causal mechanisms in the social sciences. *Annual Review of Sociology* 36:49–67.

Hox, Joop J. 2010. Multilevel Analysis: Techniques and Applications. New York: Routledge.

Hox, Joop J., Rens van de Schoot and Suzette Matthijsse. 2012. How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods* 6:87–93.

Hu, Li-Tse, and Peter M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6:1–55.

Jak, Suzanne. 2014. Testing strong factorial invariance using three-level structural equation modeling. *Frontiers in Psychology* 5:745.

Jak, Suzanne, Frans J. Oort and Conor V. Dolan. 2014. Measurement bias in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal* 21:31–39.

Jak, Suzanne, Frans J. Oort, Frans J. and Conor V. Dolan. 2013. A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal* 20:265–282.

Jowell, Roger, Caroline Roberts, Rory Fitzgerald and Gillian Eva. 2007. *Measuring attitudes cross-nationally. Lessons from the European Social Survey*. London: Sage.

Judd, Charles M., and David A. Kenny. 1981. Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review* 5:602–619.

Kaplan, David, and Sarah Depaoli. 2012. Bayesian structural equation modeling. In *Handbook of Structural Equation Modeling*, ed. Rick H. Hoyle. New York: The Guilford Press.

Kittel, Bernhard. 2006. A crazy methodology? On the limits of macro-quantitative social science research. *International Sociology* 21:647–677.

Kroneberg, Clemens. 2019. Theory development in comparative social research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für*

Soziologie und Sozialpsychologie. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meule-mann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00604-y.

Krull, Jennifer L., and David P. MacKinnon. 1999. Multilevel mediation modeling in group-based inter-vention studies. *Evaluation Review* 23:418–444.

Marsh, Herbert W., Oliver Lüdtke, Alexander Robitzsch, Ulrich Trautwein, Tihomir Asparouhov, Bengt O. Muthén and Benjamin Nagengast. 2009. Doubly-latent models of school contextual effects: In-tegrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research* 44:764–802.

Mehta, Paras D., and Michael C. Neale. 2005. People are variables too: Multilevel structural equations modeling. *Psychological Methods* 10:259.

Meuleman, Bart, and Jaak Billiet. 2009. A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods* 3:45–58.

Meuleman, Bart, and Elmar Schlüter. 2018. Explaining cross-national measurement inequivalence. A Bayesian multilevel CFA with random loadings. In *Cross-Cultural Analysis: Methods and Appli-cations*, ed. Eldad Davidov, Peter Schmidt, Jaak Billiet and Bart Meuleman, p. 363-390. New York: Routledge.

Mills, Melinda, Gerhard G. Van de Bunt and Jeanne De Bruijn. 2006. Comparative research: Persistent problems and promising solutions. *International Sociology* 21:619–631.

Muthén, Bengt O. 1989. Latent variable modeling in heterogeneous populations. *Psychometrika* 54:557–585.

Muthén, Bengt O. 1991. Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement* 28:338–354.

Muthén, Bengt O. 1994. Multilevel covariance structure analysis. *Sociological Methods & Research* 22:376–398.

Muthén, Linda K., and Bengt O. Muthén. 2002. How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling* 9:599–620.

Muthén, Linda K., and Bengt O. Muthén. 1998–2017. *Mplus User's Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.

Preacher, Kristopher J., Michael J. Zyphur and Zhen Zhang. 2010. A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods* 15:209–233.

Rabe-Hesketh, Sophia, Anders Skrondal and Andrew Pickles. 2004. Generalized multilevel structural equation modeling. *Psychometrika* 69:167–190.

Rabe-Hesketh, Sophia, Anders Skrondal and Xiaohui Zheng. 2012. Multilevel structural equation mod-eling. In *Handbook of Structural Equation Modeling*, ed. Rick H. Hoyle. New York: The Guilford Press.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*. Sage.

Ruelens, Anna, Bart Meuleman and Ides Nicaise. 2018. Examining measurement isomorphism of multi-level constructs: The case of political trust. *Social Indicators Research* 140:907–927.

Saris, Willem E. 1998. The effects of measurement error in cross cultural research. In *Cross-cultural survey equivalence. ZUMA-Nachrichten Spezial 3*, ed. Janet Harkness. Mannheim: ZUMA.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel mod-els for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schlüter, Elmar, and Ulrich Wagner. 2008. Regional differences matter: Examining the dual influence of the regional size of the immigrant population on immigrant derogation in European contexts. *International Journal of Comparative Sociology* 49:153–173.

van de Schoot, Rens, David Kaplan, Jaap Denissen, Jens B. Asendorpf, Franz J. Neyer and Marcel A. Aken. 2014. A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development* 85:842–860.

Schwartz, Shalom H. 1994. Beyond individualism/collectivism: New cultural dimensions of values. In *Individualism and collectivism: Theory, method and applications*, ed. Uichol Kim, Harry C. Triandis, Cigdem Kagitcibasi, Sang-Chin Choi and Gene Yoon. London: Sage.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized latent variable modeling: Multilevel, lon-gitudinal, and structural equation models*. Boca Raton: Crc Press.

Steenkamp, Jan-Benedict E. M., and Hans Baumgartner. 1998. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research* 25:78–90.

Stegmueller, Daniel. 2013. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* 57:748–761.

Tay, Louis, Sang E. Woo and Jeroen K. Vermunt. 2014. A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organizational Research Methods* 17:77–106.

du Toit Stephen H., and Mathilda du Toit. 2008. Multilevel structural equation modeling. In *Handbook of Multilevel Analysis*, ed. Jan de Leeuw and Erik Meijer. Springer, New York, NY.

Vandenberg, Robert J., and Charles E. Lance. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods* 3:4–69.

van de Vijver, Fons J.R. 1998. Towards a theory of bias and equivalence. In *Cross-cultural survey equivalence. ZUMA-Nachrichten Spezial 3,* ed. Janet Harkness. Mannheim: ZUMA.

Van de Vijver, Fons J.R., and Kwok Leung. 1997. *Methods and data analysis for cross-cultural research.* London: Sage.

Zhang, Zhen, Michael J. Zyphur and Kristopher J. Preacher. 2009. Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods* 12:695–719.

Bart Meuleman 1980, PhD, Associate Professor and Head of Department at the Centre for Sociological Research (CeSO), University of Leuven, Belgium. Areas of research: Cross-national analysis of public opinion, welfare state attitudes, prejudice and ethnic threat, multilevel modeling, structural equation modeling. Publications: Cross-national Analysis: Methods and Applications. London 2017 (ed. with E. Davidov, P. Schmidt and J. Billiet); The Social Legitimacy of Targeted Welfare. Attitudes to Welfare Deservingness. Cheltenham 2017 (ed. with W. van Oorschot, F. Roosma and T. Reeskens).

**Kölner Zeitschrift
für Soziologie und
Sozialpsychologie**

ABHANDLUNGEN

Comparative Analyses of Housework and Its Relation to Paid Work: Institutional Contexts and Individual Agency

Daniela Grunow

Abstract Despite major changes in gender divisions of work since the 1960s, women continue to perform a larger share of unpaid housework and care than men, whereas men continue to perform more paid work. This is true for a wide range of countries. The paper first describes respective macro-trends for women's and men's changing contributions to paid work, routine housework and child care over the past 70 years. It then focuses on the role of institutional context and individual agency in gender divisions of routine housework according to cross-national comparative research published since 2000. On the macro level, the paper identifies three main areas of investigation: the role of work–family policies, welfare state regimes, and national levels of gender equality (Gender Empowerment Measure, the Gender Development Index and the Gender Inequality Index) for men's and women's divisions of work. On the micro level, studies mainly assess theories of economic dependency and resource bargaining, time availability, doing gender and deviance neutralization. More recently, research is turning to the examination of inter-relations between the micro- and macro-level factors. According to the state of research, women are better able to enact economic and noneconomic agency in national contexts with high levels of gender equality and supportive work–family policies. This is apparent in the Scandinavian countries.

D. Grunow (✉)
Faculty of Social Science, Sociology Institute, Chair of Sociology specializing in quantitative analyses of social change, Goethe-University Frankfurt am Main
Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany
E-Mail: grunow@soz.uni-frankfurt.de

🙂 Springer

Vergleichende Analyse von Hausarbeit im Verhältnis zu bezahlter Arbeit: institutionelle Kontexte und individuelle Agency

Zusammenfassung Obwohl sich die geschlechtsspezifische Arbeitsteilung seit den
1960er-Jahren gewandelt hat, verrichten Frauen noch immer einen weitaus größeren
Anteil an unbezahlter Hausarbeit als Männer, während Männer weiterhin mehr Er-
werbsarbeit verrichten. Dieser Befund gilt für ein breites Spektrum an Ländern. In
dem vorliegenden Artikel werden zunächst die zugrunde liegenden Makrotrends der
veränderten Beiträge von Frauen und Männern zu Erwerbsarbeit, Routinehaushalts-
tätigkeiten und Kinderbetreuung in den letzten 70 Jahren beschrieben. Danach wird
auf Basis der seit dem Jahr 2000 publizierten vergleichenden Forschungsergebnis-
se die Rolle institutioneller Kontexte und individueller Agency, d. h. individueller
Handlungsspielräume, bei der Verrichtung von Hausarbeit in den Blick genom-
men. Auf der Makroebene werden in diesem Artikel drei Hauptforschungslinien
zur Arbeitsteilung von Männern und Frauen identifiziert: die Rolle von Arbeits-
und Familienpolitik, von Wohlfahrtsstaaten und von Geschlechteregalität (Gender
Empowerment Measure, GEM; Gender Development Index, GII; und Gender Ine-
quality Index, GDI). Auf der Mikroebene werden die Rolle ökonomischer Abhängig-
keiten, ökonomische Verhandlungstheorien, zeitliche Verfügbarkeit, Doing Gender
und Devianzneutralisierung untersucht. Aktuell richtet sich die Forschung zudem
verstärkt auf Wechselwirkungen zwischen diesen Mikro- und Makrofaktoren. Der
Forschungsstand zeigt, dass Frauen ökonomische und nichtökonomische Formen
von Agency besser in nationalen Kontexten realisieren können, in denen ein hohes
Maß an Geschlechteregalität besteht und in denen es eine unterstützende Arbeits-
und Familienpolitik gibt. Beide Randbedingungen sind v. a. in den skandinavischen
Ländern zu finden.

1 Introduction

Sociologists consider the range of paid and unpaid work conducted in modern soci-
eties as part of "defining experiences of individual identity, group life, and, indeed,
societies as a whole" (Smith 2006, p. 676). As paid work is generally more highly
valued and rewarded than unpaid work, the ways in which different kinds of work
are organized within a society impact almost all other areas of sociological inquiry,
including persistence and change in social inequalities, gender relations and family
life (ibid.). Which tasks constitute paid or unpaid work in society, and which societal
realm is considered its main producer—the family, the state or the market—varies
cross-nationally and over time (Bianchi et al. 2012; Gershuny and Sullivan 2003).
Since the industrial revolution there has been a rising distinction between paid mar-

ket work and unpaid family work. Unpaid family work comprises housework and care, which have long been ascribed almost exclusively to women while paid work has been ascribed almost exclusively to men (Pinchbeck 2013; Tilly 1994; Tilly and Scott 1989). These ascriptions have been problematized and renegotiated, especially since the 1960s and 1970s, with educational expansion, the rise in female (and especially maternal) employment and feminist analyses of the foundations and drawbacks of gendered divisions of paid and unpaid work for women (Budig 2004). Whereas the economic literature unanimously considers unpaid work, in particular housework, unpleasant, tiring and a type of work both men and women try to avoid (Becker 1981; Ross 1987), a certain (Marxist) stream within sociology argues that women's extensive performance of unpaid work constitutes surplus labor which is used to demonstrate a families' affluence and high social status (summarized in Thompson and Armato 2012).[1] Empirical research shows that male and female singles below age 36 spend about similar amounts of time in unpaid work (including care) in Sweden, that gender differences are small in the US and France, and substantial in Italy (Anxo et al. 2011). In addition, men and in particular women with high-income do less unpaid work than their low-income peers (Gupta et al. 2015; Heisig 2011) and among dual earner families even those with higher social status frequently struggle and fight because their homes lack basic standards of cleanliness from time to time (Alsarve et al. in press; Ruppanner 2009). This suggests that the performance of unpaid housework is still an essential though burdensome part of everyday life in contemporary societies, which limits a person's time and energy for other, more rewarding, types of work.

Early 21st century women nevertheless continue to perform much of the unpaid work, even if men have increased their involvement in this sphere (Altintas and Sullivan 2016, 2017; Hook 2006; Kan et al. 2011; Sayer 2010).

A large and growing body of cross-national comparative research has investigated possible explanations for the lack of change in gendered divisions of work,[2] despite women's increased labor force attachment and earnings potential. The cross-national perspective is important because substantial differences in gender divisions of work remain, despite cross-national convergence in the time people spend on average in paid and unpaid work (Gershuny 2018).[3] To be sure, few studies in this field focus on the total work-hour differences (paid plus unpaid work) between women and

---

[1] This literature argues that "Conspicuous care for the home is a way of demonstrating that a family is affluent enough either to free a wife from having to work in the labor market so that she can devote time to surplus labor, or, if she does work in the labor market, to pay someone else to do the surplus labor for the family. In either case, the meticulous, upkeep of a home produced by women's surplus labor is a marker of class, race, and ethnic distinction (...)" (Thompson and Armato 2012, p. 80).

[2] In this paper, if not further specified, the term 'gender divisions of work' refers to how all types of work, paid and unpaid, are divided between men and women. This perspective includes studies of gender-change in either work sphere as well as changes among women and men. Work is defined as "sets of tasks that people carry out, often for a wage, to produce goods or services for others" (Smith 2006, p. 676). Unpaid work comprises a broad range of productive and reproductive tasks, including unpaid routine housework and unpaid child care (Lachance-Grzela and Bouchard 2010).

[3] Although confidence in comparative designs is far from being uncontroversial (for a summary of the critique see for instance Goerres et al. 2019), cross-national comparisons allow for testing empirical evidence and interpretations thereof across contexts, thus providing additional opportunities to assess micro-level

men in general (Gershuny 2018; Sullivan et al. 2018). These total differences are usually minor, given that the day has 24 h and scholars have argued that a focus on total working time symmetry masks gender inequalities resulting from differences in men's and women's shares of paid and unpaid work (for a discussion see Gershuny 2018; van der Lippe et al. 2011).

The paper first provides a cross-national comparative overview of the overall trends in gender divisions of paid work, unpaid housework and childcare. It then reviews in-depth the recent comparative research on men's and women's unpaid routine housework, which is usually the focus in studies of unpaid work. Narrowing the detailed literature review to the study of routine housework is necessary, given the aim to provide a state-of-the-art account of theories, methods, and substantive results of cross-national comparative research in this special issue. First, in spite of the fact that housework, care and paid work are arguably interrelated and should ideally be assessed jointly, research in each sphere constitutes largely separate bodies of literature, with only partly overlapping theoretical frameworks (Coltrane 2010; Lachance-Grzela and Bouchard 2010). Second, studying different types of work usually requires focusing on different populations (cp. Treas 2010): housework studies include all men and women, often with a focus on couples or spouses; studies on child care work primarily focus on parents, and studies on paid work usually focus on the working population. An in-depth review of comparative research on all types of work, including paid work and child care, is thus beyond the scope of this paper (for research in related fields see Erlinghagen 2019; Hank and Steinbach 2019). This review concludes by discussing how housework, child care and paid work develop over the life courses of men and women. We argue that the development of future cross-national empirical research and theory would benefit from complementing multicountry studies more systematically with (1) designs that apply a life course perspective to gender divisions of work and (2) micro-level longitudinal data.

2 Conceptual Considerations

The body of cross-national research investigating gender divisions of work is diverse. This concerns, for instance, the construction and measurement of dependent and independent variables, whether the main theoretical focus is on macro- or micro-level determinants of gender divisions of work, and according to the number of countries being compared. The different research designs and data types relating to these various aspects have repercussions for the empirical evidence created and are discussed in this section.

---

hypotheses (Kohn 1987). This aspect is salient for studying gendered divisions of labour, because this field of research has paid a lot of attention to micro-level determinants (see Sect. 4.2).

2.1 Absolute and Relative Measures of Paid and Unpaid Work

Time spent in different types of work can be measured in absolute or relative terms. Absolute measurement of time can be either derived from time diaries or from stylized survey questions on time use (Schulz and Grunow 2011). Whereas time diaries are considered to provide more accurate descriptive information on absolute time budgets (Bonke 2005), both types of data have been found to yield rather consistent findings for both, regularly performed activities, such as cooking or cleaning, and less frequently performed activities, such as family paperwork or repair work (Schulz and Grunow 2011).

Relative measurement can be constructed from absolute measurements (for example, by dividing women's absolute reported time for housework by men's absolute reported time for housework) or surveyed directly on a scale by asking who typically performs a certain type of work. Such scales can range from 'almost/completely the respondent' over 'split about 50/50' to 'almost/completely other adults in the household', usually the respondents' partner (ibid.). Absolute measurement of the time women and men spend on certain types of work provides more detailed information than relative measurements. In the area of housework studies, however, relative measurements are more frequently utilized, especially since the 1990s, when the refinement of micro-level theories thrived and scholars started adopting couple and household perspectives to gender divisions of work (Coltrane 2000 and see Table A1, column 9 in the Online Appendix). More recently, it has been argued that using women's absolute housework hours (as dependent variable) and absolute earnings (as independent variable) should be preferred over relative measurements, as the latter tend to obscure national variation in gender-class differences (Gupta et al. 2015). Other scholars have noted that the presentation of absolute measures for only one gender would be misleading or at least inappropriate (reviewed in Coltrane 2000). Comparative studies applying both absolute and relative measures of gender differences in types of work therefore provide a more differentiated picture of gender divisions of work than studies using only one of these measures.

2.2 Macro- and Micro-level Focus and Data

Cross-national research on gender divisions of work differs with respect to the main theoretical focus. Whereas virtually all comparative studies provide some discussion of national context, there are clear differences in the degrees to which studies theorize and operationalize the role of national context. Studies with a clear macro-level focus emphasize that national context shapes gender divisions of work, even when controlling for individual or household characteristics. Studies with a clear micro-level focus, in contrast, assess individual-/couple-level theories in different national contexts. Micro-level studies thus emphasize theoretical mechanisms related to respondents', respondents' partners and household characteristics. Please note that such characteristics may be strongly influenced by national context, and vice versa (cp. Coltrane 2010). The number of studies combining both perspectives has been rising over the past decades, with the availability of cross-national and micro-level data sets.

National aggregate data limits the assessment of gender divisions of work to macro-level factors. This limitation results from the fact that aggregate data are typically based on average measures of men's and women's working time. Average measures have been found to mask within-country heterogeneity, such as effects of social class, variation by household type and change over the life course (Baxter and Tai 2016; Cooke and Baxter 2010). Aggregate measures are nevertheless useful to describe general trends over time (time-series analysis) and in cross-national comparative perspective (Altintas and Sullivan 2016, 2017; Gershuny 2018; Sayer 2010; Sullivan et al. 2014; Sullivan and Gershuny 2016). Descriptions of average *relative* differences in types of work highlight gender differences within countries, but may obscure cross-national differences in absolute housework time (Sayer 2010). For example, in countries with higher ratios of paid work for both women and men (for example Sweden, Norway, Denmark), time spent in unpaid work tends to be shorter for both sexes, leading to smaller relative gender differences in divisions of housework than in countries with lower ratios of paid work (Sayer 2010).

Comparative micro-level data comprises of information on the absolute time, or share, individuals spend in different types of work in various household contexts. Examples of cross-national surveys collecting this type of information include the Generations and Gender Survey (GSS), the European Social Survey (ESS), and the International Social Survey Programme (ISSP). These data sets are usually available for shorter time spans (and fewer time points) than national aggregate data, but they allow for a more fine-graded analysis of micro-level theories, of compositional factors underlying observed macro-level differences, and of interrelations between macro- and micro-level factors. This latter type of analysis has been maturing over the past decades, but theoretical approaches that link macro- and micro-level determinants of gender divisions of paid and unpaid work have remained underdeveloped.[4]

2.3 Longitudinal Measures of Change in Gender Divisions of Work

The majority of comparative studies investigating long-term trends of change in gender divisions of work utilize time use data from the Multinational Time Use Study (MTUS). The MTUS contains harmonized national time diary data from more than 70 randomly sampled surveys. It covers time-use, including various types of paid and unpaid work in 30 countries from the 1960s onwards (Centre for Time Use Research 2018). Whereas this data provides very detailed micro-level information on the time men and women spend in different types of work, micro-level information on potential causes of time-use patterns is rather limited, compared to cross-national surveys. Recent studies use the MTUS to create national average aggregates, sometimes for particular subgroups, to describe change in gender divisions of work over time (Altintas and Sullivan 2016, 2017; Gershuny 2018; Sullivan et al. 2018). Other macro-level comparative studies use time series data provided by national statistical

---

[4] See for instance Gershuny (2018) and Sayer (2010) for a review of housework, and Cooke and Baxter (2010) for a review of gender divisions of paid and unpaid work in families.

offices, for example EU-SILC and OECD data (Cipollone et al. 2014; Thévenon 2013).

As spelled out in the previous section, several cross-national comparative surveys (GSS, ESS, ISSP) are by now available as repeated cross-sections (also referred to as regular surveys, see Hakim 2000). This type of data allows for assessing change within countries over time. In addition, it contains rich (cross-sectional) micro-level information.

Micro-level longitudinal studies, in comparison, provide much richer information on changes at the individual or household level (ibid.). Comparative micro-level longitudinal data that allow for a detailed assessment of macro- and micro-level determinants of gender divisions of paid and unpaid work are not yet available. This type of longitudinal research is thus based on comparative case designs, using ex-post harmonized national panel study data. For example, the British Household Panel Study (BHPS), the US Panel Study of Income Dynamics (PSID), the Household, Income and Labour Dynamics in Australia (HILDA), the Swedish Level of Living Survey (SNU) and the German Socio-Economic Panel study (GSOEP) are frequently used in such comparisons. Whereas changes in paid work are usually based on episode data of employment and periods of nonemployment, comparative studies on gender divisions of housework and child care are limited to comparisons of few points in time. Change in housework is registered when the respective survey waves are collected. To the best of our knowledge, there have only been two quantitative longitudinal comparative studies on housework which were able to follow the same couples over time (Evertsson and Nermo 2004; Gershuny et al. 2005).

2.4 Comparative Case Designs and Multicountry Studies

The literature on gender divisions of work includes both comparative case studies which compare a few countries in-depth, as well as multicountry studies which compare large numbers of countries by using country-level variables. The main difference between the two concerns the way national context enters the analysis. In comparative case studies, the selection of countries is usually theoretically motivated by drawing on welfare state theory or work–family policy regimes corresponding to a theoretical sampling strategy (for a summary see Grunow 2017). Differences between country contexts, including historical specificities, within-country variation, figurations of policies and gender ideologies, are usually discussed in-depth and lead to the formulation of country-specific hypotheses (Aisenbrey et al. 2009; Bühlmann et al. 2009). In multicountry comparisons, direct tests of country-level effects and of macro–micro interactions are possible, by using methods of multilevel modelling. In recent years, the number of studies applying multilevel models to gender divisions of paid and unpaid work has increased tremendously. Most frequently scholars use random slope multilevel models to assess cross-national variation in gender divisions of work. Random slope models allow the effects of the micro-level (level-one) explanatory variables to vary across countries. The assumption of varying micro-level effect sizes across countries is often informed by the theoretical frameworks applied. For example, several studies have tested hypotheses suggesting that the equalizing effects of individual gender ideologies and relative economic resources

on gender divisions of routine housework are stronger in more egalitarian countries than in less egalitarian countries (Fuwa 2004; Knudsen and Wærness 2008). Random intercept multilevel models reflect the assumption that effects of the level-one explanatory variables on the dependent variable are the same across countries, but that these variables' intercepts may vary between countries. Whereas model selection is thus important in this research, surprisingly few studies explicitly discuss this aspect. In addition, little attention is paid to national specificities or the selection of countries. The latter is usually predetermined by the comparative data set used (for a discussion of the limitations of these nonrandom samples see Schmidt-Catran et al. 2019).

The following section provides a brief overview of dominant macro-level theories and aggregate trends in gendered divisions of work since the 1960s: paid work, unpaid routine housework work and child care. Where available, trends in men's and women's absolute and relative time are reported and discussed. The range of micro-level theories and empirical studies focusing on micro–macro interactions as predictors of housework will be reviewed in Sect. 4.

3 Macro-level Focus on Gender Divisions of Work: Emphasizing the Institutional Context

Research employing a macro-level perspective emphasizes that contextual factors shape family arrangements and patterns of individual behavior, resulting in cross-national variation in gender divisions of paid and unpaid work (Baxter and Tai 2016). The contextual factors discussed most prominently in the literature include—but are not limited to—work–family policies, welfare state regimes and national levels of gender equality (Cooke and Baxter 2010; Grunow et al. 2018; Sayer 2010). Operationalizations of these concepts will be discussed in the following paragraphs.

Work–family policies provide legal rights and infrastructure to support women and men as workers and caregivers, for example paid and/or job protected care leaves, child care support and the right to work part time. In addition, the policy frameworks in place reflect varying normative ideals concerning gender divisions of work (Grunow et al. 2018). With the expansion of welfare states, the provision of these policies increased markedly across the European countries in the late 20th century, but with variation within and between types of welfare regimes (Cooke and Baxter 2010; Mandel and Semyonov 2006).

Welfare state regime typologies assessing gender divisions of paid and unpaid work developed extensively in the 1990s, often as a critique or refinement of Esping-Andersen's (1990) well-known welfare state typology (Gornick and Meyers 2003; Lewis 1993; Orloff 1993; Sainsbury 1994, 1996; for a general discussion of welfare regime theories see Schröder 2019). Critics emphasized the importance to acknowledge gendered forms of citizenship and equal rights policies (Orloff 1993, 2008; Pfau-Effinger 2005). These criteria have not lead to a regrouping of countries along the established welfare state categories, however. Neither has it become a standard in comparative studies on gender divisions of paid and unpaid work to control for national-level equal rights policies. Based on these typologies and following Hook

(2006), researchers expect the least egalitarian divisions of paid and unpaid work in Western and Southern European countries which share a history of conservative-corporatist policies (e. g. France, Germany, Italy, the Netherlands, Spain). The most egalitarian divisions are expected in the social democratic Scandinavian countries (e. g. Sweden, Denmark and Norway), due to their comparatively long and extensive development of gender equal work-care policies and high levels of state support for working families (Hook 2006). Liberal regimes, in contrast, emphasize market-based solutions to balance different types of work as well as gender-neutrality of policies (e. g. Great Britain, the United States and Australia). These conditions are associated with heterogeneous divisions of work and care among women and men. Aggregate outcomes in the liberal regimes have been argued to range in-between the outcomes in social democratic and conservative welfare states (Baxter and Tai 2016).

Recently, research highlighted distinct work–family policy changes in countries formerly known as prime examples of conservative welfare states, such as Germany and the Netherlands (Grunow and Veltkamp 2016; Grunow et al. 2018). Even though these countries kept conservative policy elements, they introduced rather strong incentives for more gender balanced divisions of paid and unpaid work, i. e. shared parental leaves (Germany and the Netherlands) and short(er), though better paid, employment breaks for parents on leave (Germany). These examples illustrate the need to look beyond welfare state typologies to understand cross-national variation in gender divisions of work.

At present, welfare regime typologies lack proper classifications for the postsocialist transformation states of Eastern Europe (e. g. Poland, the Czech Republic, Hungary). Eastern European countries have become recognized as a group sharing a common history of high female labor force participation under socialism but adopting rather different work–family policy strategies since (Grunow and Veltkamp 2016; Hofmeister et al. 2006; Mills et al. 2006). It has also been noted that communist societies did not challenge employed women's main responsibility for unpaid family work (Reimann 2016; Robila 2004). Taking this into consideration, one may expect comparatively high levels of gender symmetry in paid work, especially under communism, and comparatively low levels of gender symmetry in unpaid work in Eastern Europe.

Different but related concepts have been used to measure *national levels of gender equality*. The most prominent measures have been devised as part of the United Nations Development Programme, including the Gender Empowerment Measure (GEM), the Gender Development Index (GDI) and the Gender Inequality Index (GII). The GEM combines estimates of gender gaps in economic income, participation in high-paying and economically powerful positions, and political and professional representation (Pillarisetti and McGillivray 2002). The GDI was devised around the same time as the GEM but lacks measures of women's empowerment. The GII includes, in addition to labor market and empowerment indicators, measures of reproductive health. The GII thus combines measures of maternal mortality ratio and adolescent birth rates (health), proportions of parliamentary seats occupied by women and proportions of adult women and men aged 25 years and older with at least some secondary education (empowerment), and labor force participation rates
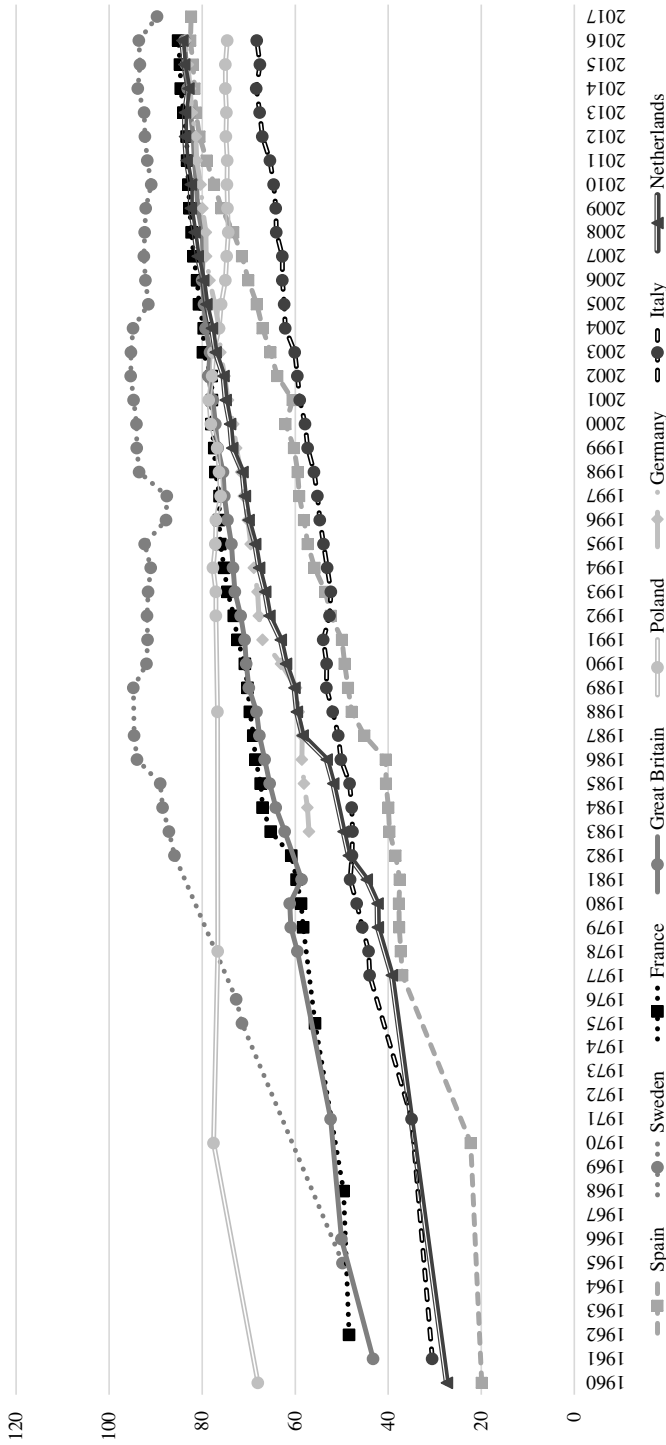
Fig. 1 Ratio of female-to-male labor force participation rates, 1960–2017. Source: World Development Indicators. http://datatopics.worldbank.org/world-development-indicators/, author's own depiction. Notes: Labor force participation rates represent the proportion of the population (age 15+) that is economically active. The ratio of female-to-male participation rates is calculated by dividing female labor force participation rate by male labor force participation rate and multiplying by 100

of female and male populations aged 15 years and older (labor market) (United Nations Development Programme 2016). According to recent cross-national GII-based rankings and with few exceptions, the Scandinavian countries score highest with respect to gender equality, followed by countries from western and southern Europe. The English-speaking liberal countries and Eastern Europe occupy much lower positions among the most highly developed countries (United Nations Development Programme 2018).

In addition, studies consider nationally dominant *gender ideologies*, sometimes also referred to as gender norms or gender culture (Pfau-Effinger 2005). These concepts capture widely shared ideas of meaning and reality about the skills and functions of women and men in society (Pfau-Effinger 2005, p. 4). Macro-level gender ideologies are frequently conceptualized in terms of aggregated individual gender attitudes. These measures consider "individuals' levels of support for a division of paid work and family responsibilities that is based on the belief in gendered separate spheres" (Davis and Greenstein 2009, p. 87). More recently, empirical research has identified newly emerging gender ideologies in western societies which cannot easily be located on the established egalitarian/traditional axis and are therefore considered multidimensional (summarized in Grunow et al. 2018). Examples include ideologies of intensive mothering (Hays 1996), intensive parenting (Wall 2010), and egalitarian essentialism (Charles and Grusky 2004; Cotter et al. 2011). The acknowledgement of multidimensional gender ideologies has not yet entered comparative analyses of gender divisions of work. Most studies employing a macro-level measure of gender ideology thus create a composite index of various items addressing gender attitudes along the egalitarian/traditional axis (for example Geist and Cohen 2011; Yodanis 2005), or use aggregated means of single items to reflect gender ideologies (for example Budig et al. 2012).

Across countries and welfare state regimes, absolute time in paid work has converged since the 1960s and now ranges "between 250 and 300 min per day" (Gershuny 2018, p. 2). Since the 1960s, women continue to perform rising shares of paid work while men's shares are declining (Gershuny 2018; based on MTUS data). On average, among 20–59 year olds, women now perform around 40% of all paid work and men perform around 60% (Gershuny 2018).

The findings from time-use data are mirrored in the ratio of female-to-male labor force participation rates for the population aged 15 or older (Fig. 1). Figure 1 shows the long-term development of the ratio of the female-to-male labor force participation rates (LFPR) for a selected set of European countries.[5] The ratio is calculated by first dividing the female LFPR by the male LFPR and then multiplying by 100. A value of 100 would mean that men's and women's labor force participation rates are similar.

In line with the expectations based on welfare regime arguments, the female-to-male LFPR ratio was as low as 20 in Spain and around 30 in the Netherlands and Italy in the 1960s. It was highest in Poland, with values of 68 and higher, whereas Great Britain, France and Sweden occupied intermediate positions with

[5] Countries have been selected to reflect the spectrum welfare state regimes discussed in the previous section, and according to the availability of time trend data.

values around 50. Until the mid-1980s, the female-to-male LFPR ratio increased steeply across countries, with the exception of Poland. Since then, the ratio has been highest in Sweden, where it fluctuated around 90 to 95%. In Poland, the ratio decreased slightly in the 2000s (from 78 in 2000 to 75 in 2016). In the other countries, the ratio increased further throughout the 2000s and 2010s. In 2016, cross-national variation in the female-to-male LFPR ratio was smaller than ever.[6] It ranged between 68 in Italy and 94 in Sweden, whereas values in the Netherlands, France, Germany, Great Britain and Spain were around 85. This pattern does not any longer reflect clear differences between the conservative, liberal and eastern European countries displayed.

Changes in the female-to-male LFPR ratio mostly result from changes in women's labor force participation (OECD employment data base). Female LFPRs increased markedly from around 40% in Germany, Great Britain and France in the 1960s and 1970s to over 55% (Germany and Great Britain), respectively slightly over 50% (France) in 2016 (ibid.). In the Netherlands, Italy and Spain, where participation rates were well below 30% in the 1960s, the increase in women's employment was even more pronounced. In Sweden, the female labor force participation rates surpassed 50% in the second half of the 1960s, continued rising steeply to a maximum of 71% around 1990 and since staggered slightly below 70% (ibid.). The OECD average of female LFPRs rose from slightly over 30% in the early 1960s to 52% in 2016 (ibid.).
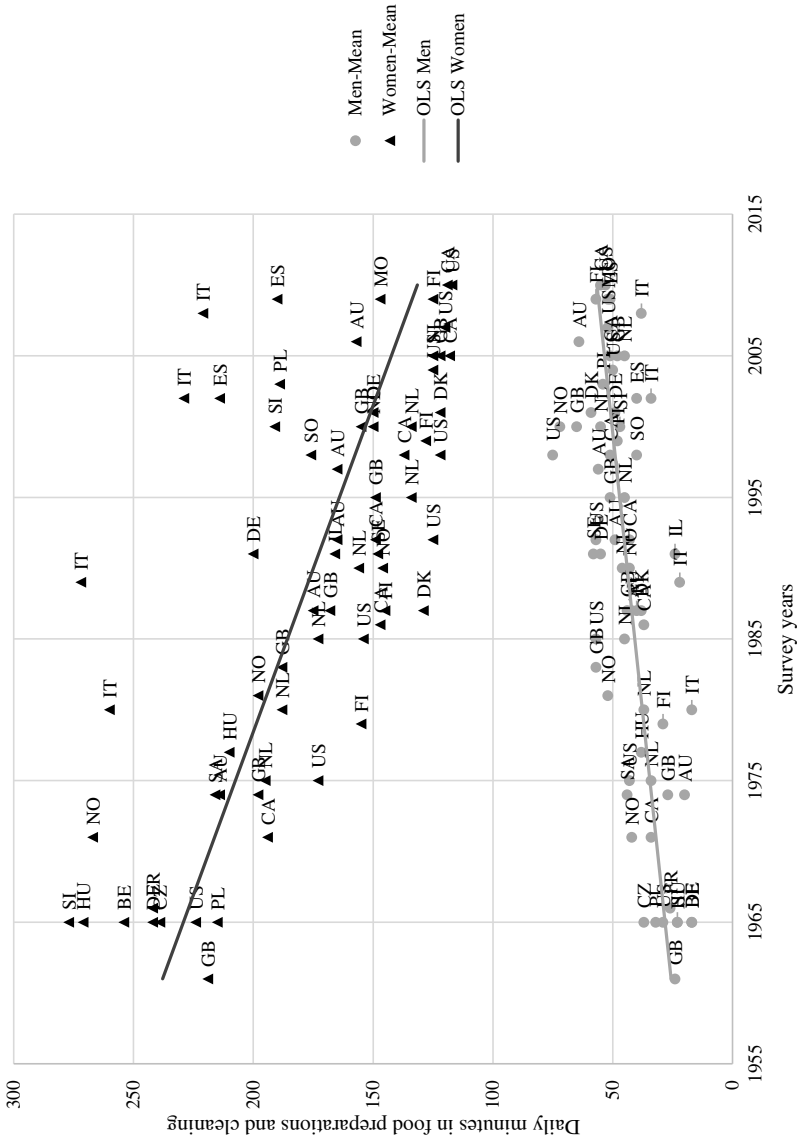
Much of the increase was due to mothers' continued employment and shorter work interruptions during family formation. A large share of the growth in female labor force participation rates thus reflects part-time work. Scholars investigating the forces behind these macro-level trends consider the spread of policies targeting maternal employment as main drivers of female employment rates in Europe and other OECD countries (Aisenbrey et al. 2009; Cipollone et al. 2014; Grunow et al. 2006; Thévenon 2013).

Concomitant with the absolute and relative increases in women's paid work time, their average time in routine housework decreased markedly across OECD countries (Fig. 2; see also Altintas and Sullivan 2016). In Germany and France, for instance, women spent more than four hours per day in routine housework in 1965 and less than three hours in the 2000s. Similar trends can be found in all developed economies, but cross-national variation of average time spent in routine housework remains high. According to time use data, routine housework hours are currently lowest in the Scandinavian countries (i.e. Denmark and Finland), the UK, the US and the Netherlands.

As more women spend time in paid work and less in routine housework, the question arises how they and their families compensate for this time in the unpaid sphere. To be sure, men's increases in routine housework are far from compensating for women's forgone routine housework time. German men, for instance, increased their average daily time in routine housework from 17 min in 1965 to 49 min in 2001 (see Fig. 2 and Altintas and Sullivan 2016, p. 457). Men in France increased their average daily minutes in routine housework from 26 min in 1966 to 52 min

---

6 Data for 2017 have been available only for Sweden (ratio of 90) and Spain (ratio of 82).

Fig. 2 Trends in gender divisions of routine housework, 1960s–2010s. Source: MTUS data. Notes: The depiction was replicated by the author from Altintas and Sullivan (2016). The figure refers to average minutes in routine housework for women and men age 19 or older. *OLS* refers to the fitted Ordinary Least Squares regression lines for men and women

in 2009. Whereas men more than doubled their time in routine housework across a broad range of countries, the gender gap in time spent in routine housework remained rather large. The gender gap in routine housework is currently lowest in the Scandinavian and the English-speaking liberal countries and highest in Eastern and Southern Europe (cp. Altintas and Sullivan 2016, p. 465).

So far, the time trends discussed covered changes in (working-age) men's and women's time use in general. In contrast, care work requires the focus to be on parents. A recent study by Altintas and Sullivan (2017) assessed change in co-resident fathers' time for childcare and routine housework, using the MTUS data. Across countries between the 1970s and 2010s, co-resident fathers of young children strongly increased their childcare time (Altintas and Sullivan 2017, p. 92). In the Nordic countries fathers' time for childcare almost doubled, from 45 to 82 min per day on average (ibid.). Fathers in Anglo-Saxon countries started out on a much lower level (with a mean value of 22 min) and meanwhile caught up with their northern peers (ibid). In the other countries, changes have been less pronounced. Here, the most recent figures (2000–2010) suggest that fathers spend on average around 70 min per day on childcare (ibid.). The trend that co-residential fathers spend more time with their children is counterbalanced on the macro-level by the fact that the number of children growing up with a co-resident father has been declining and more children are raised by single mothers (Coltrane 2010). Studies comparing mothers' and fathers' time for care document that mothers' time in childcare did not decrease as fathers' time in childcare increased (Neilson and Stanfors 2014; based on MTUS data). Instead, mothers' time in childcare increased as well, though not as strongly as fathers'. Overall, cross-national increases in parents' time with their children indicate a trend towards intensive parenting, which has also been documented in comparative studies of gender ideologies (Grunow et al. 2018; Knight and Brinton 2017).

Taken together, the findings presented suggest convergence in gender divisions of paid and unpaid work across countries. Variation across welfare regime types declined, although divisions of work continue to be most gender-balanced in the social-democratic cluster. This cluster also shows high levels of gender equality, according to national gender equality rankings. Countries usually grouped as conservative welfare regimes show the greatest heterogeneity. In part, this variation can be attributed to variation in work–family policies and levels of gender equality. Despite long-term trends of gender convergence, working age women continue to perform on average substantially more housework and childcare than working age men, whereas men currently perform on average more of all paid work than women. The remaining gender gap in divisions of paid and unpaid work has consequences for gender inequalities in earnings, human capital accumulation and further segregation of tasks within households over time (Gershuny 2018). Much research on gender divisions of work is thus driven by the aim to better understand why women continue to perform more housework (Baxter and Tai 2016). To answer this question, it is useful to broaden the macro-level perspective adopted so far and assess how micro-level forces shape gender divisions of housework. In particular, we discuss interrelations between both levels of analysis.

4 Interrelations of Macro- and Micro-level Determinants of Routine Housework

Time-use studies show that routine housework makes up a considerable fraction of all unpaid work (Gershuny 2018). Definitions of routine housework comprise traditionally *female* everyday household chores, including cooking, shopping for groceries, cleaning, caring for sick family members and laundry. Even though a focus on unpaid routine housework arguably tends to underestimate men's share of (and total time in) unpaid work, gender differences in routine housework are currently seen as a main driver of gender disparities in other spheres of life, including paid work and political representation, and thus as a reason for persistent social inequalities among men and women (Budig 2004; Treas and Drobnič 2010). First, the chores comprising routine housework are *low-schedule-control tasks* (Barnett and Shen 1997; Cunningham 2007) which limit a person's flexibility and availability for paid market work and other forms of social participation (Cunningham 2008). Second, performing large amounts of routine housework is associated with higher rates of depression (for a review see Mencarini and Sironi 2010). Third, time spent in routine housework has been argued to have detrimental effects on earnings, further career development and economic independence (summarized in Treas and Drobnič 2010). Unequal divisions of housework have thus been tied to various aspects of gender and class stratification (Cooke 2011). The divide in routine housework between men and women is thus consequential for gender inequality in other spheres of life.

4.1 Technical Summary of the Literature Reviewed

We review in this section comparative peer-reviewed articles and key comparative books on housework that have been published since 2000 (for a review of earlier works see Coltrane 2000). We limit our discussion to cross-nationally comparative peer-reviewed outlets published in English, which investigate men's and/or women's absolute or relative contributions to routine housework as dependent variable. Our literature search resulted in 49 publications, most of them articles, published between 2000 and autumn 2017.

As outlined in the conceptual section of this paper, the literature on routine housework can be grouped according to various design aspects of which Table A1 (Online Appendix) provides a technical overview (these studies' main empirical findings are summarized in Table 1). We emphasize the distinction between multicountry studies and comparative case designs, as the former enable multilevel modelling, which allows for testing macro-level effects, as well as macro–micro-level interactions (see Table 1 and Table A1 in the Online Appendix).

The housework literature is clearly dominated by comparisons of European and OECD countries and includes both comparative case studies as well as multicountry studies. The 16 comparative case studies reviewed for this paper mostly compared 2 to 3 countries, whereas the 32 multicountry studies compared between nine and 34 countries. The book edited by Treas and Drobnič contains both chapters employing comparative case designs and multicountry studies. Almost two thirds of

Table 1 Empirical findings on micro- and macro-level determinants for gender divisions of routine housework. Author's own work.

N°	Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants				Others	Significant tested interactions
		Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality		
Comparative Case Studies											
1	Bernhardt, Noack & Lyngstad	2008	✓	✓	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Gender × importance of work; Gender × importance of earning
2	Bianchi, Nazio, Lesnard & Raley	2014	N.t.	✓	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Union type c	N.t.
3	Bittman, England, Sayer, Folbre & Matheson	2003	g	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.
4	Cooke	2006	c g	N.t.	c g	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
5	Crompton, Brockmann & Lyonette	2005	N.t.	N.t.	✓	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
6	Evertsson & Nermo	2004	c	N.t.	c	c	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.
7	Gershuny, Bittman & Brice	2005	N.t.	g	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.
8	Gershuny & Sullivan	2003	c	–	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
9	Gimenez-Nadal, Molina & Ortega	2017	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Correlation between parents' housework and children's housework c	N.t.
10	Gupta, Evertsson, Grunow, Nermo & Sayer	2015	c	N.t.	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.

Table 1 (Continued)

N°		Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants					Significant tested interactions
			Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality	Others	
11	Hofäcker, Stoilova & Riebling	2013	g	c	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.
12	Lewin-Epstein, Stier & Braun	2006	c	–	✓	✓	N.t.	✓	N.t.	N.t.	Family model ✓	N.t.
13	Neilson & Stanfors	2014	c	✓	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Gender × age of youngest child in household
14	Sullivan	2010	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Educational level × survey data
15	Sullivan, Coltrane, McAnnally & Altintas	2009	N.t.	N.t.	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
16	Zabel & Heintz-Martin	2013	c	c	c	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Man's work status × woman's work status
Multicountry Studies												
17	Aassve, Fuochi & Mencarini	2014	c	N.t.	c g	✓	N.t.	c	N.t.	N.t.	N.t.	N.t.
18	Altintas & Sullivan	2017	✓	✓	N.t.	N.t.	N.t.	c	N.t.	c	N.t.	N.t.
19	Breen & Cooke	2005	N.t.	N.t.	✓	N.t.	N.t.	✓	N.t.	N.t.	N.t.	Proportion of autonomous women × proportion of adjuster men
20	Davis & Greenstein	2004	g	✓	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Gendered social structure c	N.t.

Table 1 (Continued)

N°		Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants					Significant tested interactions
			Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality	Others	
21	Dotti Sani	2014	c	c	c	N.t.	N.t.	N.t.	N.t.	c g	N.t.	N.t.
22	Fahlén	2016	✓	✓	✓	–	✓	N.t.	N.t.	✓	N.t.	N.t.
23	Geist	2005	g	g	c g	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
24	Heisig	2011	g	g	g	N.t.	N.t.	N.t.	N.t.	N.t.	Economic development and inequality ✓	N.t.
25	Kan, Sullivan & Gershuny	2011	N.t.	N.t.	✓	N.t.	✓	N.t.	N.t.	N.t.	N.t.	N.t.
26	Moreno-Colom	2017	N.t.	g	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
27	Nordenmark	2004	N.t.	✓	c g	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
28	Sullivan, Billari & Altintas	2014	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Fertility ✓	Very low fertility × time College × time
... thereof using Multilevel Analysis												
29	Altintas & Sullivan	2016	c	c	N.t.	N.t.	N.t.	c	N.t.	N.t.	N.t.	N.t.
30	Batalova & Cohen	2002	✓	✓	✓	N.t.	N.t.	N.t.	N.t.	c	Premarital cohabitation c	Gender empowerment × premarital cohabitation

Table 1 (Continued)

N°		Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants					Significant tested interactions
			Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality	Others	
31	Davis, Greenstein & Gerteisen Marks	2007	✓	✓	✓	N.t.	N.t.	N.t.	N.t.	g	Union type, g	Cohabiting × male Works full-time cohabiting × female works full-time Cohabiting × male earns more Cohabiting × female earns more Cohabiting × ideology
32	Fuwa	2004	c	c	c	N.t.	N.t.	✓	N.t.	✓	Economic development ✓	N.t.
33	Fuwa & Cohen	2007	c	✓	✓	N.t.	c	N.t.	N.t.	N.t.	N.t.	N.t.
34	Geist & Cohen	2011	g	c	c	N.t.	✓	N.t.	✓	✓	N.t.	N.t.
35	Hank & Juerges	2007	g	✓	✓	N.t.	N.t.	N.t.	N.t.	✓	N.t.	N.t.
36	Hook	2010	N.t.	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	Married women's and mothers' employment rate ✓	Hours of paid work × employed men's weekly paid hours
37	Iversen & Rosenbluth	2006	✓	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	Divorce rate ✓	N.t.

Table 1 (Continued)

N°		Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants					Significant tested interactions
			Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality	Others	
38	Knudsen & Wærness	2008	✓	✓	g	N.t.	N.t.	N.t.	N.t.	✓	Economic development; g	Female × gender ideology; Gender empowerment × relative income; Gender empowerment × gender ideology
39	Ruppanner	2010	✓	✓	✓	N.t.	N.t.	N.t.	N.t.	✓	N.t.	Gender empowerment men × gender empowerment women
40	Ruppanner & Treas	2015	N.t.	g	N.t.	N.t.	N.t.	N.t.	N.t.	N.t.	Female weekend employment; c	Female × respondent works weekends weekly; Female × spouse works weekends weekly; Female × spouse's work hours
41	Tai & Treas	2013	g	✓	N.t.	N.t.	N.t.	✓	N.t.	N.t.	N.t.	N.t.
42	Tamilina & Tamilina	2013	✓	✓	✓	N.t.	✓	✓	N.t.	N.t.	Poverty-related mechanism ✓	N.t.
43	Thébaud	2010	✓	N.t.	g	c	N.t.	N.t.	N.t.	✓	N.t.	World Culture Index × lower income
44	Treas & Tai	2016	g	✓	g	N.t.	N.t.	N.t.	N.t.	✓	N.t.	N.t.

Table 1 (Continued)

N°	Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants					Significant tested interactions
		Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality	Others	
45	2011	N.t.	N.t.	N.t.	N.t.	✓	c	N.t.	✓	Gender norms, historical changes	Child <age 5 × GNP
											Child 5–15 × GNP
											Married × GNP
											Middle-level education × GNP
											High-level education × GNP
										c	Child <age 5 × child-care expenses
											Child aged 5–15 × child-care expenses
											Married × masculinity
											Middle-level education × masculinity
											High-level education × masculinity
46	2017	g	✓	✓	✓	N.t.	N.t.	N.t.	N.t.	Unemployment	Gender × unemployed partner
										e g	Unemployment rate × unemployed
											Year (2010) × unemployed

Van der Lippe, de Ruijter, de Ruijter & Raub

Van der Lippe, Treas & Norbutas

Table 1 (Continued)

N°	Year of publication	Individual agency in micro-level theories				Institutional context & macro-level determinants					Significant tested interactions
		Economic dependency & resource-bargaining	Time availability	Doing gender & gender ideology	Deviance neutralization	Work–family policies	Welfare state regime	Gender ideologies	Gender equality	Others	
47	Voicu, Voicu & Strapcova	2009	g	✓	✓	N.t.	–	N.t.	N.t.	N.t.	Religious structure, technological development	Innovation × total housework
											✓	Innovation × women's working hours
48	Yodanis	2005	✓	N.t.	✓	N.t.	N.t.	N.t.	✓	N.t.	Divorce culture	N.t.
											✓	
Edited Book Containing Comparative Case Studies & Multicountry Studies												
49	Treas & Drobnič	2010	✓	✓	✓	✓	✓	c g	N.t.	✓	✓	N.t.

Overview of theories and findings in the studies reviewed for this paper. Full references as well as a technical appendix summarizing the data and methods used in each of the studies are provided in the online appendix. Several of the multicountry studies used theoretically relevant micro-level variables as controls, rather than systematic operationalization of theories. In these cases, the table allocates findings to the respective theories. ✓ confirmed and significant; – not confirmed; *c* partly confirmed, difference by country, significant; *g* partly confirmed, difference by gender, significant; *N.t.* not tested; *GNP* gross national product

the multicountry studies reviewed for this paper applied multilevel analysis. Time-use data are employed in both strands of research, whereas the case-design based studies employ more national time use data and multicountry comparisons draw mainly from the MTUS. Survey data are used in both strands of research as well, whereas the selection of countries and choice of data sets is more heterogeneous among the comparative case studies than the multicountry studies. The largest share of multicountry studies draws on ISSP data, followed by studies using the ESS and MTUS (see Table A1, column 6 in the Online Appendix). The ISSP collects estimates of both absolute and relative contributions to housework. The relative estimates are most frequently used in the studies reviewed here (see Table A1, column 9 in the Online Appendix). The ESS contains stylized estimates of respondent's time for housework as well as estimates for the respondent's partner. These data are frequently combined in the studies reviewed here to create relative measures of housework. The MTUS data contain absolute measures of time use, based on individual time-diaries. The MTUS-based studies tend to focus on absolute measures of housework, often investigating cross-national variation in time trends. As many of the multicountry studies use the same data, similar measures of housework and regression techniques, substantive findings should be rather well-comparable across studies.

4.2 Individual Agency in Micro-level Theories on Housework

There are many studies aiming to test micro-level theories of gendered divisions of housework in different countries, both in comparative case study designs (i.e. Bittman et al. 2003; Gupta et al. 2015) and in multicountry studies (i.e. Aassve et al. 2014; Davis and Greenstein 2004; Geist 2005; Knudsen and Wærness 2008). We argue that the concept of agency may be useful to understand the linkages between men's and women's unequal performance of housework and the internal and external constraints they are facing. Agency has been described as a person's fundamental relative autonomy in all circumstances (Hitlin and Elder 2007) and as embedded in specific institutional contexts (see also Emirbayer and Mische 1998; Mahmood 2001). The concept of agency thus describes men's and women's limited range of possible alternative actions and variations thereof in distinct contexts. It has been argued that the alternatives men and women perceive are conditioned by the individuals internalized own past experiences. To be sure, a person's gender defines or influences these experiences from birth onwards, most notably when it comes to family life (Eliot 2012; Grunow and Veltkamp 2016).

An important approach underlying the micro-level assessment of gender divisions of housework uses a concept of agency that links women's and men's focus on jointly maximizing the welfare of the household with the principle of economic rationality (Becker 1981). Following this line of reasoning, scholars have assessed whether women specialize in housework and men in paid work because men are the more productive earners and women the more productive homemakers. However, the number of studies falsifying this claim has been increasing, and more recently, research has turned to alternative explanations for gendered divisions of housework (see also Baxter and Tai 2016; Coltrane 2000). In the comparative lit-

erature published from 2000 onwards, economic rationality still plays an important role, but more attention is paid to the fact that performing housework is associated with economic dependency and that partners may have their own (economic and noneconomic) motives to do—or avoid doing—housework.

In the recent comparative literature four main theoretical mechanisms are addressed relating to individual agency in different ways (see Table 1 for an overview): First, economic dependency and resource-bargaining emphasize the existence of economic power relations in couples. These concepts suggest that the spouse with the greater earnings has more power to refrain from doing housework (reviewed in Gupta 2007). Second, the concept of time availability suggests that the partner who spends less time in paid work performs a greater share of housework because he/she has more time available (Barnett 1994; Presser 1994). This concept assumes a cooperative approach to gender divisions of housework, thus considering agency a pragmatic choice. Observed availability of time, however, may itself be a result of relative labor market resources and/or gender ideologies. Third, scholars focusing on concepts of doing gender (West and Zimmerman 1987) have argued that in contemporary societies, *femaleness* is still confirmed by doing housework and *maleness* by avoiding housework (Berk 1985). Doing gender thus explicitly relates to agency in terms of internalized gender ideologies and helps understand why some people (in particular women) may choose to perform more housework even though this may not be in their own best economic interest. Fourth, deviance neutralization has been discussed as a mechanism that operates through gendered links between relative resources and internalized gender ideologies. This mechanism suggests that women earning (or working) more than their male partners compensate for their nonnormative divisions of paid work by performing even more housework than their normative peers (Bittman et al. 2003; Greenstein 2000).

4.3 Empirical Assessment of Micro-level Theories on Housework

It is straightforward to expect cross-national variation in the empirical salience of the different micro-level theories. First, women's absolute and relative economic bargaining power should be higher (and dependency as well as time availability reduced) in countries where paid employment is more equally shared between women and men. Second, traditional gender ideologies that consider unpaid housework and care a predominantly female domain and paid work a predominantly male domain have declined considerably over the past decades (Knight and Brinton 2017). These shifts have led to more variation in gender ideologies, both within and between countries (Grunow et al. 2018; Knight and Brinton 2017).

Additionally, scholars have frequently pointed out that the micro-level theories on gender divisions of housework suggest dynamic developments of housework over the course of couple relationships whereas the vast majority of comparative studies use cross-sectional micro-level data which is unable to test micro-level sources of change over time (Baxter and Tai 2016; Grunow et al. 2012; Kühhirt 2012). This research runs the risk of confusing causes and consequences of divisions of work observed at a particular point in time. There is also a chance to falsely attribute effects that have been caused by couple-dynamics to macro-level country characteristics and

vice versa. Complementing this argument, Knudsen and Wærness (2008) claimed that the national context affected women's and men's total time in paid work and housework in the same direction, thus potentially blurring effects of national context on couples' divisions of work.

Taking both substantial and methodological arguments into account, it is perhaps not surprising that roughly between 20 and 30% of comparative micro-level cross-sectional studies found varying support by country for economic dependency, time availability and doing gender (Table 1). According to the studies employing multilevel analyses, roughly between 6 and 12% of variance in gender divisions of housework can be attributed to the country level, whereas variance explained on the country level tends to be smaller among men than women (Fuwa 2004; Fuwa and Cohen 2007; Hank and Jürges 2007; Thébaud 2010; van der Lippe et al. 2011).[7]

Economic dependency and resource bargaining theories have unconditionally been confirmed in one third of the comparative studies reviewed (Table 1). Close to one third found varying support for economic dependency and resource bargaining by country and another third by gender (cp. Table 1). Variation by gender indicates that if the female partner earns less than the male partner, his housework share is significantly lower than in couples earning approximately the same, but in the reverse case, the effect is not statistically significant. The operationalization of relative resources and dependency varied between studies. Some studies used relative income or relative occupational status (for example, Fahlén 2016; Ruppanner 2010). Others operationalized relative resources in terms of relative education (for example, Aassve et al. 2014; Lewin-Epstein et al. 2006). According to the study by Aassve et al. (2014), relative education mattered more in Western Europe and less in Eastern Europe. Multivariate models including measures of relative income in addition to relative education were supportive of economic dependency rather than relative resources (ibid.). Cross-national variation in support for economic dependency and bargaining mechanisms was reported more frequently in comparative case studies than in multicountry studies.

Support in favor of the *time availability* argument varied as well. Two of the comparative case studies, but none of the multicountry studies rejected the time availability argument. Time availability was measured rather consistently in terms of respondent's and partners' absolute or relative hours in paid work. More than half of the studies reported unconditional support for time availability. One out of five reported variation by country and a few studies reported variation by gender. Variation by gender means that women working part-time spent more time for housework than women working full-time whereas men working part-time spent less time for housework than their full-time working peers (for example Moreno-Colom 2017).

Studies seeking to address the *doing gender* theory controlled for respondent's gender ideologies thus separating, for example egalitarian-minded persons from persons with more traditional views on gender divisions of work. The doing gender hypothesis received unconditional support in close to 60% of the studies. A quarter

---

[7] Of course, variance components are further influenced by other aspects of study design, such as the selection of countries, micro-level sampling frames and construction of the dependent variable.

of the studies found variation in determining whose gender ideologies mattered for gender divisions of housework, women's or men's. Another quarter of the studies found support for doing gender in some countries but not in others and two studies reported variation by gender and country. Geist (2005), for instance, reported that women with egalitarian gender ideologies performed less housework in liberal and social-democratic regimes but not in conservative regimes, whereas men's gender ideologies consistently affected their shares of housework. In a similar vein, Nordenmark (2004) pointed to interaction effects between gender ideologies and country context (see also Cooke 2006). To summarize, both comparative case design studies and multicountry studies suggest cross-national variation in the ways gender ideologies and housework are related.

Deviance neutralization has been assessed less often in comparative research than other theories. Where assessed, the deviance neutralization hypothesis received empirical support in comparative case studies, in particular in countries that are usually classified as liberal welfare regimes (Bittman et al. 2003; Evertsson and Nermo 2004). According to multicountry studies, there is also support for deviance neutralization. Thébaud (2010) found support for resource bargaining in a sample restricted to partnered men, but reported that men's behavior was even more consistent with gender deviance neutralization in countries with traditional gender cultures. Aassve et al. (2014) investigated cross-national gender divisions of housework among coresident couples. Their findings stressed the importance of the gender ideologies held by women and men for divisions of housework and in addition supported the deviance neutralization hypothesis. Importantly, their study provided further support (but did not test) for possible interactions between gender ideologies and country context (ibid.).

To summarize we find rather stable cross-national support for the decisive role of doing gender, time availability and deviance neutralization on gendered divisions of housework. We find mixed cross-national evidence for resource bargaining and economic dependency and in addition marked variation by gender. Apparently, with respect to divisions of housework, men benefit more than women from having high relative economic resources.

4.4 Empirical Assessment of Macro–Micro Interactions

The fact that support for micro-level theories often varied between countries has fueled more systematic assessments of macro–micro interactions (Fahlén 2016; Geist and Cohen 2011; Ruppanner 2010; Tamilina and Tamilina 2014; Treas and Tai 2016). This shift in focus is further promoted by the spread of comparative data on country characteristics, such as work–family policies and gender equality measures, which can be merged with comparative micro-level data and applied in multilevel analyses.[8] Most frequently, studies of micro–macro interrelations included interaction terms

---

[8] A broad range of macro-level work–family policy indicators have become available which allow for assessing the impact of specific macro-level variables and work–family policies on gender divisions of labor (for example the OECD family data base; Multilinks Data base; various gender equality measures, developed the United Nation's Development Programme).

with national gender equality measures (see Table 1). Studies consistently reported that these macro-level indicators moderate divisions of housework beyond individual characteristics. In addition, studies continued to draw on welfare state typologies, either to group the countries assessed in respective dummy variables (for example Altintas and Sullivan 2017), or to motivate country selection and country-specific hypotheses in comparative case studies (Gupta et al. 2015).[9] Empirical support for the relevance of welfare regime differences is rather high and consistent between studies (see Table 1). Overall gender equality (i.e. measured in terms of greater aggregate equality in employment participation, political representation or GEM) has been found to increase egalitarian divisions of housework, though these effects have been rather small (Davis and Greenstein 2004; Fuwa 2004; Hook 2006; Ruppanner 2010). Significant interaction effects have been reported between GEM, women's gender ideology and time availability on men's share of housework (Fuwa 2004). Hank and Jürges (2007) reported a similar association between GEM and divisions of housework in their study of older couples. This research suggests that women's agency (be it economic or ideological) is more pronounced for those living in more egalitarian countries (see also Cooke 2006).

The salience of gender ideologies and women's economic bargaining power tend to be lower in conservative welfare states and in countries with low levels of gender equality. In the Scandinavian countries, which also score high on national gender equality measures, women's capacity to negotiate their share of housework is highest (see also Cooke and Baxter 2010).

By and large, these findings are consistently reported in comparative case designs and in multicountry studies. Whereas only multicountry studies employing multi-level analysis are able to provide rigorous tests of these micro–macro associations, longitudinal comparative case studies are needed to assess these processes from a life course perspective.

5 Gender Divisions of Paid and Unpaid Work—Evidence from a Life Course Perspective

The time women and men devote to paid and unpaid work varies over the life course. More scholars thus argue that applying a life course perspective to remaining gender differences in gender divisions of work is indispensable to investigating the inter-relations of macro- and micro-level mechanisms that contribute to the gendering of paid and unpaid work over time (Baxter et al. 2014; Baxter and Tai 2016; Coltrane 2010; Nitsche and Grunow 2016). The life course perspective highlights, first, the ways in which national institutions, such as educational and occupational systems, labor markets, and work–family policies structure gendered variation in paid and unpaid work over time (Brückner and Mayer 2005). This aspect is of particular importance for cross-national comparisons. Second, the life course perspective draws attention to path dependencies, turning points, and the role of linked lives in shaping

---

[9] Other macro-level factors observed include, for example, indicators of economic growth, economic inequality and divorce rates (Batalova and Cohen 2002; Heisig 2011).

men's and women's agency in gender divisions of work (Elder 1998).[10] It is thus not surprising that time for paid and unpaid types of work varies as individuals leave the parental home, start and end relationships, move in together, and have children. Across a wide range of countries, both sexes now have on average higher and more similar skills when they form a couple and move in together because education, school-to-work transitions, cohabitation and early employment patterns decreasingly vary by gender (Blossfeld et al. 2005, 2015; Nazio 2008). Available evidence suggests that women and men share paid and unpaid work more equally during the early stages of the life course than aggregate national averages would suggest (Blossfeld et al. 2005; Bühlmann et al. 2009). However, whereas working age men's time for paid and unpaid work remains rather stable over the life course, regardless of family transitions, women's time-use is strongly affected by family-related turning points, most importantly the birth of children (Baxter et al. 2014; Baxter and Tai 2016; Kühhirt 2012; Schober 2013). As a result, couples' divisions of paid and unpaid work become more unequal during family formation, and they often remain unequal, due to path dependencies in partners' career advancement, pay and the newly established routines in divisions of housework and care. In many countries, motherhood comes with earnings penalties, while fathers experience earnings premiums (Boeckmann and Budig 2013; Budig et al. 2012). These differences mostly apply in countries lacking public childcare for infants and toddlers and related to this, in couples with stay-at-home mothers (Boeckmann and Budig 2013; Budig et al. 2012).

Cross-national research employing a life course perspective to the transition to parenthood is still rare, but available qualitative comparative studies suggest that prevalent parenthood ideologies are powerful forces that may lead new mothers to prioritize unpaid housework and care over their equality claims and career ambitions, whereas most new fathers start perceiving themselves as main earners (Evertsson and Grunow 2016). Transitions to parenthood thus affect individual agency in gendered ways, and new parents challenging dominant parenthood ideologies report feeling penalized by relevant others, for instance relatives, colleagues or bosses (ibid.). According to this study, parenthood ideologies and related behaviors vary across work–family policy contexts, with couples in Sweden, a prime example of a social democratic welfare state, planning the most egalitarian work-care arrangements and couples in the Czech Republic, a postsocialist transition country promoting extensive maternity leave, the least egalitarian divisions (ibid.).

The claim that change and persistence of gendered divisions of work at the transition to parenthood are moderated by work–family policies is further backed up by comparative life course research using cross-sectional data (Bühlmann et al.

---

[10] Path dependencies concern established routines in everyday life as well as trajectories, such as career paths, which are unlikely to change, unless they are disrupted by biographical turning points (Nitsche and Grunow 2016). Turning points that may impact gender divisions of paid and unpaid work include the birth of children, couples' separation, illness or job loss. The notion of linked lives emphasizes that individual life courses are tied to the life courses of other people, most importantly that of partners and children (Moen 2003).

2009).[11] For couples living in social democratic regimes, shifts in gender divisions of paid and unpaid work are rarer and less persistent, whereas couples living in liberal regimes tend to accommodate to unequal divisions of work. Couples in conservative regimes experience a similar, though less pronounced traditionalization than their peers in liberal regimes (ibid.). Eastern European couples tend to return to prebirth employment symmetry while housework remains gendered. The findings are in line with conclusions drawn from quantitative longitudinal research in single countries (Baxter et al. 2014; Grunow et al. 2012; Kühhirt 2012). Taken together, these findings complement the theoretical arguments and findings reviewed in the previous sections by pointing to household dynamics of paid and unpaid work that are linked to both economic and noneconomic forms of individual agency and moderated by the national context.

6 Methodological Implications for Cross-sectional Research

The findings reviewed in the previous section illustrate the importance of theorizing and operationalizing gender divisions of labor as processes that change over the individual lifetime, not only in comparative perspective. Even though longitudinal analyses of paid employment have meanwhile become standard in comparative research, most cross-national studies of routine housework and child care are based on cross-sectional data or repeated cross-sections. The potential mechanisms at work would therefore have to be stable over the life course to allow for correct analyses of the determinants of gendered divisions of work (compare Baxter and Tai 2016). As a consequence, cross-sectional variables, such as (relative or individual) earnings and (relative or individual) employment hours have frequently been shown to be endogenous in cross-sectional research. Earnings and work hours not measured at an earlier point in time may have affected divisions of unpaid work observed later and subsequently earlier divisions of unpaid work may have affected earnings (Bloemen and Stancanelli 2014; Carlson and Lynch 2017).

Cross-sectional data also create problems for the correct assessment of the impact of macro-level variables on gender divisions of work, for instance specific work–family policies. First, work–family policies in place at one point in time target subgroups of the populations studied (for instance, parents of newborns), not parents per se or the working population as a whole. Second, policy changes may affect the behavior of target groups with delay and it may take even longer until the wider population adopts their attitudes and behaviors (Lachance-Grzela and Bouchard 2010). Cross-sectional data, though clearly dominant in comparative research, are unsuited to correctly assess these dynamics. In the absence of longitudinal data, these limitations need to be taken into account when designing comparative research and interpreting the results.

---

[11] Bühlmann et al. (2009) base their analysis on cross-sectional ESS data, but they use a life course approach to construct comparison groups reflecting different biographical stages and to inform their hypotheses.

7 Summary and Discussion

This paper provided a comparative overview of changes in gendered divisions of paid and unpaid work in Western societies throughout the late 20th and early 21st century. Starting from the notion that the performance of paid and unpaid work constitutes the basis for individual identity, interaction, social participation and inequality in the Western world, we investigated how work–family policies, welfare state regimes and national levels of gender equality affected change in men's and women's divisions of work.

We first discuss cross-national variation observed in the macro-level trends. In the 1960s and 1970s, cross-national differences in gender divisions of paid and unpaid work used to correspond rather closely to the expectations derived from welfare state theory. On average, the most gender balanced divisions of paid and unpaid work could be found in the social-democratic countries, the least gender balanced divisions in the conservative countries of southern Europe. Over time and concomitant with the expansion of welfare states, the spread of work–family supportive policies and women's more equal political and economic representation in western societies, country differences became smaller and within-regime variation more pronounced. Over the same period, men more than doubled their time for routine housework and co-resident fathers more than doubled the time spent with their children. Still, the literature consistently supports the conclusion that gender convergence in the time spent on paid and unpaid work was mostly driven by the changing employment behavior of women and that the increase in women's employment was at least in part a reaction to policy changes addressing working mothers (Cipollone et al. 2014; Cooke and Baxter 2010; Thévenon 2013).

Despite long-term trends of gender convergence across a wide range of countries in paid and unpaid types of work, working age women continue to perform substantially more housework and childcare than working age men and men perform more paid work than women. Since paid work provides direct access to material resources and is more highly valued in contemporary societies than unpaid work, the resulting gap in divisions of paid and unpaid work has consequences for gender inequalities in earnings, human capital accumulation and further segregation of tasks within households over time, to the disadvantage of women (Gershuny 2018). Most of the research reviewed for this overview article therefore aimed to better understand why women continue to perform on average more housework and under which conditions men increase their housework share (Baxter and Tai 2016; Lachance-Grzela and Bouchard 2010).

The review of comparative studies assessing micro-level theories and potential mechanisms associated with gendered divisions of housework yielded quite consistent cross-national support for the doing gender theory, operationalized in terms of individual gender ideology measures. Women holding egalitarian gender ideologies tend to perform less housework and men holding egalitarian ideologies perform more housework than their less egalitarian peers. Studies are more ambiguous concerning whose ideologies matter more, i.e. women's or men's, an individual's or their partner's. Multilevel models provided additional insights into macro–micro interrelations between individual gender ideologies and country level traits, most

importantly welfare state regimes, work–family policies and women's economic and political representation. In conservative welfare states, specifically in countries in which women's economic and political representation is lower, masculine culture prominent and minimal work–family supportive policies exist, gender divisions of housework were affected less by individual gender ideologies. In the Scandinavian countries, where these macro-level aspects support more egalitarian divisions of work and care, gender ideologies mattered more.

Studies reported mixed evidence, including variation between countries and by gender concerning the resource bargaining and economic dependency mechanisms, operationalized in terms of partners' relative earnings, educational levels or occupation. Considering variation by gender, we conclude that when it comes to arranging housework within couples, men more often than women use their higher relative economic resources to negotiate lower shares of housework. In principle, this gendered finding can be interpreted as evidence pointing to deviance neutralization. The deviance neutralization mechanism suggests that women perform more—and men less—housework to compensate for violating of traditional gender ideologies in cases where the woman earns more than her partner. Such conclusions about the gendered meaning of relative resources and economic dependency are further substantiated by studies using multilevel models. Respective analyses have shown that in countries supporting gender equality and maternal employment through work–family policies women's economic bargaining power on the micro-level was higher, and corresponded better with the theoretically expected patterns.

This evidence was backed up and complemented by the small but growing body of comparative life course research. Life course research adopting a comparative approach is still rare but available studies consistently show that men's and women's agency changes over the life course, affecting strategic economic motives, as well as internalized gender ideologies (Evertsson and Grunow 2016). Family formation thus constitutes a turning point in couples' divisions of work and care that leads to changes in gender divisions of paid and unpaid work. These changes tend to become path-dependent; an insight that helps explaining why on average across countries, women still perform more unpaid work than men and men perform more paid work. Work–family policies were found to moderate these dynamics, indicating the potential of policies to influence gender divisions of work over the life course (Bühlmann et al. 2009).

Taken together, we can conclude that women's agency to perform or avoid performing housework, whether driven by strategic economic considerations, by internalized gender ideologies, or by a mixture of both, is stronger in countries supporting higher levels of gender equality. Men are on average better able to enact economic agency, and avoid housework, in contexts favoring traditionally masculine culture and male breadwinning, while their support for gender balanced policies and practices depends on their social surroundings (Thébaud 2010; Thébaud and Pedulla 2016).

Whereas the body of research assessing the gender-division of work included both comparative case designs and multicountry studies, the latter provided explicit statistical tests over the past two decades of which aspects of country context impact on divisions of unpaid work at home. In particular, the growing use of multilevel

analyses provided new insights of how micro-level patterns of men's and women's performance or avoidance of housework were mediated by national levels of gender equality, economic inequality, and work–family policies (i. e. Dotti Sani 2014; Fahlén 2016; Knudsen and Wærness 2008; Heisig 2011).

What difference does the national context make, compared to individual factors and are context effects practically significant in terms of effect sizes? Technically, this question can be answered in a straight-forward manner by looking at the studies that have employed multilevel analyses. The studies reviewed in this paper report that roughly between 6 and 12% of variance in gender divisions of housework can be attributed to the country level and that the variance explained on the country level is higher for women than men (Fuwa 2004; Fuwa and Cohen 2007; Hank and Jürges 2007; Thébaud 2010; van der Lippe et al. 2011).

More importantly, however, scholars investigating social inequalities associated with gender divisions of labor need to consider that national context enters the analyses not only in terms of macro-level variables, it also shapes the gender inequalities which are measured on the micro level, affecting both, the dependent and independent variables considered (in a similar vein Coltrane 2010). An example would be new mothers' economic bargaining power. In countries supporting maternal employment through work–family policies, mothers will have more economic resources and thereby more bargaining power relative to their partners than in countries unsupportive of maternal employment. Work–family policies have also been argued to change gender ideologies on both the micro and macro levels (Gangl and Ziefle 2015; Grunow et al. 2018). Related to this, national context has an impact on how homogenous or heterogeneous men's and women's economic resources or gender ideologies are. If there is high variation within a country in these aspects, the share of the variance explained on the country level will be lower, and vice versa. A substantial interpretation of high and low country-level variance needs to draw on theory and substantial arguments, not only on statistical significance and effect sizes.

Earlier reviews of studies investigating gender divisions of paid and unpaid work concluded that the theorizing of macro- and micro-level interdependencies has remained surprisingly weak, given the great empirical advances that have been made in the field (Coltrane 2010; Cooke and Baxter 2010). The more recent publications reviewed here do not suggest a broad theoretical turn either. The review concludes with a call for integrating the theoretical and substantial insights gained by micro-level longitudinal research more systematically into the design and interpretation of multilevel models of gender divisions of paid and unpaid work. As pointed out in the previous section, adopting a life course perspective may go a long way, even if applied to cross-sectional data.

References

Aassve, Arnstein, Giulia Fuochi, and Letizia Mencarini. 2014. Desperate housework: relative resources, time availability, economic dependency, and gender ideology across Europe. *Journal of Family Issues* 35:1000–1022.

Aisenbrey, Silke, Marie Evertsson, and Daniela Grunow. 2009. Is there a career penalty for mothers' time out? A comparison of Germany, Sweden and the United States. *Social Forces* 88:573–605.

Alsarve, Jenny, Katarina Boye, and Christine Roman. 2019. Realized plans or revised dreams? Swedish parents' experiences of care, parental leave and paid work after childbirth. In *New parents in Europe: Work-care practices, Gender norms and Family Policies*, ed. Daniela Grunow, Marie Evertsson. Cheltenham: Edward Elgar. Forthcoming.

Altintas, Evrim, and Oriel Sullivan. 2016. 50 years of change updated: cross-national gender convergence in housework. *Demographic Research* 35:455–470.

Altintas, Evrim, and Oriel Sullivan. 2017. Trends in fathers' contribution to housework and childcare under different welfare policy regimes. *Social Politics* 24:81–108.

Anxo, Dominique, Letizia Mencarini, Ariane Pailhé, Anne Solaz, Maria Letizia Tanturri, and Lennart Flood. 2011. Gender differences in time use over the life course in France, Italy, Sweden, and the US. *Feminist Economics* 17(3):159–195.

Barnett, Rosalind C. 1994. Home-to-work spillover revisited: a study of full-time employed women in dual-earner couples. *Journal of Marriage and the Family* 56:647–656.

Barnett, Rosalind C., and Yu-Chu Shen. 1997. Gender, high- and low-schedule-control housework tasks, and psychological distress. A study of dual-earner couples. *Journal of Family Issues* 18:403–428.

Batalova, Jeanne A., and Philip N. Cohen. 2002. Premarital cohabitation and housework: couples in cross-national perspective. *Journal of Marriage and Family* 64:743–755.

Baxter, Janeen, and Tsui Tai. 2016. Inequalities in unpaid work: a cross-national comparison. In *Handbook on well-being of working women. International handbooks of quality-of-life*, ed. Mary L. Connerley, Jiyun Wu, 653–671. Dordrecht: Springer.

Baxter, Janeen, Sandra Buchler, Francisco Perales, and Mark Western. 2014. A life-changing event: first births and men's and women's attitudes to mothering and gender divisions of labor. *Social Forces* 93:989–1014.

Becker, Gary S. 1981. *A treatise on the family*. Cambridge, London: Harvard University Press.

Berk, Sarah F. 1985. *The gender factory: the apportionment of work in American households*. New York: Plenum Press.

Bianchi, Suzanne M., Liana C. Sayer, Melissa A. Milkie, and John P. Robinson. 2012. Housework: Who did, does or will do it, and how much does it matter? *Social Forces* 91:55–63.

Bittman, Michael, Paula England, Liana C. Sayer, Nancy Folbre, and George Matheson. 2003. When does gender trump money? Bargaining and time in household work. *American Journal of Sociology* 109:186–214.

Bloemen, Hans G., and Elena G.F. Stancanelli. 2014. Market hours, household work, childcare, and wage rates of partners: an empirical analysis. *Review of Economics of the Household* 12:51–81.

Blossfeld, Hans-Peter, Erik Klijzing, Melinda Mills, and Karin Kurz. 2005. *Globalization, uncertainty and youth in society: the losers in a globalizing world*. London: Routledge.

Blossfeld, Hans-Peter, Jan Skopek, Moris Triventi, and Sandra Buchholz. 2015. *Gender, education and employment: an international comparison of school-to-work transitions*. Cheltenham: Edward Elgar.

Boeckmann, Irene, and Michelle Budig. 2013. Fatherhood, intra-household employment dynamics, and men's earnings in a cross-national perspective (No. 592). LIS Working Paper Series. http://hdl.handle.net/10419/95618. Accessed 8 June 2018.

Bonke, Jens. 2005. Paid work and unpaid work: diary information versus questionnaire information. *Social Indicators Research* 70:349–368.

Brückner, Hannah, and Karl Ulrich Mayer. 2005. De-standardization of the life course: what it might mean? And if it means anything, whether it actually took place? *Advances in Life Course Research* 9:27–53.

Budig, Michelle J. 2004. Feminism and the family. In *The Blackwell companion to the sociology of families*, ed. Jacqueline Scott, Judith Treas, and Martin Richards, 416–434. Oxford: Blackwell.

Budig, Michelle J., Joya Misra, and Irene Boeckmann. 2012. The motherhood penalty in cross-national perspective: the importance of work–family policies and cultural attitudes. *Social Politics* 19:163–193.

Bühlmann, Felix, Guy Elcheroth, and Manuel Tettamanti. 2009. The division of labour among European couples: the effects of life course and welfare policy on value–practice configurations. *European Sociological Review* 26:49–66.

Carlson, Daniel L., and Jamie L. Lynch. 2017. Purchases, penalties, and power: the relationship between earnings and housework. *Journal of Marriage and Family* 79:199–224.

Centre for Time Use Research. 2018. Multinational time use study. https://www.timeuse.org/mtus. Accessed 9 June 2018.

Charles, Maria, and David B. Grusky. 2004. *Occupational ghettos: the worldwide segregation of women and men*. Stanford: Stanford University Press.

Cipollone, Angela, Eleonora Patacchini, and Giovanna Vallanti. 2014. Female labour market participation in Europe: novel evidence on trends and shaping factors. *IZA Journal of European Labor Studies* 3:18.

Coltrane, Scott. 2000. Research on household labor: Modeling and measuring the social embeddedness of routine family work. *Journal of Marriage and Family* 62:1208–1233.

Coltrane, Scott. 2010. Gender theory and household labor. *Sex Roles* 63:791–800.

Cooke, Lynn P. 2006. Policy, preferences, and patriarchy: the division of domestic labor in east Germany, west Germany, and the United States. *Social Politics: International Studies in Gender, State and Society* 13:117–143.

Cooke, Lynn P. 2011. *Gender-class equality in political economies*. Abingdon: Routledge.

Cooke, Lynn P., and Janeen Baxter. 2010. "Families" in international context: Comparing institutional effects across western societies. *Journal of Marriage and Family* 72:516–536.

Cotter, David, Joan M. Hermsen, and Reeve Vanneman. 2011. The end of the gender revolution? Gender role attitudes from 1977 to 2008. *American Journal of Sociology* 117:259–289.

Cunningham, Mick. 2007. Influences of women's employment on the gendered division of household labor over the life course: evidence from a 31-year panel study. *Journal of Family Issues* 28:422–444.

Cunningham, Mick. 2008. Influences of gender ideology and housework allocation on women's employment over the life course. *Social Science Research* 37:254–267.

Davis, Shannon N., and Theodore N. Greenstein. 2004. Cross-national variations in the division of household labor. *Journal of Marriage and Family* 66:1260–1271.

Davis, Shannon N., and Theodore N. Greenstein. 2009. Gender ideology: Components, predictors, and consequences. *Annual Review of Sociology* 35:87–105.

Dotti Sani, Giulia Maria. 2014. Men's employment hours and time on domestic chores in European countries. *Journal of Family Issues* 35:1023–1047.

Elder, Glen H. 1998. The life course as developmental theory. *Child Development* 69:1–12.

Eliot, Lise. 2012. *Pink brain, blue brain: how small differences grow into troublesome gaps—and what we can do about it*. Richmond: Oneworld.

Emirbayer, Mustafa, and Ann Mische. 1998. What is agency? *American Journal of Sociology* 103:962–1023.

Erlinghagen, Marcel. 2019. Employment and its institutional context. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00599-6.

Esping-Andersen, Gøsta. 1990. *The three worlds of welfare capitalism*. Princeton: Princeton University Press.

Evertsson, Marie, and Daniela Grunow. 2016. Narratives on the transition to parenthood in eight European countries. The importance of gender culture and welfare regime. In *Couples' transitions to parenthood: analysing gender and work in Europe*, ed. Daniela Grunow, Marie Evertsson, 269–294. Cheltenham: Edward Elgar.

Evertsson, Marie, and Magnus Nermo. 2004. Dependence within families and the division of labor: comparing Sweden and the United States. *Journal of Marriage and Family* 66:1272–1286.

Fahlén, Susanne. 2016. Equality at home—A question of career? Housework, norms, and policies in a European comparative. *Demographic Research* 35:1411–1440.

Fuwa, Makiko. 2004. Macro-level gender inequality and the division of household labor in 22 countries. *American Sociological Review* 69:751–767.

Fuwa, Makiko, and Philip N. Cohen. 2007. Housework and social policy. *Social Science Research* 36:5112–5530.

Gangl, Markus, and Andrea Ziefle. 2015. The making of a good woman: extended parental leave entitlements and mothers' work commitment in Germany. *American Journal of Sociology* 121:511–563.

Geist, Claudia. 2005. The welfare state and the home: regime differences in the domestic division of labour. *European Sociological Review* 21:23–41.

Geist, Claudia, and Philip N. Cohen. 2011. Headed toward equality? Housework change in comparative perspective. *Journal of Marriage and Family* 73:832–844.

Gershuny, Jonathan. 2018. Gender symmetry, gender convergence and historical work-time invariance in 24 countries. https://www.timeuse.org/sites/default/files/2018-02/CTUR%20WP%202%202018_0.pdf. Accessed 8 June 2018.

Gershuny, Jonathan, and Oriel Sullivan. 2003. Time use, gender, and public policy regimes. *Social Politics* 10:205–228.

Gershuny, Jonathan, Michael Bittman, and John Brice. 2005. Exit, voice, and suffering: do couples adapt to changing employment patterns? *Journal of Marriage and Family* 67:656–665.

Goerres, Achim, Markus B. Siewert and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Gornick, Janet C., and Marcia K. Meyers. 2003. Welfare regimes in relation to pad work and care. *Advances in Life Course Research* 8:45–67.

Greenstein, Theodore N. 2000. Economic dependence, gender, and the division of labor in the home: A replication and extension. *Journal of Marriage and the Family* 62:322–335.

Grunow, Daniela. 2017. Theoriegeleitetes Sampling für international vergleichende Mixed-Methods-Forschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 69:213–235.

Grunow, Daniela, and Gerlieke Veltkamp. 2016. Institutions as reference points for parents-to-be in European societies: A theoretical and analytical framework. In *Couples' transitions to parenthood: analysing gender and Work in Europe*, ed. Daniela Grunow, Marie Evertsson, 3–33. Cheltenham: Edward Elgar.

Grunow, Daniela, Katia Begall, and Sandra Buchler. 2018. Gender ideologies in Europe: A multidimensional framework. *Journal of Marriage and Family* 80:42–60.

Grunow, Daniela, Heather Hofmeister, and Sandra Buchholz. 2006. Late 20th-century persistence and decline of the female homemaker in Germany and the United States. *International Sociology* 21:101–131.

Grunow, Daniela, Florian Schulz, and Hans-Peter Blossfeld. 2012. What determines change in the division of housework over the course of marriage? *International Sociology* 27:289–307.

Gupta, Sanjiv. 2007. Autonomy, dependence, or display? The relationship between married women's earnings and housework. *Journal of Marriage and Family* 69:399–417.

Gupta, Sanjiv, Marie Evertsson, Daniela Grunow, Magnus Nermo, and Liana C. Sayer. 2015. The economic gap among women in time spent on housework in former West Germany and Sweden. *Journal of Comparative Family Studies* 46:181–201.

Hakim, Catherine. 2000. *Research design. Successful designs for social economics research*. London: Routledge.

Hank, Karsten, and Hendrik Jürges. 2007. Gender and the division of household labor in older couples. *Journal of Family Issues* 28:399–421.

Hank, Karsten, and Anja Steinbach. 2019. Families and their institutional contexts: The role of family policies and legal regulations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00603-z.

Hays, Sharon. 1996. *The cultural contradictions of motherhood*. New Haven: Yale University Press.

Heisig, Jan Paul. 2011. Who does more housework: rich or poor? A comparison of 33 countries. *American Sociological Review* 76:74–99.

Hitlin, Steven, and Glen H. Elder. 2007. Time, self, and the curiously abstract concept of agency. *Sociological Theory* 25:170–191.

Hofmeister, Heather, Hans-Peter Blossfeld, and Melinda Mills. 2006. Globalization, uncertainty and women's mid-career life courses: a theoretical framework. In *Globalization, uncertainty and women's careers: an international comparison*, ed. Hans-Peter Blossfeld, Heather Hofmeister, 3–31. Cheltenham: Edward Elgar.

Hook, Jennifer L. 2006. Care in context: men's unpaid work in 20 countries, 1965–2003. *American Sociological Review* 71:639–660.

Kan, Man Yee, Oriel Sullivan, and Jonathan Gershuny. 2011. Gender convergence in domestic work: Discerning the effects of interactional and institutional barriers from largescale data. *Sociology* 45:234–251.

Knight, Carly R., and Mary C. Brinton. 2017. One egalitarianism or several? Two decades of gender-role attitude change in Europe. *American Journal of Sociology* 122:1485–1532.

Knudsen, Knud, and Kari Wærness. 2008. National context and spouses' housework in 34 countries. *European Sociological Review* 24:97–113.

Kohn, Melvin L. 1987. Cross-national research as an analytic strategy: American Sociological Association, 1987 Presidential Address. *American Sociological Review* 52:713–731.

Kühhirt, Michael. 2012. Childbirth and the long-term division of labour within couples: How do substitution, bargaining power, and norms affect parents' time allocation in West Germany? *European Sociological Review* 28:565–582.

Lachance-Grzela, Mylène, and Geneviève Bouchard. 2010. Why do women do the lion's share of housework? A decade of research. *Sex roles* 63:767–780.

Lewin-Epstein, Noah, Haya Stier, and Michael Braun. 2006. The division of household labor in Germany and Israel. *Journal of Marriage and Family* 68:1147–1164.

Lewis, Jane E. 1993. *Women and social policies in Europe: work, family and the state*. Cheltenham: Edward Elgar.

Mahmood, Saba. 2001. Feminist theory, embodiment, and the docile agent: some reflections on the Egyptian islamic revival. *Cultural Anthropology* 16:202–236.

Mandel, Hadas, and Moshe Semyonov. 2006. A welfare state paradox: State interventions and women's employment opportunities in 22 countries. *American Journal of Sociology* 111:1910–1949.

Mencarini, Letizia, and Maria Sironi. 2010. Happiness, housework and gender inequality in Europe. *European Sociological Review* 28:203–219.

Mills, Melinda, Hans-Peter Blossfeld, and Fabrizio Bernardi. 2006. Globalization, uncertainty and men's employment careers: a theoretical framework. In *Globalization, uncertainty and men's careers: An international comparison*, ed. Hans-Peter Blossfeld, Melinda Mills, and Fabrizio Bernardi, 3–37. Cheltenham: Edward Elgar.

Moen, Pyllis. 2003. Linked lives: dual careers, gender, and the contingent life course. In *Social dynamics of the life course: transitions, institutions, and interrelations*, 237–258.

Moreno-Colom, Sara. 2017. The gendered division of housework time: analysis of time use by type and daily frequency of household tasks. *Time & Society* 26:3–27.

Nazio, Tiziana. 2008. *Cohabitation, family, and society. Routledge advances in sociology*. New York: Routledge.

Neilson, Jeffrey, and Maria Stanfors. 2014. It's about time! Gender, parenthood, and household divisions of labor under different welfare regimes. *Journal of Family Issues* 35:1066–1088.

Nitsche, Natalie, and Daniela Grunow. 2016. Housework over the course of relationships: gender ideology, resources, and the division of housework from a growth curve perspective. *Advances in life course research* https://doi.org/10.1016/j.alcr.2016.02.001.

Nordenmark, Mikael. 2004. Does gender ideology explain differences between countries regarding the involvement of women and of men in paid and unpaid work? *International Journal of Social Welfare* 13:233–243.

Orloff, Ann Shola. 1993. Gender and the social rights of citizenship: the comparative analysis of gender relations and welfare states. *American Sociological Review* 303–328. https://doi.org/10.2307/2095903

Orloff, Ann Shola. 2008. Should feminists aim for gender symmetry? Feminism and gender equality projects for a post-maternalist era. Paper presented at the annual conference of the International Sociological Association Research Committee on Poverty, Social Welfare and Social Policy, RC 19 The Future of Social Citizenship: Politics, Institutions and Outcomes. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.576.3826&rep=rep1&type=pdf (Date of access: 15 Aug. 2018)

Pfau-Effinger, Birgit. 2005. Culture and welfare state policies: reflections on a complex interrelation. *Journal of social policy* 34:3–20.

Pillarisetti, Jayasree, and Mark McGillivray. 2002. Human development and gender empowerment: methodological and measurement issues. *Development Policy Review*. https://doi.org/10.1111/1467-7679.00059.

Pinchbeck, Ivy. 2013. *Women workers in the industrial revolution*. London: Routledge.

Presser, Harriet B. 1994. Employment schedules among dual-earner spouses and the division of household labor by gender. *American Sociological Review* 59:348–364.

Reimann, Maria. 2016. Searching for egalitarian divisions of care: polish couples at the life-course transition to parenthood. In *Couples' transitions to parenthood: analysing gender and work in Europe*, ed. Daniela Grunow, Marie Evertsson, 221–242. Cheltenham: Edward Elgar.

Robila, Mihaela. 2004. Families in eastern Europe: context, trends and variations. In *Families in eastern Europe*, 1–14. Oxford: Elsevier.

Ross, Catherine E. 1987. The division of labor at home. *Social forces* 65:816–833.

Ruppanner, Leah E. 2009. Conflict and housework: does country context matter? *European Sociological Review* 26:557–570.

Ruppanner, Leah E. 2010. Cross-national reports of housework: an investigation of the gender egalitarianism measure. *Social Science Research* 39:963–975.

Sainsbury, Diane. 1994. *Gendering welfare states*. Thousand Oaks: SAGE.

Sainsbury, Diane. 1996. *Gender, equality and welfare states*. Cambridge: Cambridge University Press.

Sayer, Liana C. 2010. Trends in housework. In *Dividing the domestic: men, women, and household work in cross-national perspective*, ed. Judith Treas, Sonja Drobnič, 19–38. Stanford: Stanford University Press.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schober, Pia S. 2013. Gender equality and outsourcing of domestic work, childbearing, and relationship stability among British couples. *Journal of Family Issues* 34:25–52.

Schröder, Martin. 2019. Varieties of capitalism and welfare regime theories: Assumptions, accomplishments, and the need for different Methods. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00609-7.

Schulz, Florian, and Daniela Grunow. 2011. Comparing diary and survey estimates on time use. *European Sociological Review* 28:622–632.

Smith, Vicki. 2006. Work and employment. In *The Cambridge dictionary of sociology*, ed. Bryan S. Turner, 676–682. Cambridge: Cambridge University Press.

Sullivan, Oriel, and Jonathan Gershuny. 2016. Change in spousal human capital and housework: a longitudinal analysis. *European Sociological Review* 32:864–880.

Sullivan, Oriel, Francesco C. Billari and Evrim Altintas. 2014. Fathers' changing contributions to child care and domestic work in very low–fertility countries: the effect of education. *Journal of Family Issues* 35:1048–1065.

Sullivan, Oriel, Jonathan Gershuny, and John P. Robinson. 2018. Stalled or uneven gender revolution? A long-term processual framework for understanding why change is slow. *Journal of Family Theory & Review* 10:263–279.

Tamilina, Larysa, and Natalya Tamilina. 2014. The impact of welfare states on the division of housework in the family: a new comprehensive theoretical and empirical framework of analysis. *Journal of Family Issues* 35:825–850.

Thébaud, Sarah. 2010. Masculinity, bargaining, and breadwinning: understanding men's housework in the cultural context of paid work. *Gender & Society* 24:330–354.

Thébaud, Sarah, and David S. Pedulla. 2016. Masculinity and the stalled revolution: How gender ideologies and norms shape young men's responses to work-family policies. *Gender & Society* 30:590–617.

Thévenon, Olivier. 2013. *Drivers of female labour force participation in the OECD*. OECD Social, Employment and Migration Working Papers No. 145. Paris: OECD.

Thompson, Martha E., and Michael Armato. 2012. *Investigating gender*. Cambridge: Polity press.

Tilly, Louise A. 1994. Women, women's history, and the industrial revolution. *Social Research* 61:115–137.

Tilly, Louise A., and Joan W. Scott. 1989. *Women, work, and family*. London: Psychology Press.

Treas, Judith. 2010. Why study housework? In *Dividing the domestic: men, women, and household work in cross-national perspective*, ed. Judith Treas, Sonia Drobnič, 3–18. Stanford: Stanford University Press.

Treas, Judith, and Sonia Drobnič. 2010. *Dividing the domestic: men, women, and household work in cross-national perspective*. Stanford: Stanford University Press.

Treas, Judith, and Tsui Tai. 2016. Gender inequality in housework across 20 European nations: lessons from gender stratification theories. *Sex Roles* 74:495–511.

United Nations Development Programme. 2016. Gender Inequality Index (GII). http://hdr.undp.org/en/content/gender-inequality-index-gii. Accessed 9 June 2018.

United Nations Development Programme. 2018. Human development report: gender inequality index. http://hdr.undp.org/en/composite/GII. Accessed 9 June 2018.

Van der Lippe, Tanja, Judith de Ruijter, Esther de Ruijter, and Werner Raub. 2011. Persistent inequalities in time use between men and women: a detailed look at the influence of economic circumstances, policies, and culture. *European Sociological Review* 27:164–179.

Wall, Glenda. 2010. Mothers' experiences with intensive parenting and brain development discourse. *Women's Studies International Forum* 33:253–263.

West, Candace, and Don H. Zimmerman. 1987. Doing gender. *Gender & Society* 1:125–151.

Yodanis, Carrie. 2005. Divorce culture and marital gender equality: a cross-national study. *Gender & Society* 19:644–659.

Daniela Grunow 1975, Dr. rer. pol., is a Full Professor of Sociology specializing in Quantitative Analyses of Social Change at the Faculty of Social Sciences at the Goethe-University in Frankfurt am Main. Research expertise: Interaction of paid work, domestic work and gender relations in cross-national perspective. Publications: Gender ideologies in Europe: A multidimensional framework. Journal of Marriage and Family. 80, 2018 (with K. Begall and S. Buchler); Couples' transitions to parenthood: Analysing gender and work in Europe. Cheltenham, UK 2016 (ed., with M. Evertsson).

Kölner Zeitschrift
für Soziologie und
Sozialpsychologie

ABHANDLUNGEN

Theory Development in Comparative Social Research

Clemens Kroneberg

Abstract While questions of methodology and research design have received a lot of attention, less is known about theory development in comparative social research. As theoretical objectives and orientations are diverse, theorizing takes many forms, ranging from orienting statements and typologies to different kinds of causal propositions. After introducing different understandings of "theory" and associated types of research questions, the article discusses the interplay between empirical research and theory development in comparative social research. Using examples from different fields of application, I argue that theory development in comparative research can be hampered by placing too much emphasis on general micro-level theories, but also by a lack of theoretical abstraction, that intertwines mechanism sketches with historical and contextual details of the particular macro-level phenomena under investigation. The article calls for a greater focus on meso-level theorizing, as it has the greatest potential to produce theoretical knowledge about contextual variation in causal mechanisms and to motivate the development of theoretical models that are explicit enough to be systematically revised across studies.

Keywords Methodology · Models · Mechanisms · Cross-national research · Analytical sociology

Theorieentwicklung in der international vergleichenden Sozialforschung

Zusammenfassung Im Vergleich zu der Aufmerksamkeit, die Methodologie und Forschungsdesign erhalten, wurden allgemeine Fragen der Theorieentwicklung in

C. Kroneberg (✉)
Institute of Sociology and Social Psychology, University of Cologne
Greinstr. 2, 50939 Köln, Germany
E-Mail: c.kroneberg@uni-koeln.de

🖄 Springer

der international vergleichenden Sozialforschung bislang wenig diskutiert. Aufgrund der vielfältigen theoretischen Zielsetzungen und Orientierungen existieren viele Arten der Theoriebildung, von orientierenden Feststellungen und Typologien bis hin zu kausalen Propositionen. Ausgehend von unterschiedlichen Theorieverständnissen und entsprechenden Arten von Forschungsfragen diskutiert der Beitrag das Wechselspiel von empirischer Forschung und Theorieentwicklung in der international vergleichenden Sozialforschung. Anhand von Beispielen aus unterschiedlichen Forschungsfeldern argumentiert der Autor, dass Theorieentwicklung in der komparativen Forschung oftmals durch zwei Tendenzen behindert wird: einerseits durch die Überbetonung allgemeiner Mikrotheorien, andererseits durch den Mangel an theoretischer Abstraktion in Arbeiten, die Skizzen von Mechanismen zu eng mit historischen und kontextuellen Details konkreter Makrophänomene fusionieren. In diesem Beitrag wird dafür argumentiert, bei der Theorieentwicklung einen stärkeren Fokus auf die Mesoebene zu richten. Dies birgt das größte Potenzial, theoretisches Wissen über kontextuelle Variation in kausalen Mechanismen zu generieren sowie die Entwicklung theoretischer Modelle anzuregen, die explizit genug sind, um systematisch über verschiedene empirische Studien hinweg verbessert werden zu können.

Schlüsselwörter Methodologie · Modelle · Mechanismen · Ländervergleichende Forschung · Analytische Soziologie

1 Introduction

Cross-national comparisons are a core branch of empirical research in the social sciences. Although most research takes place in one-country settings, the claim that cross-national comparisons offer specific insights is well received throughout the social sciences. In sociology in particular, such research often attracts widespread attention and recognition, as it delivers on the discipline's promise to study "big structures," "large processes," and "huge comparisons" (Tilly 1984).

At the same time, international comparative research is diverse with regard to its methodological and theoretical orientations. Even if we limit ourselves to work in the social sciences that aims at systematic description and explanation, there are different worlds of comparative research. Ragin (1987) has contrasted a qualitative, historical, and case-oriented approach with a quantitative, abstractly causal, and variable-oriented approach, and contends that "the split between the two major research strategies is more complete and more profound in comparative social science than in most other subdisciplines" (Ragin 1987, p. viii). Others have agreed in principle with this observation but stressed that the underlying logic and challenges of causal inference remain the same (Goldthorpe 2007; King et al. 1994; critically, see Mahoney and Goertz 2006). However, these discussions have for the most part been concerned with questions of methodology. Theoretical and methodological issues are of course closely intertwined, but this often leads to neglect of the former at the expense of the more tangible issues of methodology and research design.

This article is about theory development in international comparative research. More precisely, rather than summarizing the different kinds of theory employed

(such as system theory, rational choice theory, or institutionalism; see, e.g., Peters 2013, pp. 118–145), I engage with the meta-theoretical questions regarding what kind of theoretical insights can be gained in comparative social research and what kind of theorizing allows for a more or less systematic character of theory development across studies. As exhaustive treatment of this topic is beyond the scope of a journal article, I focus selectively on a diverse set of approaches and examples that allow me to critically discuss important promises and pitfalls of theory development in international comparative social research.

The remainder of this article is structured as follows: The next section discusses different understandings of "theory" and associated common types of research questions in comparative social research. As theoretical objectives and orientations are diverse, theorizing takes many forms, ranging from orienting statements and typologies to different kinds of causal propositions. I then turn to the interplay between empirical research and theory development, and discuss different theoretical rationales for conducting comparative social research. As I will argue, theory development in comparative research can be hampered by placing too much emphasis on general micro-level theories, but also by a lack of theoretical abstraction, that intertwines mechanism sketches with historical and contextual details of the particular macro-level phenomena under investigation. Meso-level theorizing is crucial in order to avoid these problems and ease theory development and accumulation of knowledge across studies. It has the greatest potential to produce theoretical models of limited scope conditions whose primary purpose is to understand contextual variation and that are explicit enough to be systematically revised across studies.

2 Understandings of Theory and Research Questions in Comparative Research

2.1 The Generality of Theoretical Propositions

It has been repeatedly noted that social scientists can mean very different things when they use the words "theory," "theoretical," or "theorize" (Abend 2008; Turner 1991). While this is also true of international comparative research, the strong empirical orientation of this branch of the social sciences means that we can restrict our discussion to a more limited set of predominant understandings. As a starting point, consider Abend's (2008, p. 177) distinction between theory as "a general proposition, or logically-connected system of general propositions, which establishes a relationship between two or more variables" (e.g., modernization theory) and theory as "an explanation of a particular social phenomenon" (e.g., a theory about the origins of the French Revolution). These two understandings have affinities with the two major traditions in comparative social research (Ragin 1991, pp. 1–2; Mahoney and Goertz 2006): Most scholars in variable-based quantitative research subscribe to the first understanding and test relationships between two or more variables based on data from multiple countries. In comparison, proponents of case-oriented comparative research tend to shy away from reducing countries (i.e., the cases) to variables, and instead demand that countries be taken seriously as "meaningful wholes."

Most importantly, the two understandings differ in terms of how general or contextually specific the resulting theory is. The agenda of an "explanatory sociology" was long tied to the covering-law model, or Hempel–Oppenheim scheme, of explanation (Hempel and Oppenheim 1948; Homans 1967; Esser 1993; Opp 2013). According to this model, even explanations of singular events always have to apply general laws that allow deduction of the phenomenon of interest under particular initial conditions. However, the search for stable and general relationships between macro-level phenomena (e.g., economic development and fertility or democratization) has been notoriously unsuccessful. In response, proponents of explanatory sociology have come to locate causal propositions of law-like generality at the micro level—attempting to formulate a general theory of action. Even if this project were successful, however, its end product would not immediately offer theoretical insights into cross-national differences and communalities.[1]

Comparative social research aims at causal propositions that relate macro- and meso-level entities (more generally on this issue of analytical primacy, see Coleman 1990, p. 2; Lindenberg 1992; see also Andreß et al. 2019). For example, such propositions may address how educational systems affect equality of educational opportunity, how the welfare state and labor market policies shape levels and trajectories of unemployment, or which historical developments caused certain welfare states to converge or diverge on a number of characteristics. The unsuccessful search for macrosociological laws means that these causal propositions are unlikely to be law-like statements. In comparative social research, generality or particularity should itself be seen as a matter of degree. Hence, rather than assuming incompatible understandings of "theory" in variable-based and case-based research traditions, it seems more adequate to perceive the difference as one of different working assumptions. While the quest for generalizability or its implicit assumption is common in variable-based quantitative research, proponents of case-based comparative research are usually more cautious when it comes to extrapolating insights beyond the cases at hand or formulating general propositions. For example, Thelen (2002) notes with respect to "the debate about universal versus contextualized theory building":

> [...] those who strive to develop universal theory operate on the assumption that the general part of an explanation will capture a very big part of the story. [...] I think it is safe to say that most historical institutionalists would [...] be more inclined to think that what you might be able to discover at the level of universal laws may be a rather small and maybe even trivial part of the story. The search for middle range theory is thus driven less by a disdain for theory than the conviction that deeper understanding of causal relationships (i.e., good theory) can often be achieved through a more intense and focused examination of a number of carefully selected cases (Thelen 2002, p. 95).

As it is difficult to decide a priori what level of generality can be achieved or will yield the most interesting insights, we should decouple this question from our

---

[1] It should also be noted that new accounts of mechanism-based explanations in analytical sociology place less emphasis on general laws as a necessary ingredient of scientific explanations (for an overview, see Kalter and Kroneberg 2014).

understanding of theory. For the current purpose, I will therefore define theory as *a collection of interrelated concepts, assumptions, and causal propositions to explain a specified set of phenomena*. While this understanding does not require propositions to be "general," it adds the requirement that some of them be causal hypotheses, i.e., that they specify cause–effect relationships. Before elaborating on such propositions, however, the status and value of non-causal propositions should be discussed.

2.2 Non-Causal Propositions: Orienting Statements and Typologies in Comparative Research

A great deal of theorizing in comparative research is devoted to the formulation and refinement of non-causal propositions. Two types are particularly important: orienting statements and typologies. To illustrate, consider the following two classic orienting statements by Marx and Weber:

> The mode of production of material life conditions the general process of social, political and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness (Marx 1859, pp. 11–12).
> Not ideas, but material and ideal interests, directly govern men's conduct. Yet very frequently the 'world images' that have been created by 'ideas' have, like switchmen, determined the tracks along which action has been pushed by the dynamic of interest (Weber 1958, p. 280).

These statements are often presented as deep insights or bold claims, and are certainly constitutive of distinct research programs in historical and comparative sociology. Upon closer scrutiny, however, they fall short of causal propositions, as they do not specify what will change in a given direction if something else changes in a given direction (Homans 1967). This is not to say that orienting statements have no scientific value. They are heuristically useful in guiding our analytical attention by pointing out what kind of research questions we should ask and where we might be able to find answers to these more specific questions. For example, asking about the origins of capitalism as an ideology (the belief in free markets, entrepreneurship, etc.), the orienting statement by Marx would have us seek changes in a society's technological and social modes of production, while Weber's switchmen metaphor would grant an independent causal force to ideas, such as Protestantism. Guided by these orienting statements, one could then search for specific causes, and formulate propositions that specify how exactly they contribute to the emergence and diffusion of particular ideological beliefs.

Another type of general proposition found in comparative research focuses on different types of macro-entities, such as welfare states, varieties of capitalism, or whole societies (Schröder 2019). For example, comparative welfare state research has been and continues to be occupied with mapping real welfare states onto a set of ideal-typical regimes. Building on Esping-Andersen's famous *Three Worlds of Welfare Capitalism* (1990), scholars have discussed how countries outside Esping-Andersen's initial focus fit into the three types of liberal, conservative/corporatist, and the social democratic welfare regimes. A common question in this context is

whether some of the countries bear sufficient family resemblance to assign them to one of these types, or whether they constitute hybrid, i.e., in-between cases (so that the original typology could be left unchanged), or whether they constitute regimes *sui generis* that make it necessary to revise or extend Esping-Andersen's typology (e.g., Arts and Gelissen 2002).

The theoretical value of such discussions does not immediately reveal itself to social scientists who are interested in causal mechanisms. Despite the abstraction involved, typologies do not entail propositions about causal relations, and therefore appear to be at best preparatory for causal explanation, and at worst distractions from the real theoretical tasks. However, it would be wrong to think of typological work in comparative research as merely deficient when measured against the conceptions of causality and theory held by most variable-oriented quantitative researchers. First, the great amount of effort devoted to building typologies of countries originates in a distinctive emphasis on context: A common working assumption among historical institutionalism scholars is that the most informative causal propositions will almost always have limited scope conditions and hold only under more or less particular institutional configurations (Thelen 2002). Typologies of welfare states, varieties of capitalism, and the like take on an important role in identifying likely clusters of scope conditions. Second, the construction of typologies is often based on the testable assumption that particular country characteristics tend to co-occur. For example, Esping-Andersen (1990) states that the analytical objective of his welfare (state) regime approach is "to denote the fact that in the relation between state and economy a complex of legal and organizational features are systematically interwoven" (Esping-Andersen 1990, p. 2). More generally, Lange and Meadwell (1991, p. 84) submit that typologies have "an obvious attraction in being able to characterize whole systems with the related implication that different systemic features hang together." For example, it is justified to speak of a distinctive welfare state regime if there are groups of countries with similar relationships between the state, the market, and the family which give rise to particular properties of the welfare state. Such covariation hypotheses are testable propositions even if one does not provide a causal explanation that shows how an institutional regime came about, i.e., to dissect the causal processes that set countries on a path towards a regime in which the various institutions are closely related (see Ylikoski 2012).

To conclude, even if the product of this type of theorizing in comparative research is "descriptive generalizations" (Esser 2004), these propositions of covariation can provide an important background for the development and testing of causal hypotheses.[2] Of course, how informative typologies are, and how heuristically useful they are in the construction of explanation, depends on the extent to which their grouping of countries identifies which sources of (between-country) variance really matter (see Schröder 2019). Both in case-oriented as well as in variable-oriented research, studies regularly examine whether particular relationships differ across types or regimes, as these typologies would lead one to expect.

---

[2] Descriptive sentences are propositions about (singular or non-singular) states of the world, including regularities of covariations.

2.3 Causal Propositions in Comparative Research

Developing theory in the form of a system of causal propositions is naturally a key goal of comparative research. In addition to differences in generality, causal propositions differ in terms of whether they link more large-scale (or "macro") with more small-scale (or "micro") phenomena, or relate phenomena of the same scale (Ylikoski 2012). This is schematically presented in the well-known macro-micro-macro model by breaking down a macro-macro link into three steps: a macro-micro, a micro-micro, and a micro-macro link (e.g., Coleman 1990; Esser 1993). In international comparative research, the macro level usually corresponds to countries or nation states. A first type of causal proposition remains on this level and attempts to identify country-level determinants of country-level outcomes. For example, it has been argued that economic development promotes democracy, either by making a regime change towards democracy more likely, or by consolidating existing democratic regimes (Lipset 1959; Epstein et al. 2006). To become a theory, such a single causal statement needs to be supplemented by definitions of core concepts (e.g., democracy) and arguments that clarify why this causal relationship exists, under what scope conditions it holds, etc.

Such theoretical reasoning usually involves references to micro (or "meso") entities such as individuals, families, groups, or organizations that bring about the macro-level outcome. However, this is not a logical requirement. To stick with our example, in his theory of evolutionary universals, Talcott Parson (1964) argued that democracy is a universally necessary precondition if large-scale industrial societies are to move to higher levels of complexity and adaptive capacity. The more complex the society, the more it needs effective political organization; this, in turn, requires that it be based on the broad societal consensus that only widespread democratization can achieve (Parsons 1964).[3] However, as the heydays of system theory and functionalism are long over, such purely macro-level reasoning comprises only a minor fraction of contemporary comparative scholarship. Given that micro-micro relationships are not at the center of comparative research either (see Sect. 3.2 below), most causal propositions relate macro and micro levels.

An inexhaustible source of causal propositions is variable-based quantitative comparative research. This does not mean, however, that these propositions form a theory in the sense of a logically connected system. Often plausibility arguments or relatively loose frameworks and concepts provide the basis to formulate a theoretical expectation. For example, Putnam (2007) has put forward the contra-intuitive and politically provoking hypothesis that ethnic diversity undermines not only trust towards ethnic outgroups, but also among ingroup members ("hunkering down"). Notwithstanding the fact that this theoretical argument inspired a great deal of research, it has meanwhile become clear that it was poorly integrated with established

---

[3] A fundamental problem of Parson's reasoning is that it is not possible to scientifically assess the long-term adaptive capacity of societies since we cannot predict future environmental conditions (Granovetter 1979). Moreover, his structural functionalism has been criticized for not being able to understand the causes that lead to particular structures, as well as for ignoring exogenous influences stemming from other societies.
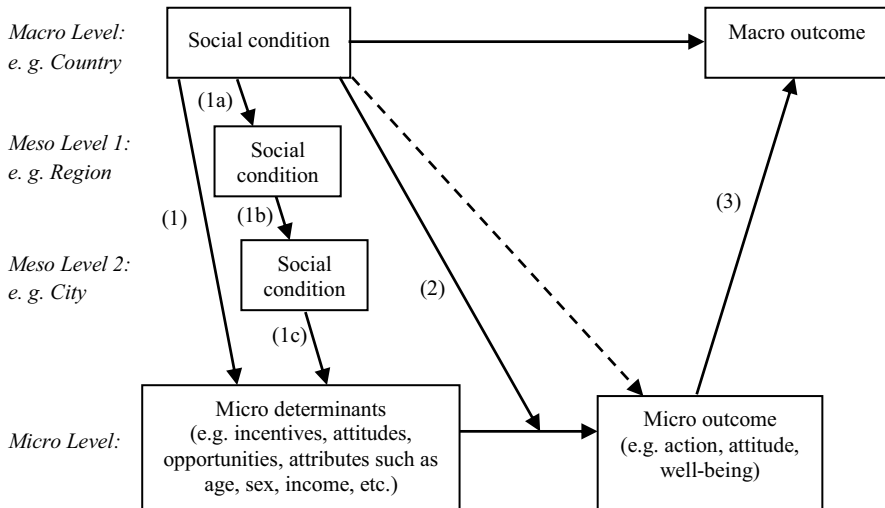
Fig. 1 An elaborated macro-micro-macro model (adapted from Nonnenmacher and Friedrichs 2013, p. 1223)

theories of intergroup contact and other causal knowledge about the consequences of neighborhood composition (Abascal and Baldassarri 2015; Christ et al. 2014).

The most important causal propositions in variable-based quantitative comparative research frequently describe contextual effects. Methodologically more advanced research uses multilevel data in which observations on individuals are nested in countries. The kinds of contextual hypotheses put forward and tested in such research can be illustrated in an elaborated macro-micro-macro model (Nonnenmacher and Friedrichs 2013, p. 1223).

In Fig. 1, the left-hand arrow represents the well-known macro-micro link in which the national context or country-wide institutions such as the type of welfare state are assumed to affect individuals' preferences, attitudes, etc. Common arguments point to effects of socialization on preferences, exposure (e.g., to ethnic diversity or social norms), psychological states (e.g., attitudes or wellbeing), institutions (e.g., labor market policies), incentives and opportunities, and more. In principle, it is also possible to assume direct effects on behavior (the dashed arrow in the middle of Fig. 1), but these will most likely be mediated through preferences, attitudes, opportunities, or some other individual-level determinant of behavior. However, indicators of concepts that tap into action-generating mechanisms are not always available. Cross-national survey research therefore often adopts a so-called indirect test strategy in which scholars assume unobserved individual-level determinants of behavior (e.g., preferences) to be systematically related to more easily observable attributes, such as family status, occupation, or ethnic origin (Kroneberg and Kalter 2012, p. 77).

In variable-oriented comparative research, there has been a recent trend towards operationalizing contextual variables at the meso level, i.e., replacing the national context with regions, cities, districts, or neighborhoods (see arrows 1a to 1c in

Fig. 1). In addition to the increased availability of more detailed contextual data, this move has been theoretically motivated by the desire to provide more direct robust evidence for assumed mechanisms. As is argued by Nonnenmacher and Friedrichs (2013, p. 1224) in their critical review of research on subjective wellbeing, "in many cases, not the entire country will be perceptible or relevant for the individual." However, care should be taken not to replace "methodological nationalism" (Wimmer and Glick Schiller 2002) with a form of methodological localism. For example, individuals with a migration background can feel subjected to "othering" discourses both at the national level (e.g., through media), and in their local surroundings. If the former impact is stronger and more permanent, retaining a focus on national-level contextual variation might yield greater explanatory power.

A second group of hypotheses assumes that the micro-level effects of individual characteristics on actors' behavior and attitudes are conditioned by macro- or meso-level contexts (see arrow 2 in Fig. 1). In statistical terms, this corresponds to a cross-level interaction. For example, it has been argued that the impact of unemployment on short- and long-term labor market success (e.g., the opportunity to get a new job and the quality of the job compared to the previous one) depends on labor market policies and welfare regulations (see e.g., Gangl 2006).

The notion that the country context conditions the relationships between individual-level variables is a key rationale for cross-national comparative survey research. To clearly distinguish between this conditioning influence and the direct effect on individual-level variables, one can use Lindenberg's social production function (SPF) theory (Lindenberg 1996). This theory views societies (or cultural and institutional contexts more generally) as chains of production functions that define the legitimate means by which actors can satisfy their human needs, such as physical and social wellbeing. Depending on their resources and restrictions, actors will invest in one or the other social production function, e.g., occupational career, leisure time activity, form of political engagement, or family status. Based on this framework, the two kinds of contextual hypotheses describe how the social context affects actors' resource endowments as well as the effectiveness and efficiency (or "relative prices") of using these resources in pursuing their goals. Note that socialization effects can also be integrated in this framework, e.g., by specifying how lower-order goals become more or less efficient in satisfying higher-order goals as a result of learning experiences.

This focus on contextual influences also clarifies why, and under what conditions, the country context is so important. For some research questions, engaging in cross-national comparisons is just a way to increase the variance on certain predictor variables (e.g., poverty rate, ethnic diversity), and the same processes could be studied by comparing different contexts such as firms or neighborhoods. At other times, however, the country context takes on unique and overwhelming importance for comparative research: This significance derives from the power of the nation state and associated institutions such as welfare states, national media, national culture and languages, and so on. Given the continued relevance and strength of nation states and their institutions, not only but particularly in Western countries, understanding their impact on individuals will continue to motivate a great deal of comparative research.

A third kind of causal proposition asks how the behavior and attitudes of individuals or groups bring about certain characteristics at the macro level of countries. In variable-oriented comparative research—e.g., multilevel models—this step is usually ignored. Implicitly, the assumption is one of simple aggregation: If we examine empirically why citizens vote for right-wing parties, couples get divorced, or individuals hold post-materialistic values, this will also allow us to explain election results, divorce rates, or the level of post-materialism at the country level. However, even in situations in which this assumption seems justified, additional insights can be gained by theorizing or explicitly modeling the micro-macro transition. This can be done in case-oriented comparative research in the form of process tracing (Bennett and Checkel 2014). In variable-oriented research, simulations can be used. For example, if we assume that the prevalence of a certain behavior or state (e.g., divorce or unemployment) will make similar behavior more likely as part of a self-reinforcing dynamic, one can model how an empirically estimated contextual effect may add up over time and lead from a small initial difference to drastically different macro outcomes—which can then again be compared to empirical distributions (see Hedström 2005, Chap. 6). Work that applies explicit theoretical models of such dynamic processes to explain puzzles of comparative research is still rare, however (for a recent exception, see Esping-Andersen and Billari 2015).

2.4 From Causal Propositions to Theoretical Models

At first sight, the theoretical status of case-oriented typological work in comparative research seems more difficult to grasp than that of quantitative variable-oriented research, with its clear focus on formulating and testing causal propositions. However, from Boudon (1976) to Manzo (2014), it has been emphasized that the statistical analysis of large-scale data as such constitutes a multivariate description. In particular, the fact that regression models force researchers to clearly assign the status of dependent and independent variables should not be taken to imply that such research automatically contributes to the formulation of a system of general propositions or an explanation of a particular social phenomenon. Quantitative studies often test three to eight hypotheses, confirming some of them and rejecting others, but without clearly articulating how this evidence contributes to the refutation, modification, or refinement of the theories that inspired these hypotheses.

The dangers are of a similar nature. The orienting statement of case-oriented research is that there are commonalities and differences between countries. If one sets out to identify similarities and differences across countries, one will surely return from this endeavor with similarities and differences. The orienting statement of variable-oriented research is that some variables are systematically related, and that empirical evidence for the existence of such relationships is likewise the regular result of statistical analyses. As has been powerfully argued by Homans (1967) in his book on the nature of social science, the theoretical progress of the discipline hinges on us moving from orienting statements to systems of propositions and explanations. More and better data, sophisticated methodology, or differentiated in-depth knowledge about cases only contribute to theory development if they are coupled with and oriented toward the improvement of explicit theoretical models.

Having discussed different forms of theory, or products of theorizing, in international comparative research, I now turn to this even more important but far less broadly discussed topic, namely the process of theory development across studies.

3 Macro and Micro Perspectives on Theory Development in Comparative Research

A main challenge for theory development in international comparative research is to formulate theoretical models that primarily aim to understand contextual variation and that are explicit enough to be systematically revised across studies. By combining these two characteristics, such models have the greatest capacity for systematic theory development in comparative research. As I will illustrate based on selected examples, choice-centered theorizing in comparative research tends to struggle with the first criterion, while phenomenon-centered macro-level theorizing often lacks the necessary explicitness.

I begin with a classic example from social policy research that would commonly be regarded as genuine macrosociology (Sect. 3.1), and subsequently turn to work in which micro-level theories form the starting point of international comparative research (Sect. 3.2). Finally, I illustrate how meso-level theorizing promotes systematic theory development in comparative research (Sect. 3.3).

3.1 Phenomenon-Oriented Theorizing and Comparative Research

Similar to Lindenberg (1996), I distinguish between choice- and phenomenon-centered work in international comparative research. Choice-centered research is heavily invested in particular theories of action, behavior, or choice, whereas phenomenon-centered studies start from a pronounced interest in a particular macro-level phenomenon. The bulk of theorizing and research in comparative research is phenomenon-oriented. Even if scholars use microfoundations as part of their explanatory efforts, these are only of secondary importance, and often remain implicit or sketchy. While this follows naturally from the interest in variation across geographical contexts and historical time, the implicit and often sketchy character of micro- and meso-level mechanisms can hinder theory development and the accumulation of knowledge across studies.

To illustrate these problems, I use the example of Korpi and Palme's (1998) seminal article "The Paradox of Redistribution and Strategies of Equality." The title-giving paradox is the finding that countries which target welfare transfers at the poor do worse in terms of poverty reduction than countries where transfers are distributed universally. To explain this macro-level association, the authors sketch a mechanism based on Korpi's earlier writings (Korpi 1980, 1983):

> This rather unexpected outcome was predicted as a consequence of the type of political coalitions that different welfare state institutions tend to generate. Because marginal types of social policy programs are directed primarily at those

below the poverty line, there is no rational base for a coalition between those above and those below the poverty line. In effect, the poverty line splits the working class and tends to generate coalitions between better-off workers and the middle class against the lower sections of the working class, something which can result in tax revolts and backlash against the welfare-state. [...] The hypothesis here is that the size of the budget available for redistribution is not fixed and that the institutional structures of welfare states are likely to affect the definitions of identity and interest among citizens (Korpi and Palme 1998, p. 663).

In a nutshell, targeting welfare transfers at the poor will drive a wedge between those above and those below the poverty line. The ensuing political coalitions will tend to produce a smaller overall welfare budget than is the case when more universal policies are in place.

Of course, Korpi and Palme's piece is much richer in conceptual and theoretical terms, and has rightfully become a classic of social policy research. More recently, Brady and Bostic (2015) revisited Korpi and Palme's question using better data and methods. Analyzing a greater number of countries and decades, Brady and Bostic found that Korpi and Palme's main findings held for the 1980s and for the smaller sample of countries on which they based their original analysis. In later decades, however, low-income targeting ceases to be negatively related to transfer share and positively to poverty transfer. This disappearance of Korpi and Palme's original paradox does not come as a surprise to Brady and Bostic:

[...] there have been fundamental changes since the mid-1980s period KP (Korpi and Palme; CK) studied. Social inequality has risen considerably in most rich democracies, and the political coalitions supporting welfare states have been transformed. Partly as a result, the welfare states of the mid-2000s are quite different from the welfare states of the mid-1980s [...]. Earlier welfare states were often still growing, rested on low unemployment and smaller elderly populations, and covered most residents. By contrast, today's welfare states face neoliberalism and austerity, a dualization of insiders and outsiders, and daunting demographics (Brady and Bostic 2015, p. 270).

The authors therefore state that "even if KP's arguments were correct in the mid-1980s, it remains an open question if they still apply today" (Brady and Bostic 2015, p. 270). In concluding, they remark that results differ due to "the relationships between key variables changing over time" (Brady and Bostic 2015, p. 291).

While the idea that new times or new cases require new theories seems to be self-evident to large sections of the social sciences, it is a source of fundamental concern for analytical sociologists (see, for example, Esser 2004). How can we give theory development a systematic character and approach the ideal of an accumulation of knowledge (within the bounds of a paradigm) if our theories are bound to change in the course of social change?

The way out of this problem can be outlined as follows: First, Korpi and Palme's paradox should be seen as a particular empirical phenomenon in need of explanation: the associations between welfare state policies, transfer shares, and poverty. As these

variables are located at the macro level, it is unsurprising that they are historically contingent. Second, the main *theoretical* contribution of Korpi and Palme's work consists of proposing causal mechanisms that may be responsible for these relationships. These mechanisms—i.e., the formation of identities, interests, and political coalitions in response to welfare state policies—can be expected to be more robust than the macro-level associations they produce. Still, however, they only operate under certain scope conditions and in tandem with other mechanisms that might undermine or modify their effects.

When basic relationships at the macro level change, this could be due to a change in the initial parameters of such a model, in the model lacking certain elements that need to be incorporated, etc. Extending the number of countries or decades should be seen first and foremost as an opportunity to test additional implications of the theoretical model, to learn about its scope conditions, or about the relevance of other mechanisms. Among the possible results are a stronger corroboration, partial refutation, extension, or refinement of an existing theory. Without a more fully specified theoretical model at hand, however, a change of macro-level relationships is almost uncontrollable or, more precisely, much more difficult to deal with in a way that contributes to cumulative theory development.

3.2 Choice-Centered Theorizing and Comparative Research

Basing comparative research on explicit microfoundations may appear to be a safeguard against these problems of phenomenon-oriented research. However, a strong micro theory does not per se guarantee theoretical progress in comparative research. On the contrary, micro theories of universalist ambition tend to produce their own, albeit different, problems in the context of comparative research. Examples of such universalist theories include Schwartz' theory about the content and structure of values (Schwartz 1992; see also Cieciuch et al. 2019), or theories of action such as rational choice theories or more recent alternatives (see Kroneberg and Kalter 2012). The relation between such theories and international comparative social research is more distant or instrumental. The explananda of theories of action, or social-psychological theories of values, are located at the micro level. Accordingly, the most conclusive tests of such theories can be designed in laboratory experiments or field experiments, or based on tailored detailed survey data that are rarely available for a larger set of countries.

Testing a micro theory with a claim to universal validity in comparative research is like negotiating an obstacle course: The researcher hopes that the theory will pass all its tests, i.e., prove applicable and valid in all countries, so that she or he can uphold its original claim to universal applicability and validity. Accordingly, deviant cases where the hypothesized relationships do not hold are often approached rather defensively by questioning measurement equivalence or pointing to other potential sources of bias—which are indeed likely to exist in cross-national multi-purpose survey data. Hence, a serious challenge consists of finding a constructive way to remain open to and constructively build on cross-national variation. Only then can comparative research play the important role of a corrective for theory development.

Let me again consider an example at greater length. Behavioral economics has become very successful in predicting behaviors in game-theoretical experiments (Camerer 2003). In particular, models of social preferences make it possible to explain behavioral regularities that escape narrower rational choice models which assume material egoism. However, it took a major effort in international comparative research to demonstrate that individuals' preferences are much more contextually contingent than was assumed based on research in Western countries. Henrich et al. (2001, 2010) conducted cross-cultural research that documented striking differences in preferences across societies. For example, in contrast to the results commonly attained in student samples in the US or Europe, subjects in some small-scale societies tended to reject not only unfair offers in the ultimatum game, but also hyper-fair offers (i.e., a share greater than 50%; Henrich et al. 2001, p. 75). Comparative research thus showed that the models of social preferences developed based on student samples in Western societies have limited scope.

In this case, comparative research not only acted as an important corrective to universalist claims, but it also inspired further theory development: At the macro level, the researchers found that individuals' concern for fairness in the game-theoretical experiments was a function of societies' degree of market integration, measured as households' average percentage of calories purchased (Henrich et al. 2010). Of course, this is just an association that is itself in need of explanation, and marks at best the beginning of the development of a theory. Remarkably, the object of such a theory would no longer be choices in behavioral games, but cross-cultural differences.

The findings might have even more fundamental consequences at the micro level. While the interdisciplinary team of anthropologists and economists might have favored different interpretations, one account suggests abandoning the hope of experimental economics to identify generalizable dispositions that allow one to continue with the practice of treating preferences as givens. Instead, the stark cultural differences lend themselves to a more sociological interpretation:

> [...] our abstract game structures may cue one or more highly context specific behavioral rules [...]. According to this interpretation, our subjects were first identifying the kind of situation they were in, seeking analogs in their daily life, and then acting in an appropriate manner. In this case, individual differences result from the differing ways that individuals frame a given situation, not from generalized dispositional differences (Henrich et al. 2004, p. 48).

While this article is not the place to discuss the potential offered by different theories of action, the example demonstrates that the relationship between international comparative research and general microfoundations can go both ways: Comparative research is not only a supplier of explananda to which theories of individual behavior can be applied, but can at times motivate scholars to modify these theories or to reconsider their status as more or less universal. However, there is nothing in

micro-level theories with a universalist ambition that encourages the development of comparative macro- or meso-level theories.[4]

3.3 Meso-Level Theorizing and Comparative Research

In the two previous sections, I pointed to problems of comparative research that is either too centered around a micro-level theory, or which blends mechanism sketches with historical and contextual details of the particular macro-level phenomena under investigation. To avoid these problems and ease theory development and accumulation of knowledge across studies, meso-level theorizing is crucial. To illustrate and elaborate this argument, I discuss Elinor Ostrom's work on the governance of common-pool resources, which is an exemplary case of a meso-level research agenda with great potential for systematic theory development in comparative research.

The classic reference point for Ostrom's work is Garrett Hardin's seminal article "The Tragedy of the Commons" (Hardin 1968), in which he argued that the users of an open-access commons might ultimately destroy the very resource on which they depend because they are trapped in a social dilemma: Individually rational, selfish actors would use a resource until the expected benefits equal the expected costs. However, as these individual decisions fail to take into account the costs imposed on others (external effects), they cumulate in overuse and the potential destruction of the commons, as in cases of overfishing, traffic congestion, or global climate change. Ostrom's own theoretical and empirical work challenges key assumptions and conclusions of Hardin's analysis. Three features of Ostrom's work are particularly important for the purpose at hand: strong microfoundations, the primacy of a meso-level agenda, and comparative research.

First, the game-theoretical analysis of the tragedy of the commons is a prime example of the virtues of microfoundations. It forces analysts to be explicit about their assumption as regards actors' preferences, action alternatives, knowledge, and beliefs. Crucially, Ostrom has been among the scholars who have broken with a rational-choice orthodoxy that at times tries to hold on to unrealistic assumptions for reasons of parsimony or analytical tractability—chief among them the assumption of rational egoism. Instead, her work forcefully illustrates the real-world significance of social preferences. However, as I argued above, explicit microfoundations alone do not make for good comparative theory building.

Second, and most importantly, Ostrom's primary analytical interest is in how the tragedy of the commons can be avoided depending on the attributes of common-pool resources (biophysical conditions), attributes of the community, and the existing institutions for ensuring fair access and sustained availability. As emphasized by Ostrom herself, theorizing at this meso level is closely tied to the theoretical model

---

[4] The theory of action to which I myself contributed provides another example: The Model of Frame Selection (Esser and Kroneberg 2015), which highlights the relevance of dual processes and framing for human behavior, does not carry in itself an agenda for comparative research. However, as the examples of Nauck (2007) and Messner (2012) show, one can combine such a theory with explicit meso- and macro-level theorizing to guide and enrich comparative research.

of interaction. With respect to the institutional rules used in governing the commons, she notes:

> „Since we had identified seven working parts of a game or action situation itself, it seemed reasonable to think of seven broad types of rules operating as external variables affecting the individual working parts of action situations […] The seven types of rules are:
>
> 1. Boundary rules that specify how actors were to be chosen to enter or leave these positions;
> 2. Position rules that specify a set of positions and how many actors hold each one;
> 3. Choice rules that specify which actions are assigned to an actor in a position;
> 4. Information rules that specify channels of communication among actors and what information must, may, or must not be shared;
> 5. Scope rules that specify the outcomes that could be affected;
> 6. Aggregation rules (such as majority or unanimity rules) that specify how the decisions of actors at a node were to be mapped to intermediate or final outcomes; and
> 7. Payoff rules that specify how benefits and costs were to be distributed to actors in positions" (Ostrom 2010, pp. 651–652).

In theorizing institutional rules, Ostrom and her collaborators were therefore guided by an explicit model of interaction, and asked "what part of an action situation is affected by a rule" (Ostrom 2010, p. 652).[5]

Ostrom's research agenda goes beyond the choice-centered research described above, given that the primary aim is not to test the validity of the underlying theory of individual behavior and model of collective action, but to formulate causal propositions at the meso or macro level. Tellingly, Ostrom notes that "(c)onceptualizing seven broad types of rules (rather than one or two) has been upsetting to scholars who wanted to rely on simple models of interactions among humans" (Ostrom 2010, p. 652). This search for meso-level propositions comes with a major shift in analytical attention, and motivates a truly comparative research agenda.

This comparative orientation is the third feature of Ostrom's work that is relevant in the present context. Ostrom and her collaborators reviewed more than 500 case studies on communities in which resource users had self-organized in order to govern common-pool resources (Poteete et al. 2010, Ch. 4). Beyond the seven types of rules identified above in theoretical terms, their meta-analysis found multiple variants of each type (e.g., 27 boundary rules and 112 different choice rules). Interestingly though, statistical analyses of these cases failed "to find rules that worked across ecological, social, and economic environments" (Ostrom 2010, p. 652). Being unable to formulate propositions that link specific rules to varying levels of success in governing the commons, Ostrom moved up a level in generality, and character-

---

[5] Of course, this heuristic strategy is well known from the macro-micro-macro model, in which these links between these institutional rules and parameters of the interaction model can be classified either as bridge assumptions or conditions of transformation.

ized more broadly what kind of institutional regularities were associated with long-sustained regimes while absent in the cases of failure (see Ostrom 1990, 2005; Cox et al. 2010). These best practices (or "design principles") include the following:

"Collective-Choice Arrangements: Most individuals affected by a resource regime are authorized to participate in making and modifying its rules."
"Graduated Sanctions: Sanctions for rule violations start very low but become stronger if a user repeatedly violates a rule."
"Conflict-Resolution Mechanisms: Rapid, low-cost, local arenas exist for resolving conflicts among users or with officials."
"Minimal Recognition of Rights: The rights of local users to make their own rules are recognized by the government" (Ostrom 2010, p. 653).

The extent to which these design principles are present affects the likelihood of sustained success or failure in governing the commons. Hence, Ostrom's theory-driven and fieldwork-intensive search leads to meso-level causal propositions. These propositions are based not only on comparative research, but could also be used to explain differences between communities. Most importantly, the fact that these propositions grow out of a research program with strong microfoundations and explicit theory development means that negative evidence is much less unsettling. When new evidence seems to contradict one or several of these propositions, one does not have to find refuge in orienting statements such as "institutions matter." Rather, there is a whole theoretical apparatus that allows one to accommodate newly uncovered or deviating relationships by extending the more general theoretical framework or revising specific theories and models (see Ostrom 2010, p. 646). It is in this way that Ostrom's research agenda illustrates how explicit meso-level theorizing ensures systematic theory development in comparative research.

Its strengths notwithstanding, the above discussion is not meant to suggest that Ostrom's work could serve as a general blueprint for comparative social research. Most importantly, comparative work in macrosociology and political science often starts out from a research question that focuses on cross-national variation without knowing all the relevant actors and the interaction system up front. The challenge in the macro-micro-macro model is then to move downward and theorize the relevant meso- and micro-level processes. The process of theory building will therefore differ, depending on the starting point of comparative research. In any case, however, theory development across studies would be eased considerably if this process were strongly oriented towards formulating explicit theoretical models of meso-level processes.

A useful heuristic device for constructing such models is still Boudon's scheme for the analysis of social processes (Boudon 1981). As shown in Fig. 2, this heuristic framework consists of three blocks: an interaction system, the relevant environment, and the outcomes of the interaction system. These blocks are linked through direct causal links (paths a and b in Fig. 2) and potential feedback processes (paths c, d, and e). The scheme therefore makes it possible to represent reproductive, cumulative, and transformative social processes. Not accidently, Ostrom's theoretical framework mirrors this scheme (see Ostrom 2005, p. 15): In her case, the environment is made up of the biophysical conditions characterizing the common-pool resources, the attributes of the community, and the institutions that regulate access and availability.

Fig. 2 Boudon's scheme for the analysis of social processes (adapted from Boudon 1981)
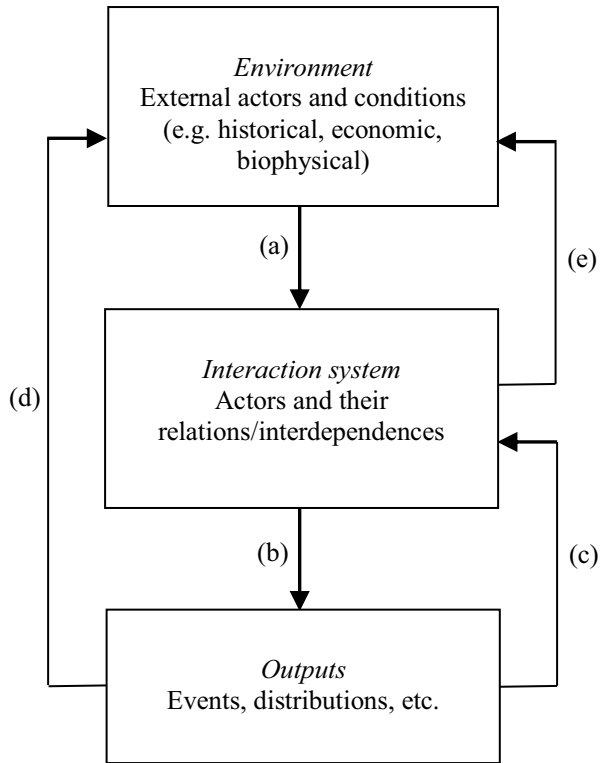
These conditions affect the interaction system, which in turn produces outcomes that potentially feed back into the interaction system or the environmental conditions.

Boudon's scheme is much less prominent than the more general macro-micro-macro model. While the two schemes are mutually consistent and belong to the same history of ideas (Raub and Voss 2017, pp. 26–27), Boudon's scheme has the crucial advantage of directing analytical attention towards meso-level processes and their interplay with larger contexts. In particular, the interaction system in Fig. 2 describes the interdependences and interactions among different categories of actors. Hence, this core building block already describes meso-level processes that could be represented as several macro-micro-macro transitions. The scheme then invites researchers to think about the wider social context in which this interaction system is embedded and how its outputs might feed back into this environment.

Thus, using Boudon's scheme as a heuristic device can help solve the problems of both phenomenon- and choice-centered comparative research. It forces phenomenon-centered analysts to specify the interaction system—i.e., the relevant actors, interdependences, and interactions—separately from the description of historically variable environmental conditions. And it does not overemphasize micro theories of univer-

salist ambition as in choice-centered research that relies exclusively on the macro-micro-macro model.[6]

4 Conclusion

A number of lessons can be drawn from our discussion of macro- and micro-level theory development in comparative research. A main problem of phenomenon-centered theorizing is that the overriding interest in particular macro phenomena leads to an inextricable fusion of conceptual and typological work, mechanism sketches, and empirical results. Such work tends to be rich in its description of national institutional contexts, the role played by various collective and corporative actors, and the feedback processes between different institutions, policies, and social dynamics. However, this line of work often lacks a more explicit formulation of theoretical models. While such models tend to be more abstract and also more artificial than the national contexts under study, they can become the object of evidence-based modification, extension, and refinement. Without such models, theoretical continuity is often sought in retaining particular vocabulary or orienting statements, such as the broader claim that dimensions of welfare institutions have effects on equality and politics. While this allows researchers in a field to retain a sense of consistency, it falls short of cumulative theory development.

On the other hand, there is choice-centered theorizing at the micro level that usually attracts scholars whose interests and expertise do not lie in international comparison or national commonalities and differences. I share Lindenberg's view that this causes a problem when "the main task is to show that choice theory can be applied to the phenomena at hand, rather than to advancing knowledge in the field that specializes in the study of the phenomenon." (Lindenberg 1996, p. 147). Of course, there are exceptions such as rational choice institutionalism or the analytic narratives project that has applied rational choice theory to gain a deeper understanding of historical events and processes (Bates et al. 1998). Generally, however, the gap between micro-level theories and the big structures and large processes of comparative sociology remains wide. From the perspective of historical or comparative sociologists, interests in abstract or formal theory too often take priority over the cases under study. Moreover, when the starting point is the belief in the universal applicability and validity of a micro-level theory, international comparative research can serve virtually no other purpose than to confirm or disprove this belief—and reactions to unexpected deviant cases rarely lead to the development of new theory that advances the macrosociological field.

Against this background, I argue that the most fruitful way forward in developing theory in international comparative research is to formulate theoretical models of

---

[6] In its most basic form, the macro-micro-macro model only requires researchers to provide microfoundations and explicit macro-micro and micro-macro links (Fig. 1 above). In choice-centered research, the micro-micro link often comes down to the explanation of individual behavior based on a theory of action. The determinants of behavior are then related to social contexts, and a more or less complex argument derives macro-level consequences from actors' choices. Hence, the simple macro-micro-macro model does not require an analysis of meso-level processes and their dynamic interplay with larger contexts.

meso-level processes, as called for by Boudon's (1981) scheme for the analysis of social processes. To return to the example of Korpi and Palme (1998): We need to theoretically model the causal mechanisms by which welfare state institutions affect the formation of interests, preferences, and coalitions among citizens. Such meso-level models neither remain on the macro level of society, nor on the micro level of individual behavior. Those who are interested in welfare state regimes will find such theories in political science, where various models of party competition, coalition formation, veto players, and more have been developed over the years (see e.g., Gehlbach et al. 2016; Laver 1998). This is not to suggest that these models can be applied off-the-shelf. Quite the contrary, they suffer from their own problems, such as multiple equilibria or other sources of indeterminacy that unfortunately tend to increase the more realistic the underlying assumptions become.

What is required is collaborative research in which those who are interested in theoretical models join forces with specialists in a particular field of comparative research. For this to occur more often, analytical sociologists have to shift their interest from generic mechanisms to more applied substantive questions, and from highly artificial worlds to the complexities of real-world institutional settings. And scholars in comparative research have to become more open to the virtues of explicit theoretical models as a way to organize their findings—even if this forces them to conduct thought experiments in more abstract artificial worlds. Formulating and revising more fully specified theoretical models of causal mechanisms promises a much greater continuity and systematic character of theory development in comparative research—whether case- or variable-based. This should be worth the journey.

References

Abascal, Maria, and Delia Baldassarri. 2015. Love thy neighbor? Ethnoracial diversity and trust reexamined. *American Journal of Sociology* 121:722–782.

Abend, Gabriel. 2008. The meaning of 'theory'. *Sociological Theory* 26:173–199.

Andreß, Hans-Jürgen, Detlef Fetchenhauer and Heiner Meulemann. Eds. 2019. *Cross-national comparative research – analytical strategies, results, and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00594-x.

Arts, Wil, and John Gelissen, 2002. Three worlds of welfare capitalism or more? A state-of-the-art report. *Journal of European Social Policy* 12:137–158.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast. 1998. *Analytic narratives*. Princeton, NJ: Princeton University Press.

Bennett, Andrew, and Jeffrey T. Checkel. Eds. 2014. *Process tracing: From metaphor to analytic tool*. Cambridge: Cambridge University Press.

Boudon, Raymond. 1976. Comment on hauser's review of education, opportunity, and social inequality. *American Journal of Sociology* 81:1175–1187.

Boudon, Raymond. 1981. *the logic of social action: An introduction to sociological analysis*. London: Routledge & Kegan Paul.

Brady, David, and Amie Bostic. 2015. Paradoxes of social policy: Welfare transfers, relative poverty, and redistribution preferences. *American Sociological Review* 80:268–298.

Camerer, Colin F. 2003. *Behavioral game theory*. Princeton: Princeton University Press.

Christ, Oliver, Katharina Schmid, Simon Lolliot, Hermann Swart, Dietlind Stolle, Nicole Tausch, Ananthi Al Ramiah, Ulrich Wagner, Steven Vertovec and Miles Hewstone. 2014. Contextual effect of positive intergroup contact on outgroup prejudice. *Proceedings of the National Academy of Sciences* 111:3996–4000.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt and René Algesheimer. 2019. How to obtain comparable measures for cross-national comparisons. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00598-7.

Coleman, James S. 1990. *Foundations of social theory*. Cambridge, MA: Belknap Press.

Cox, Michael, Gwen Arnold and Sergio Villamayor-Tomás. 2010. A review of design principles for community-based natural resource management. *Ecology and Society* 15:38.

Epstein, David L., Robert Bates, Jack Goldstone, Ida Kristensen and Sharyn O'Halloran. 2006. Democratic transitions. *American Journal of Political Science* 50:551–569.

Esping-Andersen, Gøsta. 1990. *The three worlds of welfare capitalism*. Princeton, N.J.: Princeton University Press.

Esping-Andersen, Gøsta, and Francesco C. Billari. 2015. Re-theorizing family demographics. *Population and Development Review* 41:1–31.

Esser, Hartmut. 1993. *Soziologie. Allgemeine Grundlagen*. Frankfurt a. M.: Campus.

Esser, Hartmut. 2004. Does the "new" immigration require a "new" theory of intergenerational integration? *International Migration Review* 38:1126–1159.

Esser, Hartmut, and Clemens Kroneberg. 2015. An integrative theory of action: The model of frame selection In *Order on the edge of chaos: Social psychology and the problem of social order*, Eds. Edward J. Lawler, Shane R. Thye and Jeongkoo Yoon, 63–85. New York: Cambridge University Press.

Gangl, Markus. 2006. Scar effects of unemployment: An assessment of institutional complementarities. *American Sociological Review* 71:986–1013.

Gehlbach, Scott, Konstantin Sonin and Milan W. Svolik. 2016. Formal models of nondemocratic politics. *Annual Review of Political Science* 19:565–584.

Goldthorpe, John H. 2007. *On sociology—volume one: Critique and program*. Stanford: Stanford University Press.

Granovetter, Mark. 1979. The idea of "advancement" in theories of social evolution and development. *American Journal of Sociology* 85:489–515.

Hardin, Garrett. 1968. The tragedy of the commons. *Science* 162 (3859):1243–1248

Hedström, Peter. 2005. *Dissecting the social: On the principles of analytical sociology*. Cambridge: Cambridge University Press.

Hempel, Carl G., and Paul Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science* 15:135–175.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, Herbert Gintis and Richard McElreath. 2001. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review* 91:73–78.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr and Herbert Gintis. 2004. *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer and John Ziker. 2010. Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327:1480–84.

Homans, George C. 1967. *The nature of social science*. New York: Harcourt.

Kalter, Frank, and Clemens Kroneberg. 2014. Between mechanism talk and mechanism cult: New emphases in explanatory sociology and empirical research. In *Social Contexts and Social Mechanisms, Kölner Zeitschrift für Soziologie und Sozialpsychologie,* Special Issue 54, Ed. Jürgen Friedrichs, Alexandra Nonnenmacher, 91–115. Wiesbaden: Springer VS.

King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.

Korpi, Walter. 1980. Social policy and distributional conflict in the capitalist democracies. A preliminary comparative framework. *West European Politics* 3:296–316.

Korpi, Walter. 1983. *The democratic class struggle*. London: Routledge and Kegan Paul.

Korpi, Walter, and Joakim Palme. 1998. The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the western countries. *American Sociological Review* 63:661–687.

Kroneberg, Clemens, and Frank Kalter. 2012. Rational choice theory and empirical research. Methodological and theoretical contributions in Europe. *Annual Review of Sociology* 38:73–92.

Lange, Peter, and Hudson Meadwell. 1991. Typologies of democratic systems: From political inputs to political economy. In *New directions in comparative politics*, ed. Howard J. Wiarda, 82–117. Boulder, CO: Westview.

Laver, Michael. 1998. Models of government formation. *Annual Review of Political Science* 1:1–25

Lindenberg, Siegwart. 1992. The method of decreasing abstraction. In *Rational choice theory. Advocacy and critique*, eds. James S. Coleman and Thomas J. Fararo, 3–20. Newbury Park: Sage.

Lindenberg, Siegwart. 1996. Choice-centred versus subject-centred theories in the social sciences: The influence of simplification on explananda. *European Sociological Review* 12:147–157.

Lipset, Seymour Martin. 1959. Some social requisites of democracy: Economic development and political legitimacy. *The American Political Science Review* 53:69–105.

Mahoney, James, and Gary Goertz. 2006. A tale of two cultures: Contrasting quantitative and qualitative research. *Political Analysis* 14:227–249.

Manzo, Gianluca. 2014. Data, generative models, and mechanisms: More on the principles of analytical sociology. In *Analytical sociology: Actions and networks*, ed. Gianluca Manzo, 4–52. New York: John Wiley & Sons Inc.

Marx, Karl. 1859. *A contribution to the critique of political economy*. Moscow: Progress Publishers.

Messner, Steven F. 2012. Morality, markets, and the ASC: 2011 Presidential address to the American Society of Criminology. *Criminology* 50:5–25.

Nauck, Bernhard. 2007. Value of children and the framing of fertility: results from a cross-cultural comparative survey in 10 societies. *European Sociological Review* 23:615–29.

Nonnenmacher, Alexandra, and Jürgen Friedrichs. 2013. The missing link: Deficits of country-level studies. A review of 22 articles explaining life satisfaction. *Social Indicators Research* 110:1221–1244.

Opp, Karl-Dieter. 2013. What is analytical sociology? Strengths and weaknesses of a new sociological research program. *Social Science Information* 52:329–60.

Ostrom, Elinor. 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.

Ostrom, Elinor. 2005. *Understanding institutional diversity*. Princeton, NJ: Princeton University Press.

Ostrom, Elinor. 2010. Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review* 100:641–672.

Parsons, Talcott. 1964. Evolutionary universals in society. *American Sociological Review* 29:339–357.

Peters, B. Guy. 2013. *Strategies for comparative research in political science*. Basingstoke: Palgrave Macmillan.

Poteete, Amy R., Marco Janssen and Elinor Ostrom. 2010. *Working together: Collective action, the commons, and multiple methods in practice*. Princeton, NJ: Princeton University Press.

Putnam, Robert D. 2007. E Pluribus Unum: Diversity and community in the twenty-first century the 2006 Johan Skytte prize lecture. *Scandinavian Political Studies* 30:137–174.

Ragin, Charles C. 1987. *The comparative method. Moving beyond qualitative and quantitative strategies*. Berkeley: University of Calofornia Press.

Ragin, Charles C. 1991. The problem of balancing discourse on cases and variables in comparative social science. *International Journal of Comparative Sociology* 32:1–8.

Raub, Werner, and Thomas Voss. 2017. Micro-macro models in sociology: Antecedents of Coleman's diagram. In *Social Dilemmas, Institutions and the Evolution of Cooperation. Festschrift for Andreas Diekmann*, eds. Ben Jann and Wojtek Przepiorka, 11–36. Berlin: De Gruyter.

Schröder, Martin. 2019. Varieties of capitalism and welfare regime theories: Assumptions, accomplishments, and the need for different Methods. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00609-7.

Schwartz, Shalom H. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental Social Psychology 25*, ed. Mark P. Zanna, 1–65. New York: Academic Press.

Thelen, Kathleen. 2002. The explanatory power of historical institutionalism. In *Akteure—Mechanismen—Modelle: Zur Theoriefähigkeit makro-sozialer Analysen*, ed. Renate Mayntz, 91–107. Frankfurt a. M.: Max Planck Institute for the Study of Societies.

Tilly, Charles. 1984. *Big structures, large processes, Huge comparisons*. New York: Sage.

Turner, Jonathan H. 1991. Sociological theory. Diversity and disagreement. In *The structure of sociological theory*, ed. Jonathan Turner, 1–30. Belmont, CA: Wadsworth Publishing.

Weber, Max. 1958. *Essays in sociology*. New York: Oxford University Pres.

Wimmer, Andreas, and Nina Glick Schiller. 2002. Methodological nationalism and beyond: Nation-state building, migration and the social sciences. *Global Networks* 2:301–334.

Ylikoski, Petri. 2012. Micro, macro, and mechanisms. In *The Oxford handbook of philosophy of the social sciences*, ed. Harold Kincaid, 21–45. Oxford, UK: Oxford University Press.

Clemens Kroneberg 1980, Dr. rer. soc., Professor of Sociology, Institute of Sociology and Social Psychology, University of Cologne. Areas of research: Analytical sociology, migration and integration, crime and deviance, social networks, social boundary making. Recent publications: Ethnic Discrimination in the German Housing Market. A Field Experiment on the Underlying Mechanisms. Zeitschrift für Soziologie 47, 2018 (with A. Horr and C. Hunkler); Revisiting the Role of Self-Control in Situational Action Theory. European Journal of Criminology 15, 2018 (with S. Schulz); Institutional trust and the provision of public goods: When do individual costs matter? The case of recycling. Rationality and Society 29, 2017 (with S. Rompf and T. Schlösser).

ERRATUM

**Kölner Zeitschrift
für Soziologie und
Sozialpsychologie**

Erratum zu: Individual values, cultural embeddedness, and anti-immigration sentiments: explaining differences in the effect of values on attitudes toward immigration across Europe

Eldad Davidov · Bart Meuleman · Shalom H. Schwartz · Peter Schmidt

In dem Beitrag von Eldad Davidov, Bart Meuleman, Shalom H. Schwartz und Peter Schmidt: „Individual values, cultural embeddedness, and anti-immigration sentiments: Explaining differences in the effect of values on attitudes toward immigration across Europe" (*Kölner Zeitschrift für Soziologie und Sozialpsychologie* (2014) (Suppl) 66:263–285) wurde irrtümlicherweise der Name des Mitautors Bart Meuleman mit „Bart Meulemann" falsch abgedruckt. Er lautet korrekt Bart Meuleman. Ebenso lautet sein Geburtsdatum (S. 285) nicht „1965", sondern „1980". Die Redaktion bittet den Autor und die Leser um Nachsicht.

Die Onlineversion des Originalbeitrages ist erreichbar unter doi:10.1007/s11577-014-0274-5

E. Davidov (✉)
Institute of Sociology, University of Zurich,
Andreasstraße 15,
8050 Zurich, Switzerland
E-Mail: davidov@soziologie.uzh.ch

B. Meuleman
Centre for Sociological Research, University of Leuven,
Parkstraat 45,
3000 Leuven, Belgium

S. H. Schwartz
Department of Psychology, The Hebrew University of Jerusalem,
Mount Scopus,
Jerusalem 91905, Israel

P. Schmidt
Institut für Politikwissenschaft, Justus-Liebig-Universität Gießen,
Karl-Glöckner Str. 21 E,
35394 Giessen, Germany

⌂ Springer

Employment and its Institutional Contexts

Marcel Erlinghagen

Abstract Since the late 1980s, scholars have been trying to find out if there are certain families of nation states ("regimes") that share specific features and characteristics in terms of their institutional settings. With regard to employment, different typologies have gained particular importance throughout past decades and in recent years. All these approaches suggest regime-typical patterns of employment, unemployment, and their outcomes. What can be said, however, is that the lion's share of the related quantitative empirical work relies either on the analyses of aggregated macrodata, or on the interpretative comparison of country case studies. But the important question is whether studies relying not on aggregated data but rather on internationally comparable, harmonized survey data, will still find regime-typical employment patterns. The following paper concentrates on Europe and North America, and gives an overview of this multilevel research into possible impacts of formal institutions on (a) non-standard employment and employment structure; (b) labor market mobility, job stability, and unemployment; (c) wage inequality; and (d) self-perceived job insecurity. It turns out that the concept of unique national employment models seems to be more appropriate to cover the state of as well as changes in employment than regime typologies are.

Keywords Unemployment · Flexibility · Inequality · International comparison · Multilevel analyses

M. Erlinghagen (✉)
Institute of Sociology, University of Duisburg-Essen
Lotharstraße 63, 47057 Duisburg, Germany
E-Mail: marcel.erlinghagen@uni-due.de

🖄 Springer

Erwerbsarbeit und ihre institutionellen Kontexte

Zusammenfassung Seit Ende der 1980er Jahre gibt es eine Diskussion, ob unterschiedliche nationalstaatliche Familien (‚Regime') identifiziert werden können, deren Mitglieder bestimmte zentrale institutionelle Elemente und Eigenschaften teilen. In einigen dieser Regimetypologien steht die institutionelle Rahmung von Erwerbsarbeit und Arbeitslosigkeit im Zentrum. Es wird angenommen, dass in ähnlichen institutionellen Kontexten auch ähnliche Entwicklungen am Arbeitsmarkt vorzufinden sein müssten. Bei näherer Betrachtung stellt man fest, dass der überwiegende Teil der empirischen Forschung zu diesem Thema entweder auf der Analyse aggregierter Makrodaten oder aber auf einzelnen, interpretativ verknüpften Länderfallstudien beruht. Es stellt sich jedoch die wichtige Frage, ob auch Mehrebenenanalysen, die auf international vergleichbaren, harmonisierten Individualdaten beruhen, solche regimespezifischen Erwerbsmuster nachweisen können. Der vorliegende Aufsatz gibt einen Überblick über die Befunde solcher Mehrebenenanalysen und konzentriert sich auf Studien aus Europa und Nordamerika. Dabei stehen Zusammenhänge zwischen dem institutionellen Kontext auf der einen Seite und (a) atypischer Beschäftigung und Beschäftigtenstruktur, (b) Arbeitsmarktmobilität, Beschäftigungsstabilität und Arbeitslosigkeit, (c) Lohnungleichheit und (d) subjektiver Beschäftigungssicherheit auf der anderen Seite im Fokus. Es zeigt sich, dass das Konzept letztlich einzigartiger nationaler Beschäftigungsmodelle offenbar besser als Regimetypologien geeignet ist, die spezifische Situation am und die Entwicklung des Arbeitsmarkts zu erklären.

Schlüsselwörter Arbeitslosigkeit · Flexibilität · Ungleichheit · Internationaler Vergleich · Mehrebenenanlyse

"Actors do not behave or decide as atoms outside a social context, nor do they adhere slavishly to a script written for them by the particular intersection of social categories that they happen to occupy. Their attempts at purposive action are instead embedded in concrete, ongoing systems of social relations." (Granovetter 1985, p. 487)

1 Introduction

Following a well-established argument of Hirst and Thompson (1995), at least in modern industrialized societies, the nation state and its institutions still decisively affect the decisions and behavior of individuals as well as organizations in most areas of life, although globalization is increasingly shaping people's lives all around the world. Here, institutions mean not only formal laws and regulations, but also informal norms, habits, and customs. In this respect, today's institutions are the result of protracted historical processes consisting of slow incremental changes and adaptations. Against this background, and concentrating on Europe and North America, the following paper gives an overview of recent evidence as to whether, and if so how, nation-specific institutions have an impact on employment. We will concentrate exclusively on formal institutions, and possible impacts of cultural differences

on employment are not considered here, although it is quite clear that formal and informal institutions interrelate with each other (see for example Jansen 2017). In this respect, we only take studies into account that use advanced multilevel methods and that rely on internationally harmonized individual survey data. The paper then goes on to provide an overview of the state of research on possible impacts of formal institutions on (a) non-standard employment and employment structure; (b) labor market mobility, job stability, and unemployment; (c) wage inequality; and (d) self-perceived job insecurity as important features of labor market performance and development.

The paper proceeds as follows: Sect. 2 presents several general aspects of country-specific differences in employment by briefly introducing the idea of regime typologies as an institutional framework for labor markets (2.1.), and by determining the relationship between employment and the flexibility of firms (2.2). It then goes on to give a summary overview of the selected papers included in the following review, and to describe, classify, and discuss the prevailing operationalization of institutional differences in these studies (2.3). Sect. 3 sums up the results of international comparative multilevel analyses regarding non-standard employment and the related employment structure. The paper then presents evidence regarding the relationship between institutions and labor market mobility, job stability, and unemployment (Sect. 4), between institutions and wage inequality (Sect. 5), and between institutions and self-perceived job insecurity (Sect. 6). The paper concludes with a brief summary and implications for future research (Sect. 7).

2 Country-Specific Differences in Employment

2.1 Institutions and Regime Typologies

The modernization of society is inseparably combined with the rise of nation states whose borders clearly define their territories and therefore the spatial scope of the nation-specific legal institutions (see Anderson 1983). Since societies should not be misunderstood as "national containers" (Wimmer and Glick Schiller 2002), informal norms are, by definition, not limited by strict politically defined frontiers, although they are still (at least partly) affected by the nation state. Among other reasons, this is particularly because even in the era of globalization and transnationalization, standardized national educational systems, as well as large parts of the legal system, still form and hand down certain unified beliefs and traditions. These unified beliefs and traditions help to build and stabilize certain cultures by defining themselves as nation-specific and by overlaying regional traditions (Green 1997) as well as transnational ties.

Since the late 1980s, scholars have been trying to ascertain whether there are certain families of nation states that share specific features and characteristics of their institutional settings. Two such typologies have gained particular importance throughout recent decades and years with regard to employment, namely the "welfare state regime" typology (WSR) established by Esping-Andersen (1990) and the "varieties of capitalism" approach (VOC) developed by Hall and Soskice (2001).

With regard to employment, the two approaches vary in terms of their focus on the institutional context of employment. Whereas the WSR typology is interested in distributional impacts of social policy, the VOC approach is interested in the way in which institutions shape allocative decisions of firms (see also Schröder 2019). It is suggested in both cases that this leads to regime-typical patterns of employment, unemployment, and their outcomes.

Both the WSR as well as the VOC approach have been heavily discussed, improved, extended, and criticized since their introduction (for an overview regarding WSR see Emmenegger et al. 2015 and Ferragina and Seeleib-Kaiser 2011; regarding VOC see Streeck 2012a, and various contributions in Ebenau et al. 2015; for a brief overview of both see Schröder 2019). Thelen (2014) has recently promoted a joint perspective of allocative-oriented and distributive-oriented classification of institutions (see also Schröder 2009). Such a two-dimensional "varieties of liberalization" approach (VOL) seems to be more efficient and able to analyze the institutional impacts on societies because it better captures institutional and thus societal change. As a result, Thelen (2014) suggests three types of liberalization trajectories that different countries have followed over recent years and decades in order to tackle new challenges in an increasingly globalized world.

That said, the jury is still out as to whether regime typologies constitute an adequate categorization of complex realities. With regard to the welfare state, alternative approaches have been developed in order to accommodate important gender differences (e. g., Lewis 1992; Orloff 1993) or age-related differences (e. g., Jansen 2017). Qualifying the firm-centered VOC approach, some commentators stress the importance of exports and consumption in order to gain an understanding of the economic performance and development of national economies (e. g., Baccaro and Pontusson 2016). Another criticism of any kind of regime typology is that such approaches oversimplify historically, culturally, and politically unique trajectories of institutional development in nation states (Kasza 2002). It is argued that prominent regime typologies concentrate on a more or less arbitrary selection of context factors. In this respect, and with regard to the correlation between institutions and employment, Bosch et al. (2009) stress the singularity of national employment systems with their complex and complementary development of interrelationships between institutions, organizations, and individuals.

2.2 Labor Market Flexibility and Employment

Despite the debate about the relevance and validity of any regime typology, institutions are undoubtedly an important context-forming element of labor markets. This institutional framework builds a set of more or less stringent restrictions that influence individual employees' as well as firms' decisions or behavior. Since market conditions are constantly undergoing changes, workers as well as firms need to adapt steadily in line with such changes. Starting with labor demand, we can say that whether and to what extent such adaption takes place can generally be described as firms' flexibility. With regard to employment, firms have four different flexibility alternatives to adjust the input of labor to their specific needs (Atkinson 1984): (1) Either they can react by using the external labor market by hiring or dismissing

workers to suit their needs (external numerical flexibility), or they can react with internal forms of flexibility. Internal flexibility can be achieved by (2) variations in the working tasks (functional flexibility) and/or (3) working time (working time flexibility) of their staff (Kalleberg 2001). Whereas these three flexibility strategies aim to bring about flexibilization of work input, employers can (4) react to market fluctuations by adjusting wages (wage flexibility). Against this background, firms often follow a certain core–periphery strategy of labor utilization (Doeringer and Piore 1971), using and combining internal strategies particularly for their skilled and professional staff, and external strategies for their unskilled workers (Kalleberg 2003).

Which flexibilization strategy firms actually can achieve is, however, not only a matter of the demand that they exert, but is also influenced by the number, the composition, the needs, and the power of the workforce (labor supply), as well as by labor laws, labor market policies, industrial relations, and the respective social security system (institutional context). Thus, labor market outcomes are a common product of the decisions and behavior of firms (demand) and workers (supply), embedded in historically and culturally determined and changing nation-specific varieties of labor market-related institutions. In this respect, we may anticipate different country-specific patterns in certain labor market outcomes (e.g., developments in wages, the employment structure, unemployment, job stability and mobility, job security, etc.). Barbieri (2009) distinguishes between specific labor market regimes, whereby flexibilization in liberal labor markets (e.g., UK, Ireland, and the US) is brought about by producing wage inequalities between different age groups and cohorts. In contrast, corporate labor markets (e.g., Germany, Austria, and Belgium) achieve flexibility through skill-related job security differentiations. So-called flexicurity labor markets (e.g., Denmark) adapt by wage differentiations between certain skill groups, whereas Mediterranean labor markets (e.g., Spain, Italy, and France) achieve flexibility by age-related variations in job security levels (see also Kahn 2012).

There is a huge and still growing number of international comparative papers analyzing and discussing possible impacts of institutions on labor markets and employment which cannot be valuated and summarized in one single paper. Confirming regime-type-based hypotheses, such studies constantly find evidence of the theoretically supposed regime-type pattern of country-specific differences in varying outcomes (see the references in Sect. 2.1). What can be said, however, is that the lion's share of such existing quantitative empirical work relies either on the analyses of aggregated macrodata, or on an interpretative comparison of country case studies. It is thus possible that such relationships at least partly fail to reflect "real" cause–effect relationships, but rather reflect spurious correlations (Kittel 2006). This could be the case because such analyses do not usually control for country-specific compositions of the population (micro level) and/or also usually do not control for differences in the composition of firms or industries (meso level; Cartwright 2002; for a general conceptual overview regarding macro-meso-micro analyses and their problems, see Kroneberg 2019). There are only a few studies which take the important interdependencies between different societal levels into account by analyzing (a) internationally comparable, harmonized individual survey data (b) complemented by relevant

information about the organizational and institutional context using (c) advanced quantitative multilevel methods. Thus, the literature survey below will only include studies that meet these three criteria. In addition, studies are only included if they were published after 2006. Furthermore, we restrict the following literature survey to studies dealing with Europe and North America. Subject to these restrictions, the important question will be whether such studies relying on internationally comparable, harmonized survey data will still find regime-typical patterns, at least for the labor market outcomes mentioned above. In contrast, it is also possible that in the light of the presented findings, the concept of unique national employment models (see Bosch et al. 2009) is better suited to cover the state as well as the changes in employment compared to regime typologies—even if there is no doubt that there are always some important and remarkable similarities in certain institutional settings between a number of countries.

2.3 Regime Typologies and Internationally Comparative Multi-Level Research on Individual Employment Data

The main aim of internationally comparative multilevel research on employment is not only to identify individual characteristics (e. g., gender, education, age, etc.) of such employment-related outcomes, but also to learn more about the impact of labor market-related institutions and other context factors at the macro level on individual labor market behavior and performance at the micro level (Andreß et al. 2019). Against this background, regime typologies are of importance for this kind of research for several reasons. Although this importance varies, it can be said that such regime typologies are at least important for all internationally comparative research as a kind of frame of reference that helps to formulate hypotheses about certain expected institutionally shaped patterns of employment-related outcomes such as the individual risk of becoming unemployed or feelings of insecurity regarding job loss.

Table 1 provides an overview of the selected multilevel research that will be presented below in greater detail. All such multilevel research shares the opinion that the outcome of a lower level is (partly) determined by factors of an upper level. Thus, beside individual factors (such as skills or income), workers' behaviors and decisions are also (commonly) shaped by context factors (such as institutions or general economic situations) at a higher level (Blalock 1984). There are different approaches towards analyzing such multilevel aspects (DiPrete and Foristal 1994; Guo and Zhao 2000; Heisig et al. 2017; see also Meuleman 2019; Schmidt-Catran et al. 2019; Goerres et al. 2019). We find two different primary kinds of approaches when it comes to cross-country differences in employment: cluster-robust estimation (CRE) and random coefficient models (RCM). As in other research areas within sociology (see Heisig et al. 2017), multilevel research on employment is dominated by RCM: 16 out of the 24 papers that we reviewed use RCM, five papers use CRE, and the remaining three papers follow other alternative approaches (see Table 1).

Despite such differences in research methods, the papers vary in the way in which they try to capture and operationalize the theoretically assumed impact exerted on employment by labor market-related institutions. We find three kinds of strategies:

Table 1 Synopsis of the surveyed papers. Author's own work

	Authors (Year)	Data	Regime or country dummies (RD/CD)	Institutional macro indicators	Additional macro indicators	Method	Countries
1	Kahn (2007)	IALS 1994–1998	CD	EPL, IR	None	Cluster-robust multinomial logistic regression	7
2	Kahn (2010)	ECHP 1994–2001	None	EPL	UER	Panel regression	9/14
3	Gebel and Giesecke (2011)	LFS 1992–2007	None	EPL, LMP, IR	GDP, ES, G&T	Random coefficient regression	15
4	Muffels (2015)	LFS	RD	EPL, LMP, IR	None	Cluster-robust (?) logistic regression	10
5	Green and Livanos (2017)	LFS 2006, 2007, 2009 & 2010	RD	None	GDP, UER, ES	Probit regression with Heckman correction	10
6	Muffels and Luijkx (2008)	ECHP 1994–2001	RD	None	None	Cluster-robust multinomial logistic regression	14
7	Erlinghagen and Knuth (2010)	ESS 2004 & CID 2005	CD	None	None	Cluster-robust multinomial logistic regression	6
8	De Lange, Gesthuizen and Wolbers (2014)	ESS 2002–2008	None	EPL, ET	UER, G&T	Random coefficient regression	29
9	Wulfgramm and Fervers (2015)	EU-SILC 2005–2008	RD, CD	EPL, LMP, SSS, ET	UER, GDP	Random coefficient regression	27
10	Gebel and Giesecke (2016)	LFS 1992–2012	None	EPL, LMP, IR	GDP, ES	Random coefficient regression	19
11	van Winkle and Fasang (2017)	SHARE-Life	None	EPL, LMP	None	Random coefficient regression	13
12	Simón (2010)	ESES 2002	None	None	None	Pairwise comparisons	9
13	Mandel (2012)	LIS 1991–2000, DLS 1993, NLLS 1995	RD	EPL, FAM	ES	Random coefficient regression	21
14	Bachmann, Bechara and Schaffner (2016)	EU-SILC 2004–2011	CD	None	None	Cluster-robust (?) multinomial logistic regression	24

Table 1 (Continued)

	Authors (Year)	Data	Regime or country dummies (RD/CD)	Institutional macro indicators	Additional macro indicators	Method	Countries
15	Anderson and Pontussen (2007)	ISSP (1997)	None	EPL, LMP	UER, ER	Random coefficient regression	15
16	Erlinghagen (2008)	ESS (2004)	None	EPL, SSS	GDP, UER, ES	Random coefficient regression	17
17	Chung and van Oorshot (2011)	ESS (2008)	None	EPL, LMP	GDP, UER, ER	Random coefficient regression	22
18	Mau, Mewes and Schöneck (2012)	ESS (2008)	None	EPL, SSS	GDP, UER, G&T, INEQ	Random coefficient Regression	19
19	Dixon, Fullerton and Robertson (2012)	EB (2006)	RD	IR	UER, ES, INEQ	Random coefficient regression	27
20	Esser and Olson (2012)	ESS (2004)	None	EPL, IR	ES	Random coefficient regression	19
21	Lübke and Erlinghagen (2014)	ESS (2004 & 2010)	None	EPL, LMP	GDP, UER	Random coefficient regression	19
22	Stier (2015)	ISSP (2005)	None	EPL	UER, G&T	Random coefficient regression	28
23	Chung (2016)	ESS (2008)	None	LMP, IR	UER, ES, GDP	Random coefficient regression	23
24	Balz (2017)	ESS 2004 & 2010	None	EPL	UER	Random coefficient regression	17/22

CID Citizenship, Involvement and Democracy, *DLS* Danish Leisure Study, *EB* Eurobarometer, *ECHP* European Community Household Panel, *ESES* European Structure of Earnings Survey, *ESS* European Social Survey, *EU-SILC* European Union Statistics on Income and Living Conditions, *IALS* International Adult Literacy Survey, *ISSP* International Social Survey Program, *LFS* European Labour Force Survey, *LIS* Luxembourg Income Study, *NLLS* Norwegian Level of Living Survey, *SHARElife* Survey of Health, Ageing and Retirement in Europe (retrospective survey), *EPL* employment protection legislation, *LMP* active and/or passive labor market policy, *IR* industrial relations, *SSS* social security spending, *FAM* family–work compatibility policy, *ET* education and training policy, *UER* unemployment rate, *ER* employment rate, *GDP* gross domestic product, *ES* employment structure, *G&T* globalization and technology, *INEQ* inequality

(1) include regime-type dummies, (2) include country dummies, and (3) include macro indicators in the estimated models as independent variables. Some papers follow only one of these strategies, whereas others pursue different strategies estimating alternative models: Six papers include regime dummies, five include country dummies in their estimation, but the overwhelming majority (no fewer than 19 papers) refer to selected macro indicators as independent variables in their estimations.

Those papers that rely on regime dummies provide an obvious link to the differences that are assumed to exist between institutions and their impact on labor markets and individual careers, and are directly linked to the theoretically founded corresponding hypotheses. Including country dummies also implies the assumption that there are at least country-specific institutional "packages" that trigger a certain joint country-specific impact on individuals. This seems to be very different if macro indicators were used as independent variables. This approach seems to be highly flexible at first sight and open to different possible correlations between context and workers' decisions and behavior. It is, however, surprising against the background of this general openness that all the papers surveyed share a small set of common macro indicators as independent variables in their estimations, representing differences in the institutional setting.[1] By far the most popular of these indicators represents employment protection legislation (EPL). 17 of the 24 papers include EPL in their estimations. Also very prominent are indices representing (active and/or passive) labor market policy measures (LMP; nine papers) and industrial relations (IR; seven papers). In contrast, only a small minority of the surveyed papers include variables representing social security spending (SSS; three papers), or features of the educational and training system (ET; two papers). Only one paper actually attempts to capture country-specific differences in work–family compatibility (FAM).

One could believe at first sight that in a lot of these papers, which particularly rely on macro indicators as independent variables, regime typologies play a minor role, if any at all. On a closer inspection, however, it should become clear that the repeatedly selected indicators are still much more closely connected with those regime types—at least in an indirect (and maybe sometimes unconscious) manner. Table 2 visualizes these implicit ties. If we concentrate on EPL and LMP as the two most commonly used indicators, we have to say that both represent two important facets of Esping-Andersen's typology of welfare states. The aspect of whether and how well workers are legally protected against dismissals, or how easy it is to establish temporary employment contracts, aims at the stratification of chances and risks between labor market insiders and outsiders. And labor market policy measures stand for the degree of de-commodification, so that they target the second important dimension in welfare-state regime differentiations. In contrast, industrial relations (IR), as the last of the three most commonly included indicators, refers implicitly to one important facet of the VOC approach, given that it represents the way in which employer–employee relationships are organized and coordinated. Although the three other indicators play a much smaller role in internationally comparative research, we

---

[1] Beside such indicators representing the institutional setting, most papers also include further macro indicators representing other non-institutional aspects of relevant contexts, e. g., gross domestic product or the unemployment rate (see Table 1 for details).

Table 2 Expected target and regime-specific profile of selected labor market-related institutions. Author's own work

Institution	Target	Regime-specific profile
EPL	Stratification between insiders and outsiders	Low EPL in liberal & social democratic WR, high in conservative WR
ALMP	De-commodification	High in social democratic WR, low in liberal WR
PLMP	De-commodification	High in social democratic WR, low in liberal WR
IR	Coordination of employer-employee relationship	High in coordinated ME, low in liberal ME
SSS	De-commodification	High in social democratic WR, low in liberal WR
ET	Standardization and specialization of E&T	High in coordinated ME, low in liberal ME
FAM	Work–family compatibility	High in social democratic WR, low in conservative WR

EPL employment protection legislation, *ALMP* active labor market policy, *PLMP* passive labor market policy, *IR* industrial relations, *SSS* social security spending, *FAM* family–work compatibility policy, *ET* education and training policy, *WR* welfare state regime, *ME* market economy

can also assign them to crucial features, either in the WSR or in the VOC concept. Including the absolute or relative amount of country-specific social security spending (SSS) refers to Esping-Andersen's criteria of de-commodification, whereas the degree of standardization and specialization in the national educational and training system (ET) is an important feature in the VOC approach to distinguish between coordinated and liberal market economies. The degree of work–family compatibility (FAM) is connected with the WSR concept, if only because country-specific gender roles and family norms are entangled with de-commodification strategies and policies.

Having said that, it is still an open question as to why researchers only choose such a very limited set of macro indicators. One answer could be that EPL or LMP can easily be operationalized because internationally comparable index measures provided by the OECD (Organisation for Economic Co-operation and Development), the European Union, or other supranational organizations are available.[2] But this argument is not really convincing. These supranational organizations run large and easily accessible online databases with a very broad set of harmonized statistical data for a large number of different topics. This should make it quite easy to establish at least some alternative or additional measures as independent variables in employment-related multilevel research. It could rather be speculated that, in the end, concentrating on EPL, LMP, and IR reflects only the ongoing main-

---

[2] This concentration of a very small set of indicators often produced by single one data supplier (e. g., the OECD) can cause severe problems of which we need to be aware. As Roose (2013, p. 397) points out: "If empirical knowledge in a certain field is based on a small number of data sets, then recognized as well as unrecognized systematic errors will find their way into published evidence. This can result in a multiplication of errors" (own translation). The point is therefore not only that most researchers include, for example, any EPL indicator, but also that virtually all the studies relied on the measure published by the OECD.

stream debate about employment and labor markets. Starting in the 1990s, neo-liberal commentators have successfully carried out agenda setting by simply claiming that the de-regulation of employment protection, a reduction in labor market policy, and the liberalization of industrial relations was the only viable recipe for creating economic wealth in a more and more globalized world (see Scholte 2005). In this respect, neo-liberal (inspired or provoked) arguments implicitly share with regime-typology approaches the creed of distinct and unambitious effects of single institutions or at least (regime-specific) bundles of institutions on employment. This neo-liberal agenda setting seems to be very successful because related research responded to these arguments and toiled to find evidence either in support of or refuting such arguments. This could help explain the predominance of EPL, LMP, and IR as seemingly relevant context indicators in internationally comparative employment research. Leaving aside the question of what alternative context factors should be more closely focused on instead (this is discussed later in the concluding paragraph), we have to say that against this background, one should expect deregulation to lead to a clear and unambiguous pattern in all the fields of research surveyed—even if there are debates and conflicts regarding the nature of this pattern (e.g., increasing opportunities vs. risks for employees through deregulation). And since nation-specific deregulation modes are said to be congruent for countries belonging to the same regime family, the primary conceptual differences between explicit regime-type-driven analyses and macro indicator-driven analyses actually diminish in the end.

3 Non-Standard Employment and Employment Structure

Labor law generally allows different forms of employment contracts. Thus, despite nation-specific definitions, we can distinguish between standard employment relations (SER) and non-standard employment relations (NER; atypical employment). Whereas SER usually cover full-time and permanent employment in a subordinate employment relationship, NER mainly target part-time employment, temporary employment, and temporary agency work (ILO 2016). There is no doubt that NER have been on the rise in recent decades, at least in all industrialized societies (Kalleberg 2000; ILO 2016). Three of the main drivers of this development are globalization, the ongoing transformation to a service economy, and societal modernization. Due to globalization, firms have to adapt more quickly to market changes, and therefore tend to increase their demand for certain flexible employment arrangements in the form of NER. In addition, the growing service sector also needs more part-time employees working in non-standard working hours arrangements to better meet customers' needs. It is also suggested that NER is connected with societal modernization in the form of increasing gender equality. As a result, women's labor market participation is on the rise. However, this does not mean total gender equality, since women are still said to be responsible for childcare and domestic work. Therefore, the increasing employment participation of women is linked to an increase in NER because atypical working arrangements enable women to combine paid employment

and unpaid housework. These developments are accompanied by a general trend in the deregulation of labor laws that further fosters an increase in NER (Bosch 2004a).

Although empirical evidence is of course mixed to some degree (e. g., because of differences in data, period under study, or research methods), it seems to be quite clear that legal regulations, industrial relations, the tax system, and social policies indeed shape the nation-specific prevalence of NER (Hipp et al. 2015). However, as mentioned on a general basis above, such conclusions are mainly drawn by relying on the analyses of macrodata (e. g., Busemeyer and Thelen 2015; ILO 2016) or on a comparison of single results of country case studies (see a number of papers in Koch and Fritz 2013 as well as in Eichhorst and Marx 2015; see also Buchholz et al. 2009 for a brief summary of the comprehensive results of the "Globalife Project"). We must recognize that the analyses of such aggregated information largely dominate the international comparative research on NER. There is no doubt that we can learn much from this kind of research, but when it comes to disentangling complex interrelationships between institutions, organizations, social structure, and individual behavior, we had better be on the lookout for international comparative multilevel analyses of individual survey data, combined with sufficient contextual information on institutions and organizations. However, if we restrict our focus to empirical research that meets these standards, the number of adequate studies shrinks dramatically. Longitudinal multilevel analyses regarding the correlation between institutional change and developments in NER are even fewer and farther between (see Hipp et al. 2015 for an overview). Consequently, we have only fragmented evidence of possible similar patterns of the incidence and development of NER in countries belonging to the same type of institutional regime.

By using data of the European Labour Force Survey, Green and Livanos (2017) found evidence that strong employment protection legislation (EPL) is positively correlated with the individual propensity of non-standard employment on an involuntary basis. These results are largely analogous to those produced by Kahn (2007), who analyzed data for Canada, Finland, Italy, the Netherlands, Switzerland, the UK, and the US that were collected during the period from 1994 to 1998. In a dynamic perspective relying on data of the European Community Household Panel (ECHP), Kahn (2010) was able to show that deregulation of EPL does indeed increase non-standard employment, but that "there is no evidence that such reforms raise employment. Thus, these reforms, while touted as a way of jump-starting individuals' careers in the job market, appear rather to encourage a substitution of temporary for permanent work" (Kahn 2010, p. 14; see also Gebel and Giesecke 2011).

Muffels (2015) conducted a much broader analysis of the correlation between institutional differences and the prevalence of NER in Europe. Based on European Labour Force Survey data, he shows that active labor market policies can indeed increase the number of transitions from unemployment into employment, but that these jobs are often atypical jobs. Generous unemployment benefits lead to more temporary employment, while central wage bargaining reduces the prevalence of NER. If and how certain institutions correlate with the prevalence of NER is, however, particularly a matter of specific combinations of employment-related institutions. He concluded that "low EPL for permanent workers but strict ones for temporary workers combined with highly-centralized wage bargaining coordination [...] tends

to reduce the insider–outsider divide [...] and to increase the level of employment security in a country" (Muffels 2015, p. 318).

If and how institutions shape the employment structure is usually analyzed from an individual perspective. That said, we have to keep in mind that the individual incentives for joining an NER, as well as the individual outcomes of NER, can only be understood if we take private households into account as units of common production (Becker 1965). Since individual labor supply is not always a decision that is taken alone, but often a result of intra-household negotiations (Ott 1992), we have to consider the complex interrelationship between institutions, individual resources, and family conditions (see also Grunow 2019 and Hank and Steinbach 2019). However, during the search for relevant papers for our present literature review, we were not able to identify a publication addressing this complex interrelationship by conducting multilevel analysis methods. This came as a surprise, given that taking household conditions and family structures as meso-level indicators into account seems to be a necessary and promising enterprise to understand country-specific differences in the impact of labor market institutions on employment. Although it does not apply multilevel methods, the paper by Horemans (2016) stresses the importance of closing this research gap. Relying on EU-SILC data for dual-earner couples, he shows that labor market institutions in combination with culturally determined gender roles shape labor supply decisions on the household level. The effect of this is that social inequality is affected by NER to differing degrees in different countries, while these differences are not unambiguously reflected by standard institutional regime typologies.

4 Labor Market Mobility, Job Stability, and Unemployment

Even though labor market mobility, job stability, and unemployment address different aspects of labor market performance, they are nevertheless all interrelated. All three indicators reflect the development and volatility of labor markets as a whole, as well as flexibility strategies of employers and trajectories of workers' individual careers.

Labor turnover as a measure of the numbers of employees joining and leaving firms is an important indicator of the functionality of labor markets. Fluctuations have potential advantages for employers and employees alike, but they also have drawbacks. Dismissing workers in an economic crisis, as well as hiring them when they are needed, is one facet of employers' possible flexibilization strategies as they react to changes on their key markets. However, both hiring and firing induce costs, and employers should therefore be interested in limiting fluctuations. But it is also true that too low worker turnover could endanger firms' potential for innovation, since new staff from outside the company can introduce new ideas and new knowledge, and thus help enhance productivity (Farber 1999). Employees, however, prefer long job tenure guarantees, coupled with income stability and the possibility of internal promotion. On the other hand, switching to another employer can offer new career opportunities (Ferreira 2016). Thus, from a theoretical perspective, both employers and employees should be interested in some kind of balanced turnover.

Obviously (and in addition to the general theory of labor turnover described above), institutions directly matter for labor market fluctuations because they form the legal framework for hiring, dismissals, or resignations. Institutions are also indirectly relevant because they set the economic and social context in which labor market participants have to make their decisions. The educational system and the related skill structure (see also Blossfeld et al. 2019), industrial relations, work organization, and labor law come together to build a framework of incentives that should lead to a specific pattern of labor market mobility. In addition, employment fluctuations and employment stability should vary, particularly with regard to age, gender, and education, but also with respect to specific job characteristics (industry, type of contract, etc.). Thus, the composition of the labor force, along with these characteristics, is of decisive importance when it comes to explaining between-country variations in labor market mobility. Finally, labor market mobility should be strongly influenced by economic cycles, whereas fluctuations should increase during upturns and they should decrease during downturns. This pro-cyclical development is a result of an increasing number of job opportunities during economic recovery that leads to more job-to-job moves as well as to increasing mobility from unemployment into employment. Job opportunities decline in an economic crisis and as a result, labor market mobility falls too (Schettkat 1996).

Although there is no lack of theoretical assumptions about the connection between institutions and labor market mobility, empirical cross-country comparisons are rare and concentrate primarily on the impact of EPL on job tenure (Dale-Olsen 2016). To the best of our knowledge, there are only two cross-country studies that apply multilevel methods to analyze individual survey data. The first is the study by Muffels and Luijkx (2008). They do not find clear mobility patterns for countries belonging to specific welfare or production regimes. If at all, we can see lower job stability in liberal labor market regimes (e.g., UK, Ireland, but also Denmark) and in transformation countries in Central and Eastern Europe (CEE) compared to all other countries in Europe (Muffels and Luijkx 2008). The second is a recent study by van Winkle and Fasang (2017). They analyze the employment trajectories of different birth cohorts in 14 European countries based on SHARElife data. First, it turns out that careers all over Europe have not become more dynamic, unstable, or insecure during the course of time, confirming findings of previous studies (e.g., Auer and Cazes 2000; Doogan 2001; Erlinghagen and Knuth 2004). Second, the results show that "the complexity of employment trajectories varies to a much greater extent across countries compared to change over birth cohorts" (van Winkle and Fasang 2017, p. 23). Against this background, the study was also able to show that stronger EPL decreases involuntary labor market mobility and passive labor market policy in the form of wage replacement rates increases voluntary job mobility and therefore the quality of job matches. It seems as if two contradictory forces were at work: De-commodification in the form of strong EPL lowers overall labor market mobility, while high wage replacement rates—also representing de-commodification—increases overall labor market mobility at the same time. As a result, we were able to suggest that country differences in overall labor market mobility should not be primarily a result of de-commodification differences. We can in fact hypothesize that despite regime types represented by EPL or LMP, cultural differences or

differences in the educational and training system are much more important for understanding cross-country variations in job stability or labor turnover.

With regard to unemployment, there seems to be a much clearer pattern of institutionally driven cross-country differences in labor market performance. Thus, liberal regime types with weak EPL and lower wage rigidity are more sensitive to cyclical influences resulting in higher fluctuations in and out of unemployment, as well as accounting for a significantly lower share of long-term unemployment. With regard to such general results, it is often suggested that low-skilled workers suffer most from strict EPL regulations. From a theoretical perspective, there seems thus to be a trade-off between low unemployment versus high employment protection, at least for some groups of workers (Bertola 1999; OECD 1999). Gebel and Giesecke (2011) were, however, able to show that such correlations at the macro-level are misleading. Using data from the European Labour Force Survey, they find nothing to support such a causal effect. Instead, they present evidence that lowering EPL leads employers to substitute regular employment with temporary employment, and not to increase the number of people they hire to an extent that would help to significantly reduce unemployment (see also Olsthoorn 2016; see Gebel and Giesecke 2016 for similar results regarding youth (un)employment). In addition, Wulfgramm and Fervers (2015) show that on the one hand, active labor market policy (ALMP) as well as passive labor market policy (PLMP) prolonged the re-employment periods of unemployed workers but that on the other hand, both policies significantly decrease the risk of becoming unemployed once more. Thus, the authors conclude "that cutting back the welfare state is not an easy way out of its crisis, since lower spending for ALMP and less generous unemployment benefits might go along with unstable re-employment. Nothing is gained if workfare and activation policy-induced reductions in unemployment duration are paid for by unstable re-employment and higher risk of future unemployment" (Wulfgramm and Fervers 2015).

Furthermore, beside such general conclusions, different effects and interrelations between institutions and specific social subgroups have to be taken into account. De Lange et al. (2014) stress not only that younger people particularly suffer from high EPL, but that the educational system is of decisive importance with regard to youth unemployment (see also Blossfeld et al. 2019). Relying on data from the European Social Survey (ESS), they point out: "[A]s vocational education is more specific, young people are less often in temporary employment and unemployment. In general, the link between the knowledge and skills acquired through education and its benefits in the labour market is stronger when the educational system is vocationally specific" (De Lange et al. 2014, p. 209).

One severe problem of most cross-country comparisons of unemployment incidence and rates is, however, that they do not take into account that there are possible alternative non-employment statuses (health-related incapacity for work, retirement, labor reserve, etc.), and therefore over- or underestimate the positive or negative effects of institutions on labor market integration (see as a recent example Biegert 2017). Erlinghagen and Knuth (2010) show certain country-specific pathways to deal particularly with less-favored unemployed people (the unskilled, those with health problems, older workers, etc.). Thus, lower unemployment rates and higher flows out of unemployment in more liberal labor market regimes are at least partly not an

effect of lower EPL, but instead a result of easier transitions into alternative non-employment states such as retirement or especially health-related incapacity. This transition trajectory does not, however, follow a clear-cut regime-type or deregulation pattern.

We therefore see that the connection between welfare state regimes or production regimes on the one hand, and unemployment on the other, is not as predetermined as is often suggested. Again, different historical experiences combined with certain culturally shaped beliefs and attitudes have to be taken into account if we wish to understand country-specific differences in the degree and incidence of unemployment as a social status. Further, we have good reasons to pay more attention to family relations and related institutions, as well as to variations and specific features of the educational and training system, if we seek to gain a better understanding of cross-country differences in unemployment.

5 Wage Inequality

Income consists of three different elements: (a) earned income (wages and salaries), unearned income (income on investments), and (c) transfer income (subsidies and inheritances). Income inequality is thus a product of inequality in at least one of the three individual sources of income. Analyzing and understanding income inequalities is a very challenging enterprise because individual characteristics and institutional constraints coincide and have to be understood in terms of their joint impact on all three income elements of individual or household income.

Since wages as earned income are most directly connected with employment, we will mainly concentrate below on international comparative empirical research on wage inequality. The latter has risen in almost all industrialized countries in recent decades. Some commentators argue that this is mainly because of skill-biased technological change combined with an increase in international trade (e. g., Afonso et al. 2013). Others cast doubt on this "technology-change argument," and say that the growth of wage inequality is at least also, if not mainly, driven by deregulation as well as by a loss of importance attached to collective bargaining and the shrinking power of the trade unions (e. g., Mishel et al. 2014; Stiglitz 2015). And since all industrialized countries are affected by technological change and globalization, these aspects could in any case not explain rising cross-country differences in the development of wage inequalities (e. g., Bachmann et al. 2016). Instead, these international differences could be better explained by differences in the population composition, by differences in nation-specific workplace conditions and industry structures, and by overall institutional disparities. But again, most empirical studies investigating the extent of and reasons for international differences in wage inequality are based on aggregated macrodata. Analyses employing individual microdata or information about differences on the meso-level (e. g., firms, industries, workplace arrangements) are very scarce.

Relying on aggregated data, there is ample evidence that labor market institutions and policies (such as EPL, wage-setting institutions, union density, or minimum wage regulations) explain a major part of observable cross-country differences in

wage inequality (Wallerstein 1999; Koeniger et al. 2007; Lemieux 2008; VanHeuvelen 2018). This work has been recently complemented by analyses of internationally comparable microdata. Based on longitudinal EU-SILC data, Bachmann et al. (2016) show that cross-country differences in wage inequality are partly driven by diverging employment structures, particularly regarding skills, age, and gender, resulting in different wage mobilities. In addition to such individual determinants and relying on harmonized matched employer–employee data from nine European countries, Simón (2010) emphasizes the importance attached to workplace and job characteristics in explaining cross-country differences in wage inequality. The situation is complicated by the fact that overall wage inequalities are also effected by culturally as well as politically driven gender pay differences (Mandel 2012) in combination with certain taxation and redistribution policies (Guvenen et al. 2014; DiPrete 2005). Again, the scarce empirical evidence shows no clear regime-type pattern but does stress the importance of unique country-specific combinations of institutions and population composition in helping us to understand the extent and development of wage inequality.

6 Self-Perceived Job Insecurity

There is a long tradition in analyzing possible negative psychological impacts of unemployment (e.g., Jahoda et al. 1933) that does not only affect the objective living standard, but that also (and sometimes even more importantly) causes negative stress with manifold negative consequences for health, family life, etc. Against the background of an increase in non-standard employment, the question starts to arise as to whether, and if so how, employees process the increasing flexibilization demands made by their employers. Thus, during the past ten years, researchers have become more and more interested in the question of how employees evaluate their own employment situation in terms of self-perceived job insecurity (SJI).

In an international comparative perspective, there are considerable differences in the share of employees feeling insecure about the stability of their jobs (Erlinghagen 2008; Lübke and Erlinghagen 2014; Balz 2017). With regard to the possible impacts of the institutional context on self-perceived job insecurity, there ought to be very clear correlations from a theoretical perspective: SJI should particularly increase where employment protection legislation (EPL) is lower, where labor market policy (LMP) is reduced, and where the unions are weaker.

A joint review of the existing multilevel analyses regarding self-perceived job insecurity is complicated by the fact that job insecurity is a multi-dimensional phenomenon. Anderson and Pontusson (2007) distinguish between (a) cognitive job insecurity, (b) labor market insecurity, and (c) affective job insecurity. Whereas "cognitive job insecurity" describes the (neutral) awareness of possible job loss, "labor market insecurity" seeks to subjectively evaluate future re-employment prospects, and "affective job insecurity" refers to employees' feelings of concern and fear about any future job loss (see also Chung and Mau 2014).

With regard to Europe, the most recent analyses rely heavily on data from the European Social Survey (ESS), and show no significant correlation between EPL

and cognitive job insecurity (Erlinghagen 2008; Chung and van Oorschot 2011; Mau et al. 2012; Lübke and Erlinghagen 2014) or between EPL and labor market insecurity (Lübke and Erlinghagen 2014) controlling for the composition of population, relevant differences in firm and industry structure, as well as for relevant indicators representing the current economic situation in general. In contrast, Anderson and Pontusson (2007) find a correlation between EPL and affective job insecurity when analyzing data from the International Social Survey Program (ISSP) covering not only nine European countries but also Canada, Japan, New Zeeland, and the USA.

Overall social spending (Erlinghagen 2008), as well as labor market policies (Anderson and Pontusson 2007; Chung and van Oorschot 2011; van Oorschot and Chung 2015), seem to have no significant impact on cognitive job insecurity after controlling for population composition and other context-related factors. However, Lübke and Erlinghagen (2014) show that there seems to be a delay in the effect of changes in LMP on SJI. As in previous studies, they also find that there is no significant correlation between the amount of active labor market policy and SJI in the specific survey year. Taking institutional changes into account, however, reveals that a year-on-year increase in ALMP expenditure decreases cognitive job insecurity. It seems as if employees recognize ALMP as some kind of protective shield against the negative consequences of redundancy. Conversely, it was also shown that an increase in ALMP during the previous three years increases the perceived difficulties of future job searches (labor market insecurity). Finally, it emerges that these seemingly contradictory findings can be ascribed to different skill groups: "If we take a more sophisticated look at the interaction between job search difficulties and the increase in ALMP expenditure, [...] it turns out that this correlation is less strong for people with primary education [...]. This suggests that the target group of ALMP indeed perceives these interventions as promoting security" (Lübke and Erlinghagen 2014, p. 331).

Beside EPL and labor market policy, we can also hypothesize that the kind of industrial relations should frame the subjective perception of employees' employment situation. It can be suggested that SJI should increase with weaker unions, since unions' strength should deliver additional protection against redundancy. Against this background and relying on data from the Eurobarometer, Dixon et al. (2013) find the suggested negative correlation between union density and cognitive job insecurity, but no significant result could be found with regard to labor market insecurity (see also Esser and Olson 2012). Anderson and Pontusson (2007) were able to show that union members show lower cognitive job insecurity but report higher labor market insecurity, which means that they are more worried about their future job prospects.

Depending on the country-specific institutional setting, EPL and the characteristics of industrial relations have greater or lesser importance for different groups of workers. Especially in countries with stronger unions (Chung 2016) and strong EPL (Balz 2017), temporary employees feel much more insecure compared to permanent workers. In addition, Lübke and Erlinghagen (2014) provide evidence that it is not only the current institutional context, but rather the change of institutional context through previous years (e. g., because of policy reforms) that affects SJI with a certain delay. They were able to show that an increase in EPL within the previous

five years lowers the cognitive job insecurity of permanent workers to a significantly greater extent than it does for temporary workers. In addition, workers with secondary education seem to react more sensitively to a fall in EPL, whereas low-educated as well as highly skilled employees react with a significantly lower increase of SJI to a decrease in EPL.

All in all, the evidence regarding the correlation between institutional context and SJI is somewhat mixed. But all evaluated multilevel analyses come to the conclusion that EPL and LMP as well as industrial relations have an impact on SJI, but that employees' awareness or fear of possible job losses depends even more on country-specific economic development (e. g., GDP), as well as on population and industry composition: "Economic and labour market conditions have been shown to be some of the most influential factors explaining the cross-national variation in job and employment insecurity" (Chung 2016, p. 8; see also Stier 2015). Based on the multilevel research that has been presented, we can conclude that regime concepts do not really help to understand cross-national differences in employment security. Again, it seems much more relevant to take the nation-specific economic situation as well as structural differences of employees as well as employers into account. In the end, only these unique and culturally as well as historically formed country-specific conditions help us to understand how individuals perceive their own employment situation and career prospects, and how they interpret institutional changes such as the deregulation of labor markets.

7 Summary and Conclusion

Theoretical concepts of "varieties of capitalism" (Hall and Soscice 2001) or different "worlds of welfare states" (Esping-Andersen 1990), as well as a recently developed idea of "varieties of liberalization" (Thelen 2014) are striking at first glance: Long-term historical and cultural differences between groups of similar nation states have formed certain formal and informal sets of institutions shaping both employers' and employees' beliefs and behavior in a typical manner. As a result, characteristic patterns in the employment structure, in labor market performance, in wage inequalities, and in self-perceived job insecurity should be observed that are suggested to represent certain employment regimes. As long as empirical research relies mainly on aggregated macrodata, it is indeed possible to constantly observe such patterns, and there is absolutely no doubt that such research provides important insight into cross-country differences in labor market developments and related inequalities. However, this kind of work does often not account for structural differences at the individual and organizational level. Thus, country-specific compositions of the workforce as well as diverging industry structures are commonly neglected. Analyzing internationally comparable and standardized microdata could help overcome these limitations. Since such data are very scarce, the number of this kind of study is limited.

With regard to non-standard employment, it turns out that we have to take into account decisions at the household level (household labor supply) in conjunction with certain gender norms if we wish to understand cross-country differences in

the prevalence and outcomes of atypical employment. To better understand labor market mobility patterns, we must instead concentrate much more closely on the demographic composition of the workforce, whereas international differences in unemployment should be understood as a result of a complex interplay of cyclical influences and institutions not directly connected with the labor market (e. g., educational system, social policies, etc.). The same holds true for international differences in self-perceived job insecurity. Turning to cross-country differences in wage inequalities, the studies presented stress the importance attached to differences in industry structures and workplace conditions.

Having said all this, we have to recognize that recent multilevel analyses relying on comparable international data that analyze cross-national differences regarding employment structure, labor market mobility, wage inequality, or job insecurity cast doubt on common regime classifications. Even if we take into account that regime typologies are oriented towards theoretically constructed ideal types, the empirical results do not really fit these assumptions. This also holds true if such analyses concentrate on a small set of macro indicators reflecting the institutional context instead of regime type or country dummies.

In fact, we find some kind of blurred and inconsistent evidence. Thus, labor market developments are a joint product of a complex and probably unique interplay between a broad set of historically and culturally shaped formal institutions as well as informal norms and values (macro level), certain industry structures and workplace conditions (meso level), as well as the composition of the workforce (e. g., gender, age, skills, etc.; micro level). All this is further complicated by the fact that all those levels are embedded in economic, social, and political changes. As a consequence, we have to suggest that the idea of common regime-type patterns or related unequivocal deregulation patterns is indeed very plausible and "catchy," but at least probably does not sufficiently describe or explain cross-country differences in employment (see also Kasza 2002).

That does not mean that certain similarities in labor market-related institutions or deregulation trajectories are inauthentic. We can indeed learn from research trying to identify some common historical and cultural roots of nation-specific institutions in different countries and how these roots are still visible and vivid today. What we can learn from the presented results of multilevel analyses of harmonized survey data above is, however, to sometimes be a little bit more cautious and self-effacing when it comes to uncritically reproducing mainstream wisdoms. Instead, we should be braver in scrutinizing alleged certainties (see van Kersbergen and Vis 2015)—and should encourage younger researchers to do so, even though we have to confess that this strategy could be dangerous for their career prospects. In this respect and following Bosch et al. (2009, p. 3), who suggest that regime typologies "are ultimately heuristic instruments and abstractions from national specificities," we should avoid oversimplified approaches. This also holds true for any kind of simple deregulation-related hypothesis. It is both tempting and dangerous at the same time to put forward, for example, broad regime similarities as a justification for an a priori selection of certain country cases in comparative research, or a simple interpretation of empirical evidence. Future research should improve by taking the following suggestions into account:

1. We need much more research that conducts multilevel analyses of internationally harmonized surveys, and that enables us to analyze the impact of labor market-related institutions and contexts on employment, whilst controlling for differences in population composition (e. g., education, age structure) as well as for differences at the firm level (e. g., firm size, industry).

2. Furthermore, we urgently need multilevel analyses including a much broader set of macro indicators. Instead of including EPL, LMP, or IR "by default," more research is needed that is based on theoretically founded but also creative use of alternative measures of relevant institutional differences. It can be suggested that particularly differences in family-related institutions and differences regarding the system of education and training should be taken into account to a much greater extent (e. g., Streeck 2012b; Bosch 2017). This means in practical terms that future multilevel research should particularly include indices that reflect household- or family-related issues (e. g., maternal employment rates, proportion of children in external childcare services, or paid parental leave, all provided by the OECD), as well as indices that reflect the educational system (e. g., youth unemployment rate or inactivity rate, provided by EUROSTAT).

3. Such future research should pay much more attention to previous institutional changes instead of searching for ad hoc correlations of context and employment outcomes in a single survey year (see Lübke and Erlinghagen 2014). In this context, we also need more creative ideas to map changes in the quality, importance, and meanings within institutions which have undergone (almost) no change in formal terms. As an example, traditional collective bargaining institutions today can definitely concentrate and stress different topics than in the past (e. g., Bosch 2004b; Marginson and Welz 2015).

4. We should not only concentrate on formal institutions but should draw much more attention to cultural differences in the shape of informal institutions such as norms or beliefs (see Jansen 2017). This also reminds us that future analyses should take into account that social change may alter workers' as well as employers' behavior and decisions in the course of time, although institutional settings remain unchanged (as an example see Herzog-Stein et al. (2018) on the different importance attached to temporary reductions in working hours during a recession).

5. Finally, we need broader ("representative") and internationally harmonized studies on the interrelation between macro-level institutions and the "behavior" of meso-level organizations (e. g., firms, trade unions, etc.).

This could help us to go beyond simple correlations of selected aggregated data and should improve our understanding of the complex functioning and development of national employment systems.

References

Afonso, Oscar, Ana L. Albuquerque and Alexandre Almeida. 2013. Wage inequality determinants in European Union countries. *Applied Economics Letters* 20:1170–1173.

Anderson, Benedict R. O. 1983. *Imagined communities: Reflections on the origin and spread of nationalism*. London: Verso.

Anderson, Christopher J., and Jonas Pontusson. 2007. Workers, worries and welfare states: social protection and job insecurity in 15 OECD countries. *European Journal of Political Research* 46:211–235.

Andreß, Hans-Jürgen, Detlef Fetchenhauer and Heiner Meulemann. Eds. 2019. *Cross-national comparative research—analytical strategies, results, and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00594-x.

Atkinson, John. 1984. Manpower strategies for flexible organisation. *Personal Management* 15:28–31.

Auer, Peter, and Sandrine Cazes. 2000. The resilience of the long-term employment relationship: Evidence from industrialized countries. *International Labour Review* 139:379–408.

Baccaro, Lucio, and Jonas Pontusson. 2016. Rethinking comparative political economy. *Politics and Society* 44:175–207.

Bachmann, Ronald, Peggy Bechara and Sandra Schaffner. 2016. Wage inequality and wage mobility in Europe. *Review of Income and Wealth* 62:181–197.

Balz, Anne. 2017. Cross-national variations in the security gap: Perceived job insecurity among temporary and permanent employees and employment protection legislation. *European Sociological Review* 33:675–692.

Barbieri, Paolo. 2009. Flexible employment and inequality in Europe. *European Sociological Review* 25:621–628.

Becker, Gary S. 1965. A theory of the allocation of time. *The Economic Journal.* 75:493–517.

Bertola, Guiseppe. 1999. Microeconomic perspectives on aggregate labor markets. In *Handbook of Labor Economics 3b,* eds. Orley Ashenfelter and David Card, 2986–3028. North Holland: Elsevier.

Biegert, Thomas. 2017. Unemployment in affluent democracies: The moderating role of the institutional insider/outsider divide. *American Sociological Review* 82:1037–1064.

Blalock, Hubert M. 1984. Contextual-effects models: Theoretical and methodological issues. *Annual Review of Sociology* 10:353–372.

Blossfeld, Hans-Peter, Nevena Kulic, Jan Skopek, Moris Triventi, Elina Kilpi-Jakonen, Daniela Vono de Vilhena and Sandra Buchholz. 2019. Conditions and consequences of unequal educational opportunities in the life course: Results from the Cross-National Comparative eduLIFE Project. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00595-w.

Bosch, Gerhard. 2004a. Towards a new standard employment relationship in western Europe. *British Journal of Industrial Relations* 42:617–636.

Bosch, Gerhard. 2004b. The changing nature of collective bargaining in Germany: Coordinated decentralization. In *The new structure of labor relations: Tripartism and decentralization,* eds. Harry C. Katz, Wonduck Lee and Joohee Lee, 84–118. Ithaca/New York: Cornell University Press.

Bosch, Gerhard. 2017. Different national skill systems. In *The Oxford handbook of skills and training*, eds. Chris Warhurst, Ken Mayhew, David Finegold and John Buchanan, 424–443. Oxford: University Press.

Bosch, Gerhard, Steffen Lehndorff and Jill Rubery. 2009. European employment models in flux: Pressures for change and prospects for survival and revitalization. *In European employment models in flux: A comparison of institutional change in nine European countries,* eds. Gerhard Bosch, Steffen Lehndorff and Jill Rubery, 1–56. Basingstoke: Palgrave Macmillan.

Buchholz, Sandra, Dirk Hofäcker, Melinda Mills, Hans-Peter Blossfeld, Karin Kurz and Heather Hofmeister. 2009. Life courses in the globalization process: The development of social inequalities in modern societies. *European Sociological Review* 25:53–71.

Busemeyer, Marius R., and Kathleen Thelen. 2015. Non-standard employment and systems of skill formation in European countries. In *Non-standard employment in post-industrial labour markets. An occupational perspective,* eds. Werner Eichhorst and Paul Marx, 401–429. Cheltenham: Edward Elgar.

Cartwright, Nancy. 2002. The limits of causal order, from economics to physics. In *Fact and fiction in economics: Models, realism, and social construction,* ed. Uskali Mäki, 137–151. Cambridge: Cambridge University Press.

Chung, Heejung. 2016. Dualization and subjective employment insecurity: Explaining the subjective employment insecurity divide between permanent and temporary workers across 23 European countries. *Economic and Industrial Democracy*: Online first.

Chung, Heejung, and Steffen Mau. 2014. Subjective insecurity and the role of institutions. *Journal of European Social Policy* 24:303–318.

Chung, Heejung, and Wim van Oorschot. 2011. Institutions versus market forces: Explaining the employment insecurity of European individuals during (the beginning of) the financial crisis. *Journal of European Social Policy* 21:287–301.

Dale-Olsen, Harald. 2016. Employee turnover. In *Research handbook on employee turnover,* eds. Georg Saridakis and Cary L. Cooper, 5–25. Cheltenham: Edward Elgar.

De Lange, Marloes, Maurice Gesthuizen and Maarten H. J. Wolbers. 2014. Youth labour market integration across Europe. *European Societies* 16:194–212.

DiPrete, Thomas A. 2005. Labor markets, inequality, and change—A European perspective. *Work and Occupations* 32:119–139.

DiPrete, Thomas A., and Jerry D. Forristal. 1994. Multilevel models: Methods and substance. *Annual Review of Sociology* 20:331–357.

Dixon, Jeffrey C., Andrew S. Fullerton and Dwanna L. Robertson. 2013. Cross-National differences in workers' perceived job, labour market, and employment insecurity in Europe. Empirical tests and theoretical extensions. *European Sociological Review* 29:1053–1067.

Doeringer, Peter B. and Michael J. Piore. 1971. *Internal labour markets and manpower analysis*. Lexington: Heath.

Doogan, Kevin. 2001. Insecurity and long-term employment. *Work, Employment and Society* 15:419–441.

Ebenau, Matthias, Ian Bruff and Christian May. 2015. *New directions in comparative capitalisms research*. London: Palgrave Macmillan.

Eichhorst, Werner, and Paul Marx. 2015. *Non-standard employment in post-industrial labour markets. An occupational perspective.* Cheltenham: Edward Elgar.

Emmenegger, Patrick, Jon Kvist, Paul Marx and Klaus Petersen. 2015. Three worlds of welfare capitalism: The making of a classic. *Journal of European Social Policy* 25:3–13.

Erlinghagen, Marcel. 2008. Self-perceived job insecurity and social context. A multi-level analysis of 17 European countries. *European Sociological Review* 24:183–197.

Erlinghagen, Marcel, and Matthias Knuth. 2004. In search of turbulence. Labour market mobility and job stability in Germany. *European Societies* 6:49–70.

Erlinghagen, Marcel, and Matthias Knuth. 2010. Unemployment as an institutional construct? Structural differences in non-employment in selected European countries and the United States. *Journal of Social Policy* 39:71–94.

Esping-Andersen, Gosta. 1990. *The three worlds of welfare capitalism*. Princeton: Princeton University Press.

Esser, Ingrid, and Karen M. Olsen. 2012. Perceived job quality: autonomy and job security within a multi-level framework. *European Sociological Review* 28:443–454.

Farber, Henry S. 1999. Mobility and stability: The dynamics of job change in labor markets. In *Handbook of Labor Economics 3b*, eds. Orley Ashenfelter and David Card, 2439–2483. North Holland: Elsevier.

Ferragina, Emanuele, and Martin Seeleib-Kaiser. 2011. Thematic review: Welfare regime debate: Past, present, futures? *Policy and Politics* 39:583–611.

Ferreira, Priscila. 2016. Employment, turnover and career progress. In *Research handbook on employee turnover*, eds. Georg Saridakis und Cary L. Cooper, 274–304. Cheltenham: Edward Elgar.

Gebel, Michael, and Johannes Giesecke. 2011. Labor market flexibility and inequality: The changing skill-based temporary employment and unemployment risks in Europe. *Social Forces* 90:17–40.

Gebel, Michael, and Johannes Giesecke. 2016. Does deregulation help? The impact of employment protection reforms on youths' unemployment and temporary employment risks in Europe. *European Sociological Review* 32:486–500.

Goerres, Achim, Markus B. Siewert and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Granovetter, Mark S. 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91:481–510.

Green, Andy. 1997. *Education, globalization and the nation state*. Houndsmills: Palgrave Macmillan.

Green, Anne, and Ilias Livanos. 2017. Involuntary non-standard employment in Europe. *European Urban and Regional Studies* 24:175–192.

Grunow, Daniela. 2019. Comparative analyses of housework and its relation to paid work: Institutional contexts and individual agency. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jür-

gen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00601-1.

Guo, Guang, and Hongxin Zhao. 2000. Multilevel modeling for binary data. *Annual Review of Sociology* 26:441–462.

Guvenen, Fatih, Kuruscu Burhanettin and Serdar Ozkan. 2014. Taxation of human capital and wage inequality: A cross-country analysis. *Review of Economic Studies* 81:818–850.

Hall, Peter A., and David Soskice. 2001. An introduction to varieties of capitalism. In *Varieties of capitalism: The institutional foundations of comparative advantage*, eds. Peter A. Hall and David Soskice, 1–68. Oxford: Oxford University Press.

Hank, Karsten, and Anja Steinbach. 2019. Families and their institutional contexts: The role of family policies and legal regulations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00603-z.

Heisig, Jan Paul, Merlin Schaeffer and Johannes Giesecke. 2017. The costs of simplicity: Why multilevel models may benefit from accounting for cross-cluster differences in the effects of controls. *American Sociological Review* 82:796–827.

Herzog-Stein, Alexander, Fabian Lindner and Simon Sturn. 2018. The German employment miracle in the great recession: The significance and institutional foundations of temporary working-time reductions. *Oxford Economic Papers* 70:206–224.

Hipp, Lena, Janine Bernhardt and Jutta Allmendiger. 2015. Institutions and the prevalence of nonstandard employment. *Socio-Economic Review* 13:351–377.

Hirst, Paul, and Thompson, Grahame. 1995. Globalization and the future of the nation state. *Economy and Society* 24:408–442.

Horemans, Jeroen. 2016. Polarisation of non-standard employment in Europe: Exploring a missing piece of the inequality puzzle. *Social Indicators Research* 125:171–189.

ILO. 2016. *Non-standard employment around the world: Understanding challenges, shaping prospects*. Geneva: ILO.

Jahoda, Marie, Paul Lazarsfeld and Hans Zeisel. 1933. *Marienthal: The sociography of an unemployed community*. Chicago: Aldine.

Jansen, Andreas. 2017. Work-retirement cultures: A further piece of the puzzle to explain differences in the labour market participation of older people in Europe? *Ageing and Society*. online first.

Kahn, Lawrence M. 2007. The impact of employment protection mandates on demographic temporary employment patterns: International microeconomic evidence. *Economic Journal* 117:333–356.

Kahn, Lawrence M. 2010. Employment protection reforms, employment and the incidence of temporary jobs in Europe: 1996–2001. *Labour Economics* 17:1–15.

Kahn, Lawrence M. 2012. Labor market policy: A comparative view on the costs and benefits of labor market flexibility. *Journal of Policy Analysis and Management* 31:94–110.

Kalleberg, Arne L. 2000. Nonstandard employment relations: Part-time, temporary and contract work. *Annual Review of Sociology* 26:341–365.

Kalleberg, Arne L. 2001. Organizing flexibility: The flexible firm in a new century. *British Journal of Industrial Relations* 39:479–504.

Kalleberg, Arne L. 2003. Flexible firms and labor market segmentation: Effects of workplace restructuring on jobs and workers. *Work and Occupations* 30:154–175.

Kasza, Gregory J. 2002. The illusion of welfare 'regimes'. *Journal of Social Policy* 31:271–287.

van Kersbergen, Kees, and Barbara Vis. 2015. Three worlds' typology: Moving beyond normal science? *Journal of European Social Policy* 25:111–123.

Kittel, Bernhard. 2006. A crazy methodology? On the limits of macro-quantitative social science research. *International Sociology* 21:647–677.

Koch, Max, and Martin Fritz. 2013. *Non-standard employment in Europe. Paradigms, prevalence and policy response*. Basingstoke: Palgrave.

Koeniger, Winfried, Marco Leonardi and Luca Nunziata. 2007. Labor market institutions and wage inequality. *International Labour Review* 60:340–356.

Kroneberg, Clemens. 2019. Theory development in comparative social research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00604-y.

Lemieux, Thomas. 2008. The changing nature of wage inequality. *Journal of Population Economics* 21:21–48.

Lewis, Jane. 1992. Gender and the development of welfare regimes. *Journal of European Social Policy* 3:159–173.

Lübke, Christiane, and Marcel Erlinghagen. 2014. Self-perceived job insecurity across Europe over time: Does changing context matter? *Journal of European Social Policy* 24:319–336.

Mandel, Hadas. 2012 Winners and losers: The consequences of welfare state policies for gender wage inequality. *European Sociological Review* 28:241–262.

Marginson, Paul, and Christian Welz. 2015. European wage-setting mechanisms under pressure: Negotiated and unilateral change and the EU's economic governance regime. *Transfer* 21:429–450.

Mau, Steffen, Jan Mewes and Nadine M. Schöneck. 2012. What determines subjective socio-economic insecurity? Context and class in comparative perspective. *Socio-Economic Review* 10:655–682.

Meuleman, Bart. 2019. Multilevel Structural Equation Modeling for Cross-National Comparative Research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00605-x.

Mishel, Lawrence, John Schmitt and Heidi Shierholz. 2014. Wage inequality: A story of policy choices. *New Labor Forum* 23:26–31.

Muffels, Ruud. 2015. Understanding occupation differences in flexibilisation and mobility patterns in Europe: Do institutions matter? In *Non-standard employment in post-industrial labour markets. An occupational perspective*, eds. Werner Eichhorst and Paul Marx, 298–323. Cheltenham: Edward Elgar.

Muffels, Ruud, and Ruud Luijkx. 2008. Labour market mobility and employment security of male employees in Europe: 'Trade-off' or 'flexicurity'? *Work Employment Society* 22: 221–241.

OECD. 1999. Employment protection and labor market performance. In *Employment Outlook,* published by OECD, 49–132. Paris: OECD.

Olsthoorn, Martin. 2016. Institution-driven inequalities in the risk of temporary employment: job or skill based? The relative deregulation of temporary employment and its relation to inequalities in the risk of temporary employment. *European Sociological Review* 32:517–531.

van Oorschot, Wim, and Heejung Chung. 2015. Feelings of dual-insecurity among European workers: A multi-level analysis. *European Journal of Industrial Relations* 21:23–37.

van Winkle, Zachary, and Anette Fasang. 2017. Complexity in employment life courses in Europe in the twentieth century—Large cross-national differences but little change across birth cohorts. *Social Forces* 96:1–30.

Orloff, Ann Shola. 1993. Gender and the social rights of citizenship: The comparative analysis of gender relations and welfare states. *American Sociological Review* 58:303–328.

Ott, Notburga. 1992. *Intrafamily bargaining and household decisions*. Berlin: Springer.

Roose, Jochen. 2013. Fehlermultiplikation und Pfadabhängigkeit. Ein Blick auf Schattenseiten von Sekundäranalysen standardisierter Umfragen. [Error multiplication and path-dependency. Examining the downside of secondary analyses of standardised surveys] *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 65:697–714.

Schettkat, Ronald. 1996. Flows in labor markets. Concepts and international comparative results. In *The flow analysis of labor markets*, ed. Ronald Schettkat, 14–36. London: Routledge.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Scholte, Jan Aart. 2005. *The sources of neoliberal globalization. United Nations Research Institute for Social Development (UNRISD) Programme Paper*. Geneva: UNRISD.

Schröder, Martin. 2009. Integrating welfare and production typologies: How refinements of the varieties of capitalism approach call for a combination with welfare typologies. *Journal of Social Policy* 38:19–43.

Schröder, Martin. 2019. Varieties of capitalism and welfare regime theories: Assumptions, accomplishments, and the need for different Methods. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00609-7.

Simón, Hipolito. 2010. International differences in wage inequality: A new glance with European matched employer-employee data. *British Journal of Industrial Relations* 48:310–346.

Stier, Haya. 2015. The skill-divide in job quality: A cross-national analysis of 28 countries. *Social Science Research* 49:70–80.

Stiglitz, Joseph E. 2015. *New theoretical perspectives on the distribution of income and wealth among individuals: Part I. The Wealth Residual. NBER Working Paper 21189*. Cambridge: NBER.

Streeck, Wolfgang. 2012a. How to study contemporary capitalism? *European Journal of Sociology* 53:1–28.

Streeck, Wolfgang. 2012b. Skills and politics: General and specific. In *The political economy of collective skill formation*, eds. Marius R. Busemeyer and Christine Trampusch, 317–352. Oxford: Oxford University Press.

Thelen, K. 2014. *Varieties of liberalization and the new politics of social solidarity*. Cambridge: Cambridge University Press.

VanHeuvelen, Tom. 2018. Within-group earnings inequality in cross-national perspective. *European Sociological Review* 34:286–303.

Wallerstein, Michael. 1999. Wage-setting institutions and pay inequality in advanced industrial societies. *American Journal of Political Science* 43:649–680.

Wimmer, Andreas, and Nina Glick Schiller. 2002. Methodological nationalism and beyond: Nation-state building, migration and the social sciences. *Global Networks* 2:301–334.

Wulfgramm, Melike, and Lukas Fervers. 2015. Unemployment and subsequent employment stability: Does labour market policy matter? *Socio-Economic Review* 13:791–812.

Marcel Erlinghagen 1971, Dr. sc. pol., Professor for Sociology, Institute for Sociology, University of Duisburg-Essen. Areas of research: labor market mobility, job security, volunteering, migration. Publications: Führungspositionen im Ehrenamt – ein weiterer Bereich der Benachteiligung von Frauen? Kölner Zeitschrift für Soziologie und Sozialpsychologie 68, 2016 (with B. Şaka and I. Steffentorweihen); Self-perceived job insecurity across Europe over time: Does changing context matter? Journal of European Social Policy 24, 2014 (with Ch. Lübke).

ABHANDLUNGEN

The Role of Welfare State Characteristics for Health and Inequalities in Health from a Cross-National Perspective: A Critical Research Synthesis

Timo-Kolja Pförtner · Holger Pfaff · Frank J. Elgar

Abstract Research in public health, epidemiology, and medical sociology has debated the health impacts of income inequality. A common hypothesis is that socioeconomic contexts in neighborhoods, regions, or entire nation states shape individuals' health and inequalities in health. Several reviews have concluded that the distribution of income within a society contributes to poor health. More recent research has focused on welfare state arrangements and their relevance to cross-national variations in health and inequalities in health. We argue that the welfare state determines and mediates the extent of inequalities in health through healthcare, social policy and public health. Many studies have been published in the last decade that have tried to examine the role played by these influences on health and health inequalities. This review will (1) summarize the extant research on the association between welfare state factors and health outcomes, (2) discuss how research theorizes the role of welfare state characteristics for between- and within-country differences in health, (3) present different approaches to empirically investigate the association between welfare state, health and inequalities in health, and (4) provide methodological considerations in this field of research.

T.-K. Pförtner (✉) · H. Pfaff
Institute for Medical Sociology, Health Services Research and Rehabilitation Science (IMVR),
University of Cologne
Eupener Str. 129, 50933 Cologne, Germany
E-Mail: timo-kolja.pfoertner@uk-koeln.de

H. Pfaff
E-Mail: holger.pfaff@uk-koeln.de

F. J. Elgar
Institute for Health and Social Policy, McGill University
130 Pine Avenue West, H3A 1A3 Montreal, QC, Canada
E-Mail: frank.elga@mcgill.ca

Springer

Der Stellenwert wohlfahrtsstaatlicher Arrangements für die Gesundheit und gesundheitliche Ungleichheit aus länderübergreifender Perspektive: eine kritische Forschungssynthese

Zusammenfassung Sowohl in der medizinischen Soziologie, Public-Health-Forschung als auch in der Epidemiologie wird seit Jahrzehnten diskutiert, inwieweit eine ungleiche Einkommensverteilung mit der gesundheitlichen Lage von Individuen und gesamten Gesellschaften zusammenhängt. Auf der Ebene von Nachbarschaften, Regionen oder Nationalstaaten soll der sozioökonomische Kontext einerseits die Gesundheit einzelner, andererseits auch die ungleiche Verteilung von Gesundheit insgesamt maßgeblich bestimmen. Empirische Studien zeigen, dass die Einkommensungleichheit mit einer schlechteren Gesundheit auf der Individual- wie Populationsebene assoziiert ist. Neuere Forschungsarbeiten konzentrieren sich auf den Wohlfahrtsstaat und auf wohlfahrtsstaatliche Arrangements als Determinanten von Gesundheit und gesundheitlicher Ungleichheit. Neben wohlfahrtsstaatlichen Arrangements haben nach Ansicht jüngerer Forschungsansätze das Gesundheitswesen wie auch Public-Health-Programme wesentlichen Einfluss auf die Verteilung von Gesundheit innerhalb von Nationen. In den letzten Jahren sind zahlreiche Studien veröffentlicht worden, die sich mit dem Stellenwert wohlfahrtsstaatlicher Arrangenemts, des Gesundheitswesens und von Public-Health-Programme für die Gesundheit und gesundheitliche Ungleichheit befasst haben. Der vorliegende Beitrag fasst die vorliegende Studienlage zum Stellenwert von Wohlfahrtsstaat und wohlfahrtsstaatlichen Arrangements, des Gesundheitswesens und von Public-Health-Programmen für die Verteilung von Gesundheit innerhalb und zwischen Nationen zusammen. Hierbei wird die Studienlage dargestellt und diskutiert. Darüber hinaus werden methodische Forschungsansätze zur Thematik vorgestellt und kritisch evaluiert.

1 Introduction

Societies throughout the developed and developing world have made dramatic progress in health over the past 100 years. Mortality rates have fallen sharply, and life expectancy has increased in most countries. According to the UN's World Population Prospects, life expectancy has increased since the 1950s, and will continue to increase on every continent (see Fig. 1). However, significant differences in life expectancy both between and within societies illustrate the importance of context in a population's health. One of the first scholars to focus on welfare state outcomes in the explanation of national differences in population health was Richard G. Wilkinson (1990, p. 392), who stated in the light of his own findings and findings from
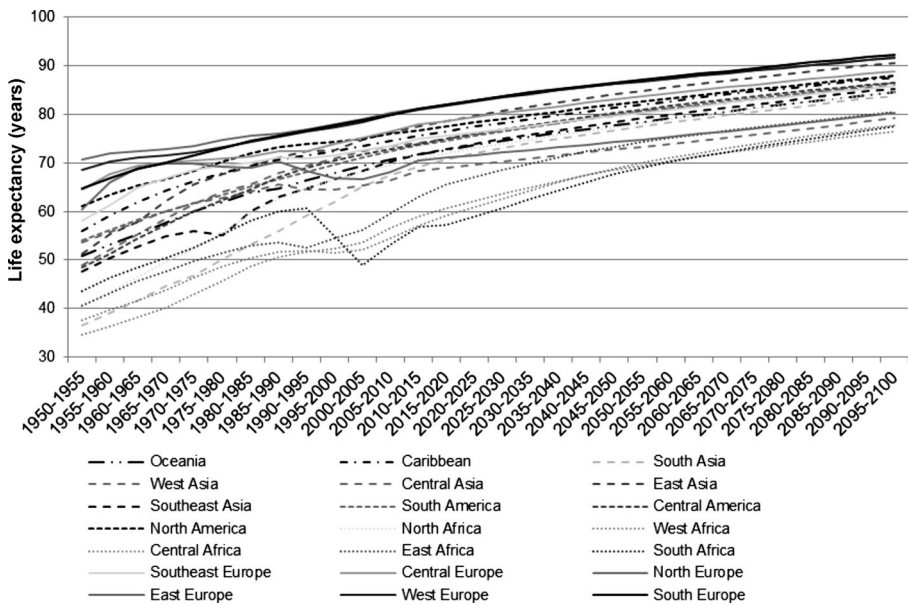
Fig. 1 Life expectancy at birth (both sexes combined) by aggregate, 1950–2100 (years). Source: World Population Prospects: The 2015 Revision—Special Aggregates: Publication List: Ecological—Special

Rodgers ([1979]) that "*as countries get richer the relationship between life expectancy and average income appears to weaken and be replaced by the growing influence of income distribution*". According to Wilkinson, this "epidemiological transition" marks a point in global development when the material conditions that support health are sufficient for rifts to emerge between socioeconomic strata due to psychosocial consequences of inequality.

To explain differences in population health, Wilkinson and Pickett ([2011]) suggests that, among affluent nations, life expectancy is lower in more unequal societies because income inequality in itself has a negative effect on members' health. In contrast to this *contextual effect* of income inequality, others have suggested that the relationship between income inequality and life expectancy results from a *composition effect* where the income inequality level of a country embodies the sum of individual life chances within a society (Gravelle [1998]; Jen et al. [2009]). This composition effect pointed to an important determinant of health on the individual level: socioeconomic status. Socioeconomic status reflects a variety of resources such as money, knowledge, prestige, power, and beneficial social connections that make it possible to protect and improve health, and is often measured by income, education and occupational status (Link and Phelan [1995]). Therefore, less privileged members of a society live in worse health than more privileged members due to the difference in resource ownership (Phelan et al. [2010]), which in turn is reflected by the level of income inequality. According to general medical sociological theories, socioeconomic status contributes to individual health through material (e.g. physical working conditions, neighborhood conditions), psychosocial (e.g. financial strain, deprivation, psychosocial working conditions), and behavioral pathways (e.g. smok-

ing behavior, physical activity, nutrition) (see Fig. 2) (Galobardes et al. 2006; Elo 2009; Braveman et al. 2011; Braveman and Gottlieb 2014; Moor et al. 2017). While this debate is mainly about the mechanisms explaining differences in population health by means of macro-level (income inequality) and micro-level factors (socioeconomic status)[1], only a few studies have investigated the underlying patterns of socioeconomic inequalities (Maio 2012, p. 41). Coburn (2000, p. 136) criticized the narrow focus on the social determinants of health by arguing that there is "*an overwhelming tendency to focus on the possible social/psycho-biological mechanisms through which social factors might be tied to health* ... [and] *a startling lack of attention to the social/political/economic context of SES or income inequality-health status relationships*".

In line with Coburn's criticism, the body of cross-national research into health and health inequalities has placed a larger focus on the wider policy context determining education, work and income within a country (Bergqvist et al. 2013, pp. 1–2; Woolf and Braveman 2011). In particular, a welfare state perspective has been developed to understand the *causes of the causes* of health and health inequalities (Beckfield et al. 2015, p. 228). The welfare state has been considered as a relevant macro-level factor determining and mediating the extent of socioeconomic inequalities in health by healthcare, social policy and public health (Thomson et al. 2016). Many studies have been published in the last decade trying to unravel the role of welfare states for health and health inequalities. The current review summarizes the extant research on the association between welfare state factors and health outcomes, discusses how research theorizes the role of welfare state characteristics for between- and within-country differences in health and inequalities in health, presents different approaches to empirically unravel the association between welfare state, health and inequalities in health, and provides methodological considerations in this field of research.

2 Public Health and the "Three Worlds of Welfare"

Cross-national research into health and health inequalities has increasingly examined the role of the welfare state as a broad determinant of health. Based on Esping-Andersen's "*welfare state as a system of stratification*" (Esping-Andersen 1990, pp. 69–77), studies on adults and adolescents have investigated whether and how the welfare state is linked to within- and between-country differences in health and inequalities in health. Although researchers have suggested various classifications and approaches to measure welfare states (Bambra 2007), Esping-Andersen's 'three worlds of welfare' are still highly influential in the literature because it explains *how* the welfare state shapes population health and socioeconomic inequalities in health

---

[1] The validity of Wilkinson's inequality-hypothesis is still highly debated (Lynch et al. 2004; Macinko et al. 2003; Maio 2012; Pickett and Wilkinson 2015; Subramanian and Kawachi 2004; Wilkinson and Pickett 2007), and more sophisticated studies indicate that the relationship between income inequality and health at the individual level is at least small or inconsistent (Leigh et al. 2011). In particular, social epidemiologists doubt that there is a context effect of income inequality on individual health, and emphasize the need for better data and methods (Leigh et al. 2011; Subramanian and Kawachi 2004).
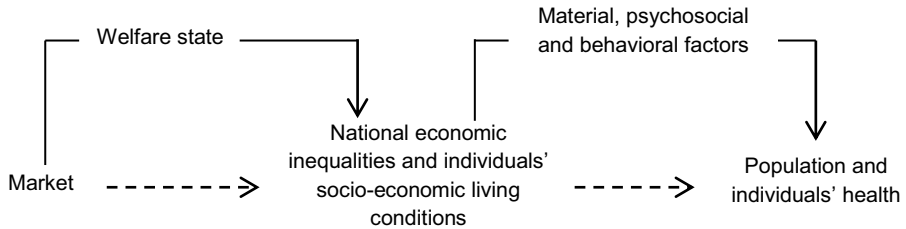
Fig. 2 Simple conceptual model on mediating market-derived socioeconomic inequalities in health chances via welfare services

by mediating individuals' life chances via the welfare services that they receive (see Fig. 2).

In his influential work, Esping-Andersen (1990) provides a typology of welfare states based on three dimensions: decommodification (the extent and generosity of welfare state services that determine individuals' dependency on the market, particularly in terms of pensions, unemployment benefit and sickness insurance), social stratification (conditions under which welfare state policies contribute to an equalization of opportunity structures), and the private–public mix (the relative contributions of the state, family, the voluntary sector and the market in welfare state provision) (Bambra 2007; Hurrelmann et al. 2011; Schröder 2019). By applying these dimensions, Esping-Andersen was able to define three ideal regime types: liberal, conservative and social democratic. In liberal-regime countries[2] (such as Australia, New Zealand, Canada, Ireland, the UK and the USA), decommodification is minimal, and welfare benefits are modest, often based on strict entitlement criteria and means tested. The conservative regime type (including countries such as Finland, France, Germany, Italy or Switzerland) is characterized by its "status-differentiating" welfare programs, in which welfare benefits depend on work-based insurance contributions. These status-differentiating welfare programs result in differing decommodification effects and contribute to the consolidation of social stratification. Finally, the social democratic regime type[3] (including countries such as Austria, Belgium, Denmark, Norway or Sweden) provides universal and comparatively generous benefits, and the state plays a strong interventional role promoting equality in different forms of income redistribution (such as progressive taxation, minimum wage policies, tax credits, and cash transfers to lower income groups).

Researchers have suggested modifications to Esping-Andersen's regime typology (Arts and Gelissen 2002, 2010; Bambra 2007; Eikemo and Bambra 2008). One of the most important modifications relates to the inclusion of other regime types such as the 'Latin-rim' (Leibfried 1992) or 'Southern' regime type (Bonoli 1997; Ferrera

---

[2] Clustering the Antipodean countries (e.g. Australia and New Zealand) into the liberal regime type has been criticized due to a more particular and a more inclusive approach to social protection than the standard liberal form (Arts and Gelissen 2002).

[3] Other welfare state typologies cluster the Scandinavian countries (Norway, Finland, Sweden and Denmark) into a single "Scandinavian" regime type (Leibfried 1992; Ferrera 1996; Bonoli 1997) to emphasize their exceptional universalistic and generous welfare regimes aiming to provide full employment and social equality.

1996; Saint-Arnaud and Bernard 2003) (characterized by a fragmented system of welfare provision and high reliance on the family and voluntary sector), the 'Confucian' regime type (Karim et al. 2010) (characterized by a minimal welfare state with a strong emphasis placed on the family and on the voluntary sector based on Confucian ethics), and the 'Eastern European' regime type (Fenger 2007) (characterized by its unique political background and a shift from a Communist towards a somewhat more liberal welfare state focusing on marketization and decentralization). Further modifications of the regime typology have been made in public health using the actual extent of services provided by different welfare states (Bambra 2005a, 2005b), or extending the focus to politics and policies as determinants of welfare state programs (Navarro 1999; Navarro and Shi 2001; Navarro et al. 2006).

Although the regime typology of Esping-Andersen has been modified and extended, the logic linking welfare state characteristics with population health and inequalities in health is still based on the idea of the welfare state as a system of stratification. For example, in a study by Eikemo et al. (2008b, p. 2282), the authors state that "*welfare states provide a variety of social transfers (such as housing-related benefits, unemployment, pensions, and sickness and disability benefits) as well as key services (most notably healthcare or social services), which together mediate the relationship between socioeconomic position and health.*" Accordingly, it has been assumed that population health and inequalities in health vary by welfare regime type with better health outcomes in the most generous regime (social democratic regime) compared to types with lower levels of decommodification and welfare benefits (such as in the liberal regime type).

A number of empirical reviews have focused on the association of welfare regime types with health and inequalities in health (Beckfield and Krieger 2009; Bergqvist et al. 2013; Brennenstuhl et al. 2012; Muntaner et al. 2011). On a national level, reviews indicate that the Scandinavian regime type generally shows better health outcomes than other regime types in infant and child mortality, but not in other health outcomes such as in (working-age/old-age/all-age) mortality, life expectancy or self-rated health. Moreover, given studies thus far, early research reviews conclude that although the Scandinavian regime type has relatively low levels of economic inequality, it does not consistently have low levels of socioeconomic inequalities in health. According to a review by Bergqvist et al. (2013), only one out of four studies using Esping-Andersen's typology or a modified typology found the smallest inequalities in health in social democratic countries in men but not in women. In other studies, inequalities in mortality and self-rated health were lower in the conservative-regime countries compared to the social democratic countries. This "*Scandinavian welfare paradox of health*" was also found in other studies using different regime typologies (Beckfield and Krieger 2009; Bergqvist et al. 2013; Brennenstuhl et al. 2012; Muntaner et al. 2011). When explaining this empirical puzzle, several theories have been proposed, drawing on methodological issues (see Sect. 5) and existing theories of inequalities in health (Mackenbach 2012; Bambra 2011). Therefore, although the application of welfare regime typologies is still very popular, the evidence regarding the influence of welfare regimes on health and inequalities in health is inconclusive and mixed.

3 Going Beyond the Regime Typology—The Institutional Approach

The '*Scandinavian Puzzle*' and the general criticism towards using regime typologies have led to alternative approaches towards unraveling the role of welfare state characteristics for health and inequalities in health. As Lundberg (2008, p. 1106) puts it, using regime typologies "*may be helpful for descriptive purposes, they are much less useful if we really want to open the black box and analyze what aspects of welfare state are of importance*". Instead, Lundberg (2008) suggests studying specific welfare programs and their influences on health and inequalities in health. Therefore, it is necessary to consider the link between specific welfare institutions (e.g. unemployment benefits) and the health of specific social groups (e.g. the unemployed). Studies on specific welfare programs (social policy, healthcare and public health) could be classified into five groups: studies focusing on family benefits, pension benefits, economic assistance and unemployment benefits, access to health/medical care and public health interventions.

The first group analyzes the relevance of *family benefits* (e.g. family cash and tax benefits, paid parental leave, childcare support, child allowance, parental insurance, childcare leave) for the individual health of specific social groups such as lone mothers or children in poor households (Bergqvist et al. 2013 ; Hank and Steinbach 2019). For example, Aitken et al. (2015) reviewed seven studies analyzing the relevance of paid maternity leave for mothers' health, and found a positive correlation between paid maternity leave and mental and physical health. They conclude that paid leave may protect mothers from financial strain and enables them to spend more time away from the workplace and to recover from the physical effects of childbirth. Similar studies indicate that generous and universal family policies are beneficial for family members' health and socioeconomic living conditions (Bergqvist et al. 2013). In contrast, recent studies focusing on the impact of income support programs for low-income families (e.g. the Earned Income Tax Credit (EITC) in the USA or the Family Tax Credit (FTC) in New Zealand) on individual health found contradictory results. Studies showed a positive effect of income support programs on maternal and child health (Hamad and Rehkopf 2016; Strully et al. 2010), whereas others revealed no or only moderate effects for adult health (Larrimore 2011; Pega et al. 2013, 2014, 2016). For example, Wicks-Lim and Arno (2017) applied difference-in-difference analysis to measure low birth weight rate, prenatal health and asthma-related pediatric hospitalization in 90 low- and middle-income neighborhoods before and after the expansion of the New York State and New York City EITC policy between 1997 and 2010. The results showed contradictory findings of reduction on low birth weight rate by increasing EITC benefits, but no effects for prenatal health or asthma-related pediatric hospitalization.

The second group of studies focuses on the relevance of *pension benefits* for individual health in old ages (Bergqvist et al. 2013). Studies hypothesized that the level of generosity and kind of public benefits (basic pensions of persons with no or low earnings; income pensions given on the basis of work contribution) are related to the health of the older people and retirees in particular. Accordingly, it is proposed that more generous pension benefits and universal basic pensions relate to higher incomes that enable investing in health-enhancing products and protect against poverty-re-

lated health burdens (Norström and Palme 2010; Esser and Palme 2010). The few studies published so far indicate that more generous pensions and universal basic pensions (in particular for women) predict better self-rated health and well-being (Esser and Palme 2010). Only few studies have yet been published focusing on the role of pension benefits for the individual health of different socioeconomic groups in old age. Farrants (2017) investigated associations of net replacement rates of pensions with inequalities in the self-rated health of pensioners using the Health Survey for England and the Swedish Study of Income and Living Conditions from 1991 to 2011. Accordingly, annual inequalities in self-rated health by education of pensioners (log odds) were modeled against net replacement rates in pensions by linear regression. Results were contradictory, and indicated a negative association between the net replacement rates of pensions and the magnitude of inequalities in health in England, but no significant association for Sweden. In an ecological study[4] of 16 European countries that was carried out from 2004 to 2014, Reeves et al. (2017) found that greater public pensions entitlement is associated with decreased unmet medical need due to financial reasons. This was observed in particular among the poorest income group. Results also showed that the association between out-of-pocket payment and unmet medical need due to financial reasons was mitigated by higher levels of public pensions entitlement. Therefore, public pensions enable people to meet the various costs involved in seeking healthcare and in keeping healthy, particularly for the lowest socioeconomic groups in old age.

The third group of studies analyzes the influence of *economic assistance and unemployment benefits* on the health of the unemployed and of those in financial need (Bergqvist et al. 2013). The rationale behind these studies is that the mechanisms linked to the generosity of unemployment benefits (e. g. wage replacement rate, duration of coverage, flexibility in the entitlement and maintenance of unemployment benefits) could act as a buffer against loss of wages, alleviate poverty, and subsequently protect individuals from the negative health consequences of unemployment (Cylus et al. 2015). According to a recent review, the generosity of unemployment benefits does have a positive effect on the health of the unemployed, as is shown for mental health, subjective well-being and financial strain (O'Campo et al. 2015). For example, in a current study by Vahid Shahidi et al. (2016) using the European Sociological Survey from 2012, the association of national unemployment insurance replacement rates with the self-rated health of the unemployed and the employed was analyzed using cross-level interactions in a multilevel framework. Study results indicated that more generous levels of unemployment benefits were significantly associated with narrowed inequality in self-rated health between the unemployed and those in employment. Therefore, the increased risks of a poor self-rated health status among the unemployed were lower in those countries that have higher levels of unemployment benefits.

Bergqvist et al. (2013) identified a fourth group of studies that analyze inequalities in *access to health/medical care*. Several determinants on the country level have been found to be associated with decreased socioeconomic inequalities in access

---

[4] According to Levin (2003, p. 108), an ecological study is "*an observational study defined by the level at which data are analyzed, namely at the population or group level, rather than individual level.*"

to healthcare such as public financing and provision of services, medical density and the existence of a gatekeeping system that coordinates patients' care paths via general practitioners (Or et al. 2008). Further, Jones et al. (2006) found with data from the European Community Household Panel (1994–1997) that supplementary private insurances are associated with income, better self-rated health and higher rates of specialist visits indicating that those on a high income might benefit from the existence of private supplementary insurance when it comes to access to healthcare systems as well as to health. According to a review from Huber and Mielck (2010), a similar situation is observable for Germany that allows people to opt out of public insurance and into private insurance if their income exceeds a certain level. Accordingly, the review indicates that the benefits of private insurance with regard to new, innovative drugs, organ transplantations, financial burden due to copayments, waiting times, and communication between patient and physician compared to those with public insurance are more generous. However, whether these observed benefits in access and use of healthcare of higher socioeconomic groups affect inequalities in health is highly debated in public health and health services research, and existing evidence is limited (Davis 1991; Oliver and Mossialos 2004; Pfaff and Pförtner 2016). An exemplary study that tries to unravel the relevance of healthcare-related factors for inequalities in health has been conducted by Klein et al. (2016). They analyzed the influence of disease-related (tumor stage, biological characteristics), patient-related (comorbidity, health behavior, psychosocial characteristics), and healthcare-related factors (treatment, screening uptake, medical expertise) on socioeconomic inequalities in health-related quality of life among patients with prostate cancer six months after radical prostatectomy using a prospective observational study among 246 patients. A stepwise approach was conducted comparing changes in the association between socioeconomic status and health-related quality of life when explanatory factors were included in the regression model. The result indicated a strong association between health-related quality of life six months after treatment and lower socioeconomic status measured by income, occupation and education. Socioeconomic inequalities in quality of life changed only slightly when explanatory factors were considered in the regression model with stronger explanatory power of patient- and healthcare-related factors. Thus, other potential social determinants of health might play an important role in inequalities in patient health which are not related to the healthcare system, such as equity in early life, labor market disadvantages, psychosocial burdens at work, and material deprivation. Therefore, public health research describes the healthcare system not as the primary factor accounting for inequalities in health, but rather as a moderator of levels of inequalities in health (Marmot and Allen 2014). For example, Banks et al. (2006) showed with cross-national data from England and the US that universal access to health/medical care is associated with better health outcomes, but nonetheless found differences in health outcomes in the top socioeconomic groups in England and the US. They concluded that other social factors might contribute to inequalities in health that cannot be explained by the healthcare system alone.

Additionally, as a fifth stream, research aims to assess how welfare states influence inequalities in health institutionally through *public health interventions* (Thomson et al. 2016). Public health interventions play an important role in reducing inequal-

ities in health, as they aim to prevent the occurrence (primary prevention) or the development (secondary prevention) of specific diseases. They are implemented at different levels (national, regional, local, or individual level), focusing on upstream or downstream determinants of public health. Accordingly, upstream interventions focus on fundamental structures and economic conditions influencing individuals' health and health behavior by state or institutional regulations (e.g. tobacco price regulations or smoking bans). Downstream interventions focus directly on the individual and include behavioral approaches for prevention or disease management (e.g. media campaigns against smoking, and smoking cessation programs).[5] Although public health interventions are intended to improve or sustain health, they can also have unintended adverse effects. According to Lorenc and Oliver (2014), public health interventions could lead to direct harm, psychological harm, group and social harm, opportunity cost harm and equity harm. The latter is of particular interest as some interventions may generate socioeconomic inequalities in health when privileged groups benefit from interventions more than disadvantaged ones do.[6] According to a rapid overview of systematic reviews by Lorenc et al. (2013), downstream preventive interventions (media campaigns on smoking, and workplace smoking bans) seemed to be more likely to increase health inequalities than upstream interventions (structural workplace interventions; provision of resources; and fiscal interventions, such as tobacco pricing). Similarly, in a recent systematic review by McGill et al. (2015), they found that price interventions (upstream) to promote healthy eating are more effective among socioeconomically disadvantaged groups than are person-specific interventions (downstream), which were more effective among higher socioeconomic groups. Similar conclusions on the equity effect of interventions were made from systematic reviews on interventions tackling obesity (Beauchamp et al. 2014) or smoking (Hill et al. 2014).

The application of specific welfare programs (social policy, healthcare and public health) are an adequate answer to the use of welfare regime typologies, which are rather unspecific when it comes to explaining health and inequalities in health. Although the application of the institutional approach is very complex and broad, it overcomes the strong theoretical and methodological amendments that are necessary when applying welfare regime typologies. The focus on specific policies allows more precise conclusions to be drawn about their role in affecting the health of different socioeconomic groups. The evidence on social policies indicates that generous benefits are associated with better health, and with lower inequalities in health, whereas empirical findings on the effect of the healthcare system on inequalities in health are rather small and inconclusive. The same is true for the role of public health

---

[5] In 1975, John McKinlay (1979) introduced the terms upstream and downstream when describing his frustration with medical practice. In his analogy of a river that represents diseases, he said that physicians are so busy constantly rescuing victims from the river that they have no time to look *upstream* to check who is pushing their patients into the river. Instead, health professionals face challenges with *downstream* endeavors that are short-term, problem-specific and individual-based.

[6] According to the inverse equity hypothesis of Victora et al. (2000), which is related to Rogers' theory of diffusion (2005), when new public health interventions are implemented, higher socioeconomic groups will initially benefit, and health inequities will widen, but if coverage increases over time, the disadvantaged groups can eventually catch up and health inequities can be decreased.

interventions, which is also a very broad and complex topic with mixed findings for different settings, risks behaviors, and health outcomes.

4 New Developments in Theorizing the Association Between Welfare States and Inequalities in Health

In theorizing the association between welfare state arrangements and inequalities in health, new explanatory approaches have emerged that consider the complex structure and processes linking different welfare policies to health and inequalities in health (Hurrelmann et al. 2011; Beckfield et al. 2015; Gkiouleka et al. 2018). These approaches take into account multiple levels, the specificity of diseases and social groups, and the interdependence of welfare state arrangements in their effect on the health of different social groups. Therefore, these approaches overcome the abovementioned linear logic in the relationship between welfare state arrangements, socioeconomic conditions and health, and the narrowed focus on distal variables on a high level of aggregation (Hurrelmann et al. 2011).

Beckfield et al. (2015) present an institutional theory of the welfare state that simultaneously considers welfare state effects of redistribution, compression and mediation on health and its social determinants. Accordingly, inequalities in health are a function of the institutional welfare effects of redistribution (institutional arrangements which redistribute the social determinants of health such as income, wealth, living standards or education), compression (institutional arrangements providing a limit of healthcare for citizens), and mediation (reducing/limiting inequalities in the social determinants of health such as income or education). These institutional effects of welfare state arrangements on health and its social determinants could interact and operate in multiple domains and at multiple levels ('institutional imbrications'). They might also overlap and interact, and could have a direct effect on health, but could also indirectly influence health via the socioeconomic living conditions. For example, the distribution of health within a society can be influenced directly through healthcare institutions (healthcare and public health), and indirectly through institutional effects on the social determinants of health such as on income via economic assistance and unemployment benefits (social policy). Importantly, Beckfield et al. (2015) emphasize that the specific effects of institutions need to consider knowledge obtained from disease etiology, life course research, and (historical) institutional changes, which assigns individuals with a specific disease to a specific welfare state life course (Halfon and Hochstein 2002; Bambra et al. 2010; Levecque et al. 2011). The institutional effects of healthcare should thus be restricted to avoidable diseases that can be avoided through optimal-quality healthcare.[7] When considering the multiple and complex mechanisms of welfare

---

[7] To capture the contribution of healthcare to population health, the American Working Group on Preventable and Manageable Diseases introduced the concept of amendable mortality, which refers to deaths that could have been avoided by providing effective medical care in good time. In contrast, preventable (causes of) mortality refers to deaths that could have been avoided through timely and effective public health interventions (Nolte and McKee 2004).

arrangements on health and its social determinants, they also suggested considering resources and barriers on the individual (e. g. knowledge with regard to taking up welfare state benefits) and social levels (e. g. stigmatization of welfare recipients) that enable specific social groups to benefit from welfare arrangements, or hinder them from doing so.

In addition, Hurrelmann et al. (2011) proposed a comprehensive model to explain the association between welfare state regimes and inequalities in health, based on the previous work of Navarro et al. (2006) and Esping-Andersen (1990). This model considers a structural (macro), organizational (meso) and individual (micro) level. On the macro-level, the architecture of welfare policy is characterized by the dominance of market, civil networks or state, egalitarianism in civil and human rights, universalism in the provision of social services for citizens, and the level of decommodification. The meso-level is directly influenced by the architecture of welfare policy, and mediates the relationship between the macro and individual levels. It is characterized by several factors such as economic inequality, levels and availability of educational and occupational training, social integration and cohesion, degree of political participation, cultural integration of migrants, religious tolerance, criminality and antisocial behaviors, sense of control of social environment, availability of good food and water, and shelter from environmental contaminants. On the micro-level, the health status of individuals and populations is located and characterized by the quality of the objective and subjective well-being of the entire population, the health quality of disadvantaged and vulnerable groups, and levels of health inequality. Furthermore, Hurrelmann et al. (2011) take into account an intermediate entity that represents general welfare policy (public policy) and healthcare, as well as their combination and where they overlap. The specific characteristics of welfare state and healthcare policies are determined by the type of welfare regimes, and could have both a direct and an indirect effect on health by shaping healthcare institutions and organizations, as well as by influencing individuals' social determinants of health.

5 Methodological Considerations in Research into the Role Played by Welfare State Characteristics for Health and Inequalities in Health

The empirical evaluation of the role of welfare state characteristics for health and inequalities in health is associated with several methodological and theoretical shortcomings which restrict the comparability of the high number of existing studies in this field (Bergqvist et al. 2013). Accordingly, we observed strong variations in the contextual levels, statistical methods, health outcomes and indicators of socioeconomic status in existing studies (detailed information about the reviewed articles can be found in the Appendix Tables). Although a large number of studies have been published so far, there still is a lack of sufficient data making it possible to empirically unravel the role of the welfare state and welfare state arrangements for the individual health of different socioeconomic groups.

This insufficiency of data availability is in particular observable in ecological studies (Navarro and Shi 2001; Navarro et al. 2003; Bambra 2006; Kangas 2010; Karim et al. 2010; Granados 2010; Regidor et al. 2011). Almost all existing eco-

logical studies relied on data from the OECD, UN or WHO, which are restricted to a specific set of countries and health outcomes such as life expectancy or (infant) mortality. For example, most ecological studies relied on data from the OECD, which is currently one of the best data sources for this kind of study. Having said that, and similar to the field of psychology drawing most samples from Western, Industrialized, Rich, and Democratic (WEIRD) societies with the critical assumption that members of WEIRD societies are representative populations (Henrich et al. 2010), the generalizability of OECD-based studies in the field of public health should also be approached with caution.

Moreover, as most ecological studies rely on small sample sizes, results are often sensitive toward outliers (see also Goerres et al. 2019). For example, the US has been identified as a statistical outlier as many values, such as for healthcare expenditures, differ considerably from other societies (Lorenzoni et al. 2014). Chung and Muntaner (2006) discussed this issue extensively, and decided to include the US as these outlying values are not the result of any fault in the sampling process. However, the effect of outliers should not be underestimated when it comes to interpreting the results, as these can change dramatically if outliers are excluded from the analyses.[8] In addition, a small sample size in ecological studies increases the statistical relevance of single countries. For example, in the ecological study by Bambra (2006), analyzing the association between welfare state regime types and infant mortality, the liberal regime type was measured by only one country (UK), and the generalizability to other countries clustered into the liberal welfare regime type should therefore be made with caution.

Furthermore, the restrictions in the availability of adequate health outcomes result in theoretical amendments that are highly debatable (Bergqvist et al. 2013). For example, linking welfare regime types and infant mortality needs strong theoretical presumptions with regard to the mechanisms on the individual level, which most studies do not sufficiently present. However, even with the existence of good theoretical justifications, ecological studies are subject to a fundamental problem: the ecological bias. The ecological bias describes an information loss in the aggregation process that reduces information and prevents associations of interest being identified in the underlying individual-level model (Wakefield 2008). In terms of health inequality research, average levels are not able to provide information on health and its socioeconomic determinants on the individual level, as these aggregate measures mask part of the range of inequality present in the population (Murray et al. 1999). Therefore, critical reviewers might perceive these ecological studies as rather data driven, not taking into account adequate theoretical justifications, confounding factors, and the possibility of an ecological bias.

Although the application of multilevel analyses in ecological studies allows one to control for confounding factors to a certain extent (Chung and Muntaner 2007), ecological studies are still faced by the problem of strong theoretical presumptions when it comes to explaining how welfare regimes might influence the health of

---

[8] Therefore, in the study by Chung and Muntaner (2006) that relies on time-series data from 19 wealthy OECD countries for the years from 1960 to 1994, analyses were conducted with and without the US to detect possible outlier effects.

certain social groups. Therefore, the consideration of individual-level data that make it possible to empirically model the association between indicators of socioeconomic status and health by welfare regime types and facets has been conducted by several studies. These studies were most often based on individual-level data taken from the European Sociological Survey (ESS), the European Union Statistics on Income and Living Conditions (EU-SILC), the World Health Survey (WHS), the Survey of Health, Ageing and Retirement in Europe (SHARE) for Elderly (aged 50+), and the Health Behaviour in School-aged Children (HBSC) for adolescents (aged 11–15 years).

A first type of studies using individual-level data rather descriptively compares inequalities in health by welfare regime types, without controlling for the within-country correlation of observations and other important determinants of health on the contextual level, which increase the risks of biased standard errors ('pooled regression studies') (Zambon et al. 2006; Eikemo et al. 2008a, 2008c; Bambra et al. 2009, 2010; Guarnizo-Herreño et al. 2013, 2014; Alvarez-Galvez et al. 2014; Bambra and Eikemo 2009; Moortel et al. 2015). An exemplary study has been published by Eikemo et al. (2008a). They analyzed income-related inequalities in self-reported health with individual-level data of the ESS from 2004. Individual-level data of countries were pooled by regime types, and income-related inequalities in health (odds ratios) were descriptively compared between regime types. Although the authors highlighted the problem of the within-country correlation of observations, and applied sensitivity analyses by means of multilevel analysis, the final models were conducted without controlling for within-county correlation and, therefore, the results should be interpreted with caution. Moreover, these kinds of studies are unable to empirically quantify differences between welfare regime types in the association of socioeconomic status with individual health, and are unable to control for important control factors on the contextual level.

Therefore, multilevel analysis was conducted by novel studies identified as a second type of studies using individual-level data and the welfare regime type approach[9] (Eikemo et al. 2008b; Dragano et al. 2011; Richter et al. 2012; van der Wel et al. 2012; Witvliet et al. 2012; Chung et al. 2013; van de Velde et al. 2014; Niedzwiedz et al. 2014a, 2014b, 2015; Rathmann et al. 2015; Alvarez-Galvez 2016; Leão et al. 2018). Although the application of a multilevel approach overcomes the abovementioned methodological shortages, other general methodological issues do arise. The application of multilevel analyses is based on strong assumptions that are most often not discussed in studies from the field of public health or epidemiology. For example, it is assumed that errors on the macro level are normally distributed, are independent across contexts and are not correlated with individual-level errors (Diez-Roux

---

[9] The institutional approach is subject to similar methodological shortages when it comes to the application of individual-level data, ecological data or multilevel data analyses. However, studies using an institutional approach are superior to studies using the typology approach as they allow one to analyze the association and causal influence of specific welfare programs on specific health outcomes of certain social groups. The application of the institutional approach is still in its infancy, and various studies have been published so far that differ in their methodological and theoretical settings. This diversity complicates the interpretation and comparison of findings from these studies, and further research needs a stronger theoretical justification with regard to how specific welfare programs influence the health of specific social groups.

2000). This implies that the countries that are under consideration in the data were selected at random, which is most often not the case. For example, the EU-SILC predominantly considered EU Member States, and countries are therefore closely related to each other, thus violating the assumption of independency in multilevel analyses. Bryan and Jenkins (2016) discussed a further problem that arises in most multilevel studies. They argue that the small number of countries in most 'multi-country data sets' limits the ability to detect robust country (cross-level) effects, such as of welfare regime types or other institutional characteristics (see also Schmidt-Catran et al. 2019). These methodological problems should be at least considered when conducting multilevel analyses with data sets which have a small sample size on the contextual level.[10]

Furthermore, analyses based on individual-level data relied most often on self-reported health. A very large number of studies in this field have focused on self-rated health and/or chronic illness. Although self-reported health indicators have been shown to be valid and reliable indicators for mortality (Benyamini 2011; Lima-Costa et al. 2012), research indicates that also social, psychosocial, and survey measurement factors are associated with respondents' self-rated health responses (also described as Measurement Invariance, Reporting Heterogeneity or Different Item Functioning) (Garbarski 2016; see also Cieciuch et al. 2019). A similar issue is true for the measurement and understanding of indicators of socioeconomic status in cross-national research, such as for educational attainment in adult cross-country surveys (Schneider 2010), or for family affluence in cross-country surveys among adolescents (Makransky et al. 2014).

The lack of adequate data is a fundamental problem of the interdisciplinary research into the role of welfare characteristics for health and inequalities in health. For individual-level data, we observe a general inconsistency in the availability of adequate data sources that include objective health measures and data from social sciences. Most analyses in this context stem from social science data that do no not include objective health data for several reasons such as a lack of financial, personal or structural resources to survey such measures. In contrast, epidemiological data most frequently include objective health measures, but contain insufficient information on the socioeconomic living contexts of respondents. Against this background, the cross-national SHARE data are innovative as they provide both socioeconomic background information, and subjective and objective health measures from a number of countries. This allows one to analyze associations between welfare state policies/regimes and inequalities in subjective and objective health. Romaniuk (2014) analyzed differences in socioeconomic inequalities in subjective (self-rated health) and objective health (handgrip strength) from a welfare state regime perspective, including 16 countries from SHARE. Results of multilevel analyses suggested only weak variation for good self-rated health (4.7%) and for normal/strong hand grip

---

[10] Bryan and Jenkins (2016) suggest considering at least 25 countries for linear models and at least 30 countries for logit models in order to derive accurate estimates. However, they also add that this rule of thumb should not be applied blindly but on the basis of the model that is being estimated and the effects in which the researcher is primarily interested. Complicated models that include multiple country-level or cross-level effects should include more countries in order to obtain unbiased and accurate effects.

strength (0.8%) across countries. Socioeconomic inequality in objective health and its association with welfare regime types were lower compared to subjective health, which might be related to cultural differences in answering questions on self-rated health. However, this is the first known study focusing on differences in socioeconomic inequalities in objective and subjective health by welfare regime types, and is associated with the abovementioned methodological shortages such as a low number of observations on the country level.

These methodological shortages must be acknowledged when interpreting and comparing study results relating to the role of welfare characteristics for health and inequalities in health. In particular, ecological studies and pooled regression studies suffer from severe methodological limitations when attempting to explain and empirically unravel the association between welfare characteristics, population health, and the health of specific social groups. Although the application of advanced methods such as multilevel analysis is more accurate for proofing the specific theoretical assumptions on the influence of welfare characteristics on individuals' health, they are nonetheless faced by methodological issues that need to be considered when interpreting study results such as the limited number of countries on the contextual level or the lack of adequate health data.

6 Discussion

In the explanation of between- and within-country differences in health, theories from medical sociology (including areas from public health research, epidemiology, political economy of health and health services research) offer new insights into the causes of the causes of health inequalities: the welfare state. The welfare state has been identified as a relevant macro-level determinant of health that shapes individuals' lives and health chances through healthcare, social policy and public health (Thomson et al. 2016). Several approaches have been proposed with a view to linking welfare state characteristics with health and inequalities in health.

According to the regime approach, the welfare state shapes population health and inequalities in health by mediating market-derived socioeconomic inequalities in individuals' life chances via welfare services. It has been assumed that more generous welfare state regime types such as the Scandinavian regime type show better population health and have the smallest inequalities in health compared to all other regime types (liberal, conservative, southern or eastern regime type) (Eikemo et al. 2008b). However, the evidence does not consistently show smaller inequalities in health in the Scandinavian regime type. Given this Scandinavian puzzle (Mackenbach 2012; Bambra 2011), and the general critique of aggregating countries into different regime types (Lundberg 2008; Hurrelmann et al. 2011) by distal variables, an institutional approach has been developed. Accordingly, empirical research focused on the influence of specific welfare institutions (social policy, healthcare and public health) on the health of specific vulnerable groups. The evidence has shown a beneficial effect of family benefits, pension benefits and economic assistance and

unemployment benefits for the health of vulnerable groups.[11] Moreover, research discussed the importance of access to healthcare and public health interventions, and found that universal access to healthcare and upstream public health interventions are associated with lower inequalities in health.

As both approaches lack a strong theoretical foundation on how welfare states' policies are historically established and intertwined, influenced and determined by political, economic and social dynamics, and are most often not able to explain the extent and social distribution of specific diseases, new complex theoretical approaches have emerged. The innovative approaches of Beckfield et al. (2015) and Hurrelmann et al. (2011) combine the influence of different welfare state and public health interventions in one approach, and describe their interconnectedness and simultaneous influence on health and the social determinants of health. Moreover, they also take into account different contextual levels in the explanation of the influence of welfare state characteristics on health and inequalities in health, and focus on the influence of healthcare on amendable diseases.

These complex approaches might be interpreted as a further step into the theoretical foundation on how the welfare state—including social policy, healthcare and public health—is responsible for shaping and constraining population health, and the extent of social inequalities in health.[12] Further theoretical debates should explore how to include factors such as life course transitions (Bambra et al. 2010), gender (Bambra et al. 2009), migration (Castañeda et al. 2015), adolescence, and later life (Rathmann et al. 2015; Richter et al. 2012; Dahl and Birkelund 1997). Moreover, as most welfare approaches on health and health inequalities focus on income-relevant issues, there is still a need to take into account the impact had by welfare policies on other relevant determinants of health such as the education system (Allmendinger and Leibfried 2016; Rathmann et al. 2016) or the labor market (Muntaner et al. 2010; Julià et al. 2017) for inequalities in health by educational status, occupation or employment status. However, we need to acknowledge that these approaches are empirically tested and proposed only for developed (Western) countries, and therefore the generalizability to other countries such as in Asia, South America or Africa is limited, and needs to be addressed by further theoretical debates (Karim et al. 2010; Chung et al. 2013; Ng et al. 2016).

Finally, the empirical verification of mechanisms linking welfare state characteristics with the health of specific social groups requires advanced methods. Looking back into history in this field of research, a development in the application of methods is visible. Most studies initially relied on aggregate data and conducted rather descriptive and bivariate analyses. These ecological and pooled regression stud-

---

[11] According to Bergqvist et al. (2013), an alternative approach (expenditure approach) focuses on the influence of welfare state spending (social and health spending) on health and inequalities in health. According to their review, welfare state spending contributes to a specific level of spending on health, and is beneficial in terms of bringing about lower levels of inequalities in health.

[12] A project on Health inequalities in European welfare states (HiNEWS) (2015–2018) that is currently underway focuses on the determinants of inequalities in health in European welfare states, the refinement, testing and development of social inequalities in health theory, the identification of policies and interventions with the potential of reducing health inequalities, and a new policy agenda on how health inequalities can be reduced most effectively.

ies have been replaced by conducting advanced studies that allow one to explicitly model associations between different contextual levels or over time. In particular, the application of multilevel analysis has become popular, as it allows one to empirically meet the requirements from theory by modeling the association of welfare state characteristics on the contextual level with the health of different socioeconomic groups on the individual level. By doing so, research has shown that the welfare regime and the generosity of social benefits matters for inequalities in health and with regard to socioeconomically disadvantaged groups. Moreover, they also highlighted the strong theoretical and methodological amendments of past studies that were not able to conduct multilevel analyses, and also point to the general issue in this field of research: the limited availability of adequate data and the theoretical shortages. To further improve this field of research, the interdisciplinary work between scholars from public health, epidemiology, political economy of health, and health services research needs to be intensified and supported by scholars from sociology, political science, and psychology.

References

Aitken, Zoe, Cameryn C. Garrett, Belinda Hewitt, Louise Keogh, Jane S. Hocking and Anne M. Kavanagh. 2015. The maternal health outcomes of paid maternity leave: a systematic review. *Social Science and Medicine* 130:32–41.

Allmendinger, Jutta, and Stephan Leibfried. 2016. Education and the welfare state: the four worlds of competence production. *Journal of European Social Policy* 13:63–81.

Alvarez-Galvez, Javier. 2016. Discovering complex interrelationships between socioeconomic status and health in Europe. A case study applying Bayesian Networks. *Social Science Research* 56:133–143.

Alvarez-Galvez, Javier, Maria L. Rodero-Cosano, Carlos García-Alonso and Luis Salvador-Carulla. 2014. Changes in socioeconomic determinants of health. Comparing the effect of social and economic indicators through European welfare state regimes. *Journal of Public Health* 22:305–311.

Arts, Will, and John Gelissen. 2002. Three worlds of welfare capitalism or more? A state-of-the-art report. *Journal of European Social Policy* 12:137–158.

Arts, Will, and John Gelissen. 2010. Models of the Welfare State. In *The Oxford Handbook of the Welfare State*, eds. Francis G. Castles, Stephan Leibfried, Jane Lewis, Herbert Obinger and Christopher Pierson, p. 569-583. Oxford: Oxford University Press.

Bambra, Clare. 2005a. Cash versus services. Worlds of welfare and the decommodification of cash benefits and health care services. *Journal of Social Policy* 34:195–213.

Bambra, Clare. 2005b. Worlds of welfare and the health care discrepancy. *Social Policy and Society* 4:31–41.

Bambra, Clare. 2006. Health status and the worlds of welfare. *Social Policy and Society* 5:53.

Bambra, C. 2007. Going beyond The three worlds of welfare capitalism: Regime theory and public health research. *Journal of Epidemiology and Community Health* 61:1098–1102.

Bambra, Clare. 2011. Health inequalities and welfare state regimes: Theoretical insights on a public health 'puzzle'. *Journal of Epidemiology and Community Health* 65:740–745.

Bambra, C., and T. A. Eikemo. 2009. Welfare state regimes, unemployment and health. A comparative study of the relationship between unemployment and self-reported health in 23 European countries. *Journal of Epidemiology and Community Health* 63:92–98.

Bambra, Clare, Gopalakrishnan Netuveli and Terje A. Eikemo. 2010. Welfare state regime life courses: The development of western European welfare state regimes and age-related patterns of educational inequalities in self-reported health. *International Journal of Health Services* 40:399–420.

Bambra, C., D. Pope, V. Swami, D. Stanistreet, A. Roskam, A. Kunst and A. Scott-Samuel. 2009. Gender, health inequalities and welfare state regimes: a cross-national study of 13 European countries. *Journal of Epidemiology and Community Health* 63:38–44.

Banks, James, Michael Marmot, Zoe Oldfield and James P. Smith. 2006. Disease and disadvantage in the United States and in England. *JAMA* 295:2037–2045.

Beauchamp, A., K. Backholer, D. Magliano and A. Peeters. 2014. The effect of obesity prevention interventions according to socioeconomic position. A systematic review. *Obesity Reviews* 15:541–554.

Beckfield, Jason, and Nancy Krieger. 2009. Epi + demos + cracy: linking political systems and priorities to the magnitude of health inequities—evidence, gaps, and a research agenda. *Epidemiologic Reviews* 31:152–177.

Beckfield, Jason, Clare Bambra, Terje A. Eikemo, Tim Huijts, Courtney McNamara and Claus Wendt. 2015. An institutional theory of welfare state effects on the distribution of population health. *Social Theory and Health* 13:227–244.

Benyamini, Yael. 2011. Why does self-rated health predict mortality? An update on current knowledge and a research agenda for psychologists. *Psychology and Health* 26:1407–1413.

Bergqvist, Kersti, Monica A. Yngwe and Olle Lundberg. 2013. Understanding the role of welfare state characteristics for health and inequalities—an analytical review. *BMC Public Health* 13:1234.

Bonoli, Giuliano. 1997. Classifying welfare states. A two-dimension approach. *Journal of Social Policy* 26:351–372.

Braveman, Paula, and Laura Gottlieb. 2014. The social determinants of health: It's time to consider the causes of the causes. *Public Health Reports* 129:19–31.

Braveman, Paula, Susan Egerter and David R. Williams. 2011. The social determinants of health: Coming of age. *Annual Review of Public Health* 32:381–398.

Brennenstuhl, Sarah, Amélie Quesnel-Vallée and Peggy McDonough. 2012. Welfare regimes, population health and health inequalities: A research synthesis. *Journal of Epidemiology and Community Health* 66:397–409.

Bryan, Mark L., and Stephen P. Jenkins. 2016. Multilevel modelling of country effects. A cautionary tale. *European Sociological Review* 32:3–22.

Castañeda, Heide, Seth M. Holmes, Daniel S. Madrigal, Maria-Elena D. Young, Naomi Beyeler and James Quesada. 2015. Immigration as a social determinant of health. *Annual Review of Public Health* 36:375–392.

Chung, Haejoo, and Carles Muntaner. 2006. Political and welfare state determinants of infant and child health indicators. An analysis of wealthy countries. *Social Science and Medicine* 63:829–842.

Chung, Haejoo, and Carles Muntaner. 2007. Welfare state matters: A typological multilevel analysis of wealthy countries. *Health Policy* 80:328–339.

Chung, Haejoo, Edwin Ng, Selahadin Ibrahim, Björn Karlsson, Joan Benach, Albert Espelt and Carles Muntaner. 2013. Welfare state regimes, gender, and depression: A multilevel analysis of middle and high income countries. *International Journal of Environmental Research and Public Health* 10:1324–1341.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt and René Algesheimer. 2019. How to obtain comparable measures for cross-national comparisons. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00598-7.

Coburn, David. 2000. Income inequality, social cohesion and the health status of populations. The role of neo-liberalism. *Social Science and Medicine* 51:135–146.

Cylus, Jonathan, Maria Glymour and Mauricio Avendano. 2015. Health effects of unemployment benefit program generosity. *American Journal of Public Health* 105:317–323.

Dahl, Espen, and Gunn E. Birkelund. 1997. Health inequalities in later life in a social democratic welfare state. *Social Science and Medicine* 44:871–881.

Davis, Karen. 1991. Inequality and access to health care. *The Milbank Quarterly* 69:253.

Diez-Roux, A. V. 2000. Multilevel analysis in public health research. *Annual Review of Public Health* 21:171–192.

Dragano, Nico, Johannes Siegrist and Morten Wahrendorf. 2011. Welfare regimes, labour policies and unhealthy psychosocial working conditions. A comparative study with 9917 older employees from 12 European countries. *Journal of Epidemiology and Community Health* 65:793–799.

Eikemo, T. A., and C. Bambra. 2008. The welfare state: A glossary for public health. *Journal of Epidemiology and Community Health* 62:3–6.

Eikemo, T. A., C. Bambra, K. Joyce and Espen Dahl. 2008a. Welfare state regimes and income-related health inequalities. A comparison of 23 European countries. *European Journal of Public Health* 18:593–599.

Eikemo, Terje A., Clare Bambra, Ken Judge and Kristen Ringdal. 2008b. Welfare state regimes and differences in self-perceived health in Europe: A multilevel analysis. *Social Science and Medicine* 66:2281–2295.

Eikemo, Terje A., Martijn Huisman, Clare Bambra and Anton E. Kunst. 2008c. Health inequalities according to educational level in different welfare regimes. A comparison of 23 European countries. *Sociology of Health and Illness* 30:565–582.

Elo, Irma T. 2009. Social class differentials in health and mortality. Patterns and explanations in comparative perspective. *Annual Review of Sociology* 35:553–572.

Esping-Andersen, Gosta. 1990. *The three worlds of welfare capitalism*. London: Polity Press.

Esser, Ingrid, and Joakim Palme. 2010. Do public pensions matter for health and wellbeing among retired persons? Basic and income security pensions across 13 Western European countries. *International Journal of Social Welfare* 19:S103–S120.

Farrants, Kristin. 2017. Recommodification and the social determinants of health. Unemployment benefits, pensions and health inequalities in Sweden and England, 1991–2011. *Journal of Public Health* 39:661–667.

Fenger, H. J. Menno. 2007. Welfare regimes in Central and Eastern Europe: Incorporating post-communist countries in a welfare regime typology. *Contemporary Issues and Ideas in Social Sciences* 3:1–30.

Ferrera, Maurizio. 1996. The 'Southern Model' of welfare in social Europe. *Journal of European Social Policy* 6:17–37.

Galobardes, Bruna, Mary Shaw, Debbie A. Lawlor, John W. Lynch and George Davey Smith. 2006. Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health* 60:7–12.

Garbarski, Dana. 2016. Research in and prospects for the measurement of health using self-rated health. *Public Opinion Quarterly* 80:977–997.

Gkiouleka, Anna, Tim Huijts, Jason Beckfield and Clare Bambra. 2018. Understanding the micro and macro politics of health. Inequalities, intersectionality & institutions—A research agenda. *Social Science and Medicine* 200:92–98.

Goerres, Achim, Markus B. Siewert and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Granados, José A. T. 2010. Politics and health in eight European countries. A comparative study of mortality decline under social democracies and right-wing governments. *Social Science and Medicine* 71:841–850.

Gravelle, H. 1998. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *British Medical Journal* 316:382–385.

Guarnizo-Herreño, Carol C., Richard G. Watt, Hynek Pikhart, Aubrey Sheiham and Georgios Tsakos. 2013. Socioeconomic inequalities in oral health in different European welfare state regimes. *Journal of Epidemiology and Community Health* 67:728–735.

Guarnizo-Herreño, Carol C., Richard G. Watt, Hynek Pikhart, Aubrey Sheiham and Georgios Tsakos. 2014. Inequalities in oral impacts and welfare regimes. Analysis of 21 European countries. *Community Dentistry and Oral Epidemiology* 42:517–525.

Halfon, Neal, and Miles Hochstein. 2002. Life course health development. An integrated framework for developing health, policy, and research. *The Milbank Quarterly* 80:433–479.

Hamad, Rita, and David H. Rehkopf. 2016. Poverty and child development: A longitudinal study of the impact of the earned income tax credit. *American Journal of Epidemiology* 183:775–784.

Hank, Karsten, and Anja Steinbach. 2019. Families and their institutional contexts: The role of family policies and legal regulations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00603-z.

Henrich, Joseph, Steven J. Heine and Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466:29.

Hill, Sarah, Amanda Amos, David Clifford and Stephen Platt. 2014. Impact of tobacco control interventions on socioeconomic inequalities in smoking. Review of the evidence. *Tobacco Control* 23:e89–97.

Huber, J., and A. Mielck. 2010. Morbidität und Gesundheitsversorgung bei GKV- und PKV-Versicherten. Forschungsstand empirischer Studien (Comparing statutory and private health insurance in terms of morbidity and healthcare. A stocktake of empirical research). *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 53:925–938.

Hurrelmann, Klaus, Katharina Rathmann and Matthias Richter. 2011. Health inequalities and welfare state regimes. A research note. *Journal of Public Health* 19:3–13.

Jen, Min H., Kelvyn Jones and Ron Johnston. 2009. Compositional and contextual approaches to the study of health behaviour and outcomes: Using multi-level modelling to evaluate Wilkinson's income inequality hypothesis. *Health and Place* 15:198–203.

Jones, Koolman, and van Doorslaer. 2006. The impact of having supplementary private health insurance on the use of specialists. *Annales d'Économie et de Statistique* 83/84:251–275.

Julià, Mireia, Laia Ollé-Espluga, Christophe Vanroelen, Deborah de Moortel, Sarah Mousaid, Stig Vinberg, Vanessa Puig-Barrachina, Esther Sánchez, Carles Muntaner, Lucía Artazcoz and Joan Benach. 2017. Employment and labor market results of the SOPHIE project: Concepts, analyses, and policies. *International Journal of Health Services* 47:18–39.

Kangas, Olli. 2010. One hundred years of money, welfare and death. Mortality, economic growth and the development of the welfare state in 17 OECD countries 1900–2000. *International Journal of Social Welfare* 19:S42–S59.

Karim, Syahirah A., Terje A. Eikemo and Clare Bambra. 2010. Welfare state regimes and population health: integrating the East Asian welfare states. *Health Policy* 94:45–53.

Klein, Jens, Kerstin Hofreuter-Gätgens, Daniel Lüdecke, Margit Fisch, Markus Graefen and von dem Knesebeck, Olaf. 2016. Socioeconomic status and health-related quality of life among patients with prostate cancer 6 months after radical prostatectomy: A longitudinal analysis. *BMJ Open* 6:e010968.

Larrimore, Jeff. 2011. Does a higher income have positive health effects? Using the earned income tax credit to explore the income-health gradient. *The Milbank Quarterly* 89:694–727.

Leibfried, Stephan. 1992. Towards a European Welfare State? On integrating poverty regimes into the European community. In *Social policy in a changing world*, eds. Zsuzsa Ferge and Jon E. Kolberg, pp. 245–279. Frankfurt a. M.: Campus.

Leigh, Andrew, Christopher Jencks and Smeeding, Timothy M. 2011. Health and economic inequality. In *The Oxford handbook of economic inequality*, eds. Brian Nolan, Wiemer Salverda and Timothy M. Smeeding, pp. 384–404. Oxford: Oxford University Press.

Levecque, Katia, Ronan van Rossem, Katrien de Boyser, Sarah van de Velde and Piet Bracke. 2011. Economic hardship and depression across the life course: The impact of welfare state regimes. *Journal of Health and Social Behavior* 52:262–276.

Levin, Kate A. 2003. Study design VI—Ecological studies. *Evidence-based dentistry* 7:108.

Leão, Teresa, Inês Campos-Matos, Clare Bambra, Giuliano Russo and Julian Perelman. 2018. Welfare states, the Great Recession and health. Trends in educational inequalities in self-reported health in 26 European countries. *PloS one* 13:e0193165.

Lima-Costa, Maria F., Cibele C. Cesar, Dora Chor and Fernando A. Proietti. 2012. Self-rated health compared with objectively measured health status as a tool for mortality risk screening in older adults: 10-year follow-up of the Bambuí Cohort Study of Aging. *American Journal of Epidemiology* 175:228–235.

Link, Bruce G., and Jo Phelan. 1995. Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior* extra issue, 80–94.

Lorenc, Theo, and Kathryn Oliver. 2014. Adverse effects of public health interventions. A conceptual framework. *Journal of Epidemiology and Community Health* 68:288–290.

Lorenc, Theo, Mark Petticrew, Vivian Welch and Peter Tugwell. 2013. What types of interventions generate inequalities? Evidence from systematic reviews. *Journal of Epidemiology and Community Health* 67:190–193.

Lorenzoni, Luca, Annalisa Belloni and Franco Sassi. 2014. Health-care expenditure and health policy in the USA versus other high-spending OECD countries. *The Lancet* 384:83–92.

Lundberg, Olle. 2008. Commentary: Politics and public health—some conceptual considerations concerning welfare state characteristics and public health outcomes. *International Journal of Epidemiology* 37:1105–1108.

Lynch, John, George D. Smith, Sam Harper, Marianne Hillemeier, Ross, Nancy, Kaplan, George A and Michael Wolfson. 2004. Is income inequality a determinant of population health? Part 1. A Systematic Review. *The Milbank Quarterly* 82:5–99.

Macinko, James A., Leiyu Shi, Barbara Starfield and John T. Wulu. 2003. Income inequality and health: A critical review of the literature. *Medical Care Research and Review* 60:407–452.

Mackenbach, Johan P. 2012. The persistence of health inequalities in modern welfare states: The explanation of a paradox. *Social Science and Medicine* 75:761–769.

Maio, Fernando de. 2012. Advancing the income inequality—health hypothesis. *Critical Public Health* 22:39–46.

Makransky, Guido, Christina W. Schnohr, Torbjørn Torsheim and Candace Currie. 2014. Equating the HBSC Family Affluence Scale across survey years. A method to account for item parameter drift using the Rasch model. *Quality of Life Research* 23:2899–2907.

Marmot, Michael, and Jessica J. Allen. 2014. Social determinants of health equity. *American Journal of Public Health* 104 Suppl 4:S517–9.

McGill, Rory, Elspeth Anwar, Lois Orton, Helen Bromley, Ffion Lloyd-Williams, Martin O'Flaherty, David Taylor-Robinson, Maria Guzman-Castillo, Duncan Gillespie, Patricia Moreira, Kirk Allen, Lirije Hyseni, Nicola Calder, Mark Petticrew, Martin White, Margaret Whitehead and Simon Capewell. 2015. Are interventions to promote healthy eating equally effective for all? Systematic review of socioeconomic inequalities in impact. *BMC Public Health* 15:457.

McKinlay, James. 1979. A case for refocusing upstream: The political economy of illness. In *Patients, physicians and illness: A sourcebook in behavioral science and health*, ed. J. Gartley, pp. 9–25. New York: Free Press.

Moor, Irene, Jacob Spallek and Matthias Richter. 2017. Explaining socioeconomic inequalities in self-rated health: A systematic review of the relative contribution of material, psychosocial and behavioural factors. *Journal of Epidemiology and Community Health* 71:565–575.

Moortel, Deborah de, Laia Palència, Lucía Artazcoz, Carme Borrell and Christophe Vanroelen. 2015. Neo-Marxian social class inequalities in the mental well-being of employed men and women. The role of European welfare regimes. *Social Science and Medicine* 128:188–200.

Muntaner, Carles, Joan Benach, Haejoo Chung, N. G. Edwin and Ted Schrecker. 2010. Welfare state, labour market inequalities and health. In a global context: An integrated framework. SESPAS report 2010. *Gaceta sanitaria* 24 Suppl 1:56–61.

Muntaner, Carles, Carme Borrell, Edwin Ng, Haejoo Chung, Albert Espelt, Maica Rodriguez-Sanz, Joan Benach and Patricia O'Campo. 2011. Politics, welfare regimes, and population health: Controversies and evidence. *Sociology of Health and Illness* 33:946–964.

Murray, Christopher J.L., Gakidou, Emmanuela, E. and Julio Frenk. 1999. Health Inequalities and Social Group Differences: What Should We Measure? *Bulletin of the World Health Organization* 77:537–543.

Navarro, Vicente. 1999. The political economy of the welfare state in developed capitalist countries. *International Journal of Health Services* 29:1–50.

Navarro, Vicente, and Leiyu Shi. 2001. The political context of social inequalities in health. *International Journal of Health Services* 31:1–21.

Navarro, Vicente, Carme Borrell, Joan Benach, Carles Muntaner, Agueda Quiroga, Maica Rodríguez-Sanz, Nuria Vergés, Jordi Gumá and M. I. Pasarín. 2003. The importance of the political and the social in explaining mortality differentials among the countries of the OECD, 1950–1998. *International Journal of Health Services* 33:419–494.

Navarro, Vicente, Carles Muntaner, Carme Borrell, Joan Benach, Águeda Quiroga, Maica Rodríguez-Sanz, Núria Vergés and M. I. Pasarín. 2006. Politics and health outcomes. *The Lancet* 368:1033–1037.

Ng, Edwin, Carles Muntaner and Haejoo Chung. 2016. Welfare states, labor markets, political dynamics, and population health: A time-series cross-sectional analysis among East and Southeast Asian nations. *Asia-Pacific Journal of Public Health* 28:219–231.

Niedzwiedz, Claire L., Srinivasa V. Katikireddi, Jill P. Pell and Richard Mitchell. 2014a. Socioeconomic inequalities in the quality of life of older Europeans in different welfare regimes. *European Journal of Public Health* 24:364–370.

Niedzwiedz, Claire L., Srinivasa V. Katikireddi, Jill P. Pell and Richard Mitchell. 2014b. The association between life course socioeconomic position and life satisfaction in different welfare states. European comparative study of individuals in early old age. *Age and Ageing* 43:431–436.

Niedzwiedz, Claire L., Jill P. Pell and Richard Mitchell. 2015. The relationship between financial distress and life-course socioeconomic inequalities in well-being. Cross-national analysis of European welfare states. *American Journal of Public Health* 105:2090–2098.

Nolte, Ellen, and McKee, Martin. 2004. *Does health care save lives? Avoidable mortality revisited*. London: Nuffield Trust.

Norström, Thor, and Joakim Palme. 2010. Public pension institutions and old-age mortality in a comparative perspective. *International Journal of Social Welfare* 19:S121–S130.

O'Campo, Patricia, Agnes Molnar, Edwin Ng, Emilie Renahy, Christiane Mitchell, Ketan Shankardass, Alexander St John, Clare Bambra and Carles Muntaner. 2015. Social welfare matters: A realist review of when, how, and why unemployment insurance impacts poverty and health. *Social Science and Medicine* 132:88–94.

Oliver, Adam, and Elias Mossialos. 2004. Equity of access to health care. Outlining the foundations for action. *Journal of Epidemiology and Community Health* 58:655–658.

Or, Zeynep, Florence Jusot and Yilmaz, Engin. 2008. *Impact of health care system on socioeconomic inequalities in doctor use*. IRDES Working Paper No 17.

Pega, Frank, Kristie Carter, Ichiro Kawachi, Peter Davis, Fiona I. Gunasekara, Olle Lundberg and Tony Blakely. 2013. The impact of in-work tax credit for families on self-rated health in adults: A cohort study of 6,900 New Zealanders. *Journal of Epidemiology and Community Health* 67:682–688.

Pega, Frank, Kristie Carter, Ichiro Kawachi, Peter Davis and Tony Blakely. 2014. The impact of an unconditional tax credit for families on self-rated health in adults: Further evidence from the cohort study of 6,900 New Zealanders. *Social Science and Medicine* 108:115–119.

Pega, Frank, Tony Blakely, Maria M. Glymour, Kristie N. Carter and Ichiro Kawachi. 2016. Using marginal structural modeling to estimate the cumulative impact of an unconditional tax credit on self-rated health. *American Journal of Epidemiology* 183:315–324.

Pfaff, Holger, and Timo-Kolja Pförtner. 2016. Der Beitrag der Versorgungsforschung zur Chancengerechtigkeit in der Gesundheit und in der Versorgung. (The contribution made by healthcare research to equal opportunities in health and healthcare) *Das Gesundheitswesen* 78:91–96.

Phelan, Jo C., Bruce G. Link and Parisa Tehranifar. 2010. Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. *Journal of Health and Social Behavior* 51:28–40.

Pickett, Kate E., and Richard G. Wilkinson. 2015. Income inequality and health: A causal review. *Social Science and Medicine* 128:316–326.

Ratharina, Katharina, Irene Moor, Anton E. Kunst, Nico Dragano, Timo-Kolja Pförtner, Frank J. Elgar, Klaus Hurrelmann, Lasse Kannas, Tibor Baška and Matthias Richter. 2016. Is educational differentiation associated with smoking and smoking inequalities in adolescence? A multilevel analysis across 27 European and North American countries. *Sociology of Health and Illness* 38:1005–1025.

Rathmann, Katharina, Veronika Ottova, Klaus Hurrelmann, Margarethe de Looze, Kate Levin, Michal Molcho, Frank Elgar, Saoirse N. Gabhainn, van Dijk, Jitse P and Matthias Richter. 2015. Macro-level determinants of young people's subjective health and health inequalities: A multilevel analysis in 27 welfare states. *Maturitas* 80:414–420.

Reeves, Aaron, Martin McKee, Johan Mackenbach, Margaret Whitehead and David Stuckler. 2017. Public pensions and unmet medical need among older people. Cross-national analysis of 16 European countries, 2004–2010. *Journal of Epidemiology and Community Health* 71:174–180.

Regidor, Enrique, Cruz Pascual, David Martínez, María E. Calle, Paloma Ortega and Paloma Astasio. 2011. The role of political and welfare state characteristics in infant mortality. A comparative study in wealthy countries since the late 19th century. *International Journal of Epidemiology* 40:1187–1195.

Richter, Matthias, Katharina Rathman, Saoirse Nic Gabhainn, Alessio Zambon, William Boyce and Klaus Hurrelmann. 2012. Welfare state regimes, health and health inequalities in adolescence: A multilevel study in 32 countries. *Sociology of Health and Illness* 34:858–879.

Rodgers, Gerry B. 1979. Income and inequality as determinants of mortality: An international cross-section analysis. *Population Studies* 33:343–351.

Rogers, Everett M. 2005. *Diffusion of innovations*. 5th ed. New York: Free Press.

Romaniuk, Joanna. 2014. Subjective and objective health of the population aged 50+ and the welfare state regimes in 16 European countries. *International Journal of Management and Economics* 39:82–107.

Saint-Arnaud, Sébastien, and Paul Bernard. 2003. Convergence or resilience? A hierarchical cluster analysis of the welfare regimes in advanced countries. *Current Sociology* 51:499–527.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schneider, Silke L. 2010. Nominal comparability is not enough. (In-)equivalence of construct validity of cross-national measures of educational attainment in the European Social Survey. *Research in Social Stratification and Mobility* 28:343–357.

Schröder, Martin. 2019. Varieties of capitalism and welfare regime theories: Assumptions, accomplishments, and the need for different Methods. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00609-7.

Strully, Kate W., David H. Rehkopf and Ziming Xuan. 2010. Effects of prenatal poverty on infant health. State earned income tax credits and birth weight. *American Sociological Review* 75:534–562.

Subramanian, S. V., and Ichiro Kawachi. 2004. Income inequality and health: What have we learned so far? *Epidemiologic Reviews* 26:78–91.

Thomson, Katie, Clare Bambra, Courtney McNamara, Tim Huijts and Adam Todd. 2016. The effects of public health policies on population health and health inequalities in European welfare states: Protocol for an umbrella review. *Systematic Reviews* 5:57.

Vahid Shahidi, Faraz, Arjumand Siddiqi and Carles Muntaner. 2016. Does social policy moderate the impact of unemployment on health? A multilevel analysis of 23 welfare states. *European Journal of Public Health* 26:1017–1022.

van de Velde, Sarah, Clare Bambra, Koen van der Bracht, Terje A. Eikemo and Piet Bracke. 2014. Keeping it in the family. The self-rated health of lone mothers in different European welfare regimes. *Sociology of Health and Illness* 36:1220–1242.

Victora, Cesar G., J. Patrick Vaughan, Fernando C. Barros, Anamaria C. Silva and Elaine Tomasi. 2000. Explaining trends in inequities. Evidence from Brazilian child health studies. *The Lancet* 356:1093–1098.

Wakefield, Jonathan. 2008. Ecologic studies revisited. *Annual Review of Public Health* 29:75–90.

van der Wel, Kjetil A., Espen Dahl and Karsten Thielen. 2012. Social inequalities in "sickness". Does welfare state regime type make a difference? A multilevel analysis of men and women in 26 European countries. *International Journal of Health Services* 42:235–255.

Wicks-Lim, Jeannette, and Peter S. Arno. 2017. Improving population health by reducing poverty. New York's earned income tax credit. *SSM—Population Health* 3:373–381.

Wilkinson, Richard G. 1990. Income distribution and mortality. A 'natural' experiment. *Sociology of Health and Illness* 12:391–412.

Wilkinson, Richard G., and Pickett, Kate. 2011. *Spirit level. Why equality is better for everyone.* London: Penguin Books Ltd.

Wilkinson, Richard G., and Kate E. Pickett. 2007. The problems of relative deprivation: Why some societies do better than others. *Social Science and Medicine* 65:1965–1978.

Witvliet, Margot I., Anton E. Kunst, Karien Stronks and Onyebuchi A. Arah. 2012. Assessing where vulnerable groups fare worst. A global multilevel analysis on the impact of welfare regimes on disability across different socioeconomic groups. *Journal of Epidemiology and Community Health* 66:775–781.

Woolf, Steven H., and Paula Braveman. 2011. Where health disparities begin: The role of social and economic determinants—and why current policies may make matters worse. *Health Affairs* 30:1852–1859.

Zambon, Alessio, Will Boyce, Ester Cois, Candace Currie, Patrizia Lemma, Paola Dalmasso, Alberto Borraccino and Franco Cavallo. 2006. Do welfare regimes mediate the effect of socioeconomic position on health in adolescence? A cross-national comparison in Europe, North America, and Israel. *International Journal of Health Services* 36:309–329.

Timo-Kolja Pförtner 1981, Dr. rer. Pol., research assistant, Institute of Medical Sociology, Health Services Research, and Rehabilitation Science of the University of Cologne (Germany). Areas of research: Social determinants of health, cross-national comparisons in health and health inequalities, work and health, poverty and relative deprivation, quantitative methods in empirical social research. Publications: In-work poverty and self-rated health in a cohort of working Germans: A hybrid approach for decomposing within-person and between-persons estimates of in-work poverty status. American Journal of Epidemiology 185, 2017 (with A. Schmidt-Catran); Socioeconomic inequalities in the impact of tobacco control policies on adolescent smoking. A multilevel study in 29 European countries. Addictive Behaviors 2016 (with A. Hublet, C.W. Schnohr, M. de Looze, T. Baška, M. Molcho, K. Lasse, A.E. Kunst, M. Richter); Widening inequalities in self-rated health by material deprivation? A trend analysis between 2001 and 2011 in Germany. Journal of Epidemiology and Community Health 70, 2016 (with F.J. Elgar).

Holger Pfaff 1956, Dr. phil., Professor for Quality Improvement and Evaluation in Rehabilitation at the University of Cologne, Director of the Institute for Medical Sociology, Health Services Research, and Rehabilitation Science (IMVR), Director of the Centre of Health Services Research Cologne (ZVFK). Areas of research: Care, context of care and quality, Organizations, work and health. Publications: Lehrbuch Versorgungsforschung: Systematik—Methodik—Anwendung. Stuttgart 2017 (ed. with E. Neugebauer, M. Schrappe, G. Glaeske); The relationship between social capital and quality management systems in European hospitals: A quantitative study. PLoS One; 8, 2013 (with A. Hammer, O. A. Arah, M. Der-Sarkissian, C. A. Thompson, R. Mannion, C. Wagner, O. Ommen, R. Sunol); Who does it first? The uptake of medical innovations in the performance of thrombolysis on ischemic stroke patients in Germany: a study based on hospital quality data. Implementation Science 10, 2015 (with N. Scholten, H. C. Lehmann, G. R. Fink, U. Karbach).

Frank J. Elgar 1973, PhD, associate professor of psychiatry, Institute for Health and Social Policy, McGill University (Montreal, Canada). Areas of research: Social and structural determinants of adolescent health, health inequalities, school violence, relative deprivation, adolescent mental health. Publications: Relative deprivation and disordered gambling in youths. Journal of Epidemiology and Community Health 72, 2018 (with N. Canale, M. Wohl, M. Lenzi, A. Vieno); Early-life income inequality and adolescent health and well-being. Social Science and Medicine 174, 2017 (with G. Gariepy, T. Torsheim, C. Currie); Socioeconomic inequalities in adolescent health 2002–2010: a time-series analysis of 34 countries participating in the Health Behaviour in School-aged Children study. The Lancet 385, 2015 (with T. K. Pförtner, I., Moor, B. De Clercq, G. Stevens, C. Currie).

Kölner Zeitschrift für Soziologie und Sozialpsychologie

Policy Effects on Political Engagement

Conrad Ziller

Abstract Public policies addressing societal problems typically aim to change citizens' behaviors and attitudes. While scholars frequently link cross-national differences in specific policies with individuals' attitudinal or behavioral outcomes, the specific operating mechanisms often remain veiled. The policy feedback literature provides an explanatory framework for how policies affect citizens' political orientations and behavior, which in turn has an impact on subsequent policy developments. This article provides an overview of policy feedback mechanisms, and reviews comparative empirical studies that link policy indicators and forms of individual-level political engagement. As illustrative examples, I have focused on social policies and immigrant integration policies as two widely publicized policy areas. Since the identification of policy effects is challenging for a number of methodological reasons, I also discuss advances related to empirical design.

Keywords Policy feedback · Welfare state · Social policy · Immigrant integration policy · Political support

Einfluss von Policy auf politisches Engagement

Zusammenfassung Politische Maßnahmen, die gesellschaftliche Probleme angehen, zielen in der Regel auf Verhaltens- und Einstellungsänderungen der Bürgerinnen und Bürger ab. Auch werden innerhalb der empirisch-vergleichenden politischen Soziologie länderübergreifende Unterschiede in Institutionen und Policies zur Erklärung von Einstellungs- und Verhaltensunterschieden genutzt. Allerdings werden dabei die Vermittlungsmechanismen zwischen Politik und Individuum oft-

C. Ziller (✉)
Institute of Sociology and Social Psychology, University of Cologne
Albertus-Magnus-Platz, 50923 Cologne, Germany
E-Mail: ziller@wiso.uni-koeln.de

⚫ Springer

mals nicht hinreichend spezifiziert und getestet. Die Policy-Feedback-Literatur bietet einen Theorierahmen, um Policies mit politischen Orientierungen und Verhaltens-weisen zu verknüpfen. In dem vorliegenden Artikel wird ein Überblick über Policy-Feedback-Mechanismen und entsprechende empirische Studien gegeben. Zur weite-ren Verdeutlichung werden vergleichende Untersuchungen zur Wirkung von wohl-fahrtsstaatlicher Sozialpolitik sowie Integrationspolitik auf politisches Engagement vorgestellt und diskutiert. Da die Identifizierung kausaler Effekte im Rahmen der Po-licy-Forschung eine Herausforderung darstellt, wird abschließend auf methodische Weiterentwicklungen eingegangen.

1 Introduction

Public policy can be defined as the outcome of political processes designed and implemented in order to address societal problems. Policymakers have an interest in the effectiveness of the policies that they implement, and thus try to anticipate the specific functioning and potential effects at the policy design stage (Knill and Tosun 2012). For example, the enactment of smoking bans in the German *Länder* (states) in the mid-to-late 2000s aimed to protect workers in bars and restaurants from second-hand smoking and its deleterious effects on health. Empirical evaluations of the effectiveness of these policies indeed find evidence of improved health among non-smokers in states where smoking has been restricted in public areas, bars, and restaurants (Kuehnle and Wunder 2017). Whether intended or not, anti-smoking policies may also shape public opinion towards smokers. Pacheco (2013) finds that people in US states which have implemented smoking bans (compared to those that have not) began to view second-hand smoking as more harmful and to regard smokers in increasingly negative terms. These changes in attitudes also yielded an effect on how receptive people were towards additional smoking restrictions. The way that various policies influence (political) attitudes and behaviors of individuals and interest groups, which then shape subsequent courses of policy development, is the central topic of the policy feedback literature (Campbell 2012; Mettler and Soss 2004; see also Schmitt-Beck 2019).

This article examines how public policies influence people's political engage-ment, which includes policy-related attitudes, attitudes towards the political system, and political behavior such as voting. I have focused specifically on the following overarching questions: Through which mechanisms do policies affect political en-gagement? What conditions can be identified that reinforce or attenuate a policy impact on political engagement? Do policies affect target populations only, or the general public as well? And, how best to empirically assess policy effects in com-parative research? To illustrate these questions substantially, I focus on two policy areas that are highly publicized and can be convincingly linked to public prefer-ences. First, I focus on social welfare policies and social security programs and relate them to the political engagement of both target groups and members of the

general population. Welfare is one of most frequently studied policy areas, as corresponding social policy measures usually have a direct impact on citizens' needs and resources (for welfare regime theories see also Schröder 2019). Moreover, changes in social policies are widely publicized through political debate and media coverage, and a high degree of visibility enhances the potential for linking changes in policy with changes in citizens' behaviors and attitudes (Campbell 2012, p. 338).

As a second policy area, this review looks at immigrant integration policies which regulate immigrants' rights and access to participate in economic, political, cultural, and social life. Immigrant integration is a critical and timely topic that has engendered extensive empirical research on the potential for governing integration outcomes. Moreover, immigrant integration is an issue that is symbolically charged and has been widely publicized. Again, this enables the study of policy effects on both immigrants, as the target group, and the general public.

In terms of theoretical framework, I draw on the policy feedback approach and related arguments (Campbell 2012; Mettler and Soss 2004). Policy feedback posits that enacted policies influence citizens' attitudes and behaviors, which in turn "feed back" (sometimes also referred to as "feed forward" to emphasize temporal ordering) to influence subsequent courses of policy formulation and implementation. Hence, policy and citizens' attitudes and behaviors are explicitly considered to influence each other in reciprocal causal processes. While this is a theoretically plausible assumption, reversed causality represents a source of endogeneity that should be taken into account in research designs, as it potentially biases the results from empirical models.

In the remainder of the article, I first outline the theoretical framework (Sect. 2). Second, I summarize and discuss findings from comparative studies that inform debates on how policy relates to political engagement (Sect. 3). The article concludes by discussing methodological issues of comparative research and potential advancements.

2 Conceptual Framework

2.1 Theories of Political Engagement

Political engagement can be broadly defined as cognitive and emotional involvement in political matters, which manifests itself in individual political interest, political knowledge, political opinions, or political attitudes (Barrett and Brunton-Smith 2014, p. 6). In broad conceptualizations (e. g., Zukin et al. 2006), political engagement also encompasses forms of political participation, which refers to citizens' actions aimed at influencing decisions of public representatives and officials (Brady 1999, p. 738; Verba et al. 2002). Specifically, this includes voting,[1] running for office, contacting politicians, membership in political and civic organizations, and non-conventional activities such as protesting. Applying a broad definition of political engagement

---

[1] For determinants of electoral participation from the perspective of comparative political systems, see also the chapters by Spies and Franzmann (2019) and Schmitt-Beck (2019) in this special issue.

entailing both attitudinal and behavioral dimensions is also suitable for a review of existing studies that differ in conceptual terms.

As explanatory approaches of political engagement, sociological, psychological, economic, and political theories can be distinguished (Druckman and Lupia 2000; Norris 2002). According to traditional sociological accounts, individual political orientations and participation are explained by people's positioning within a social structure and the influence of social networks (Lazarsfeld et al. 1948). If group interests and communication processes in social networks shape political views, then patterns of political engagement should be largely homogeneous within social groups, while cleavages predominantly exist between groups. Moreover, societal transformation processes (e.g., modernization, expanding educational opportunities, value change, or digitalization) have become increasingly relevant factors in explaining why and how citizens participate (Inglehart 1997; Welzel and Dalton 2014). The psychological perspective on political attitudes and behavior emphasizes the relevance of individual differences, including personality traits, education, partisan identification, and political efficacy (e.g., Campbell et al. 1960). Similarly, economic accounts conceptualize political preferences and participation as a product of individual choices, driven by self-interest, as well as rational and goal-directed behavior (Downs 1957; but see Sears and Funk 1991).

Political theories focus on the role of political institutions, organized interests, mobilization, and communication flows. A number of accounts argue that mobilization by political organizations and social groups critically determines individuals' political behavior (Verba et al. 2002). Over and above this, political actors and organizations influence political views by determining the framing of issues and communication flows that citizens receive (Nelson and Kinder 1996; Zaller 1992). The key point propounded by political accounts is that political engagement does not simply emerge from individual characteristics or social group membership, but is essentially shaped by political processes and how they are received by the public (Dalton 2014; Druckman and Lupia 2016). Such political processes refer not only to institutional rules, expressive government actions, political communication flows, and mobilization efforts on the part of political actors, but also policies that implement incentives and sanctions, or convey informational content.

2.2 Classic Accounts of Policy Feedback

According to common conceptualizations, politics refers to the process that determines "who gets what, when, how" (Lasswell 1936). Policies are outcomes of politics, while the term polity refers to structural aspects of a political system and its power relations. The basic idea of policy feedback is that policies not only represent a result of politics, but also have the potential to transform the political process and thus affect subsequent courses of policy development. In short, the policy feedback literature investigates how "new policies create new politics" (Schattschneider 1935). Policies may create feedback by affecting elites and bureaucrats, mass publics, as well as specific target groups. Mechanisms through which policy effects operate include the restructuring of authority and power relations, the redistribution of resources, and the informational or normative reframing of preferences and identities.

Classic work on policy feedback has largely focused on the ways in which social policies and social security programs generate constituencies of supporters, which in turn affect subsequent policy development. In *Protecting Soldiers and Mothers,* Theda Skocpol (1992) examined welfare provision from a historical perspective, showing that nineteenth and early twentieth century social programs for veterans and women in the US had a strong impact on subsequent developments in social welfare. In his seminal study entitled *Dismantling the Welfare State*, Pierson (1994) shows that attempts towards welfare retrenchment under the Reagan administration in the US provoked considerable mobilization among senior citizens who were opposed to scaling back pensions. Consequently, social security measures related to pensions remained largely intact. In contrast, cutbacks in the UK pension system under Thatcher were more easily accomplished, given that it was fragmented into different programs. This meant that retrenchment was not opposed by a well-organized group of beneficiaries. An important lesson to be learned from Pierson's classic work is that each policy creates its own constraints on change by virtue of its specific structure.[2] This implies that in order for policy to change, the support (or opposition) of citizens plays a critical role—in addition to the support of public officials and interest groups, and beyond mere path dependency (i.e., costs associated with policy shifts). Programs that create sufficient support among beneficiaries are likely to be left in place or even expanded, while those that do not do so might be cut.

As operating mechanisms, Pierson identifies resource distribution and symbolic politics as two major drivers in garnering public support for policies as well as political engagement. Resource effects refer to redistributive elements of specific policies and to the material self-interest of individuals and groups that benefit from them. Symbolic or interpretative effects refer to the informational and normative content of policies that influences people's attitudes and behavior. Pierson's study focuses on the retrenchment of social policies, and represents a theoretical framework rather than a comprehensive system-level theory. Nonetheless, the specified feedback conditions and mechanisms linking policy and citizens' behaviors and attitudes go far beyond earlier system approaches such as Easton's diagram of the political system. Moreover, it is a simple task to transfer the policy feedback approach to policy areas other than welfare. In this vein, Pierson's work has inspired numerous studies investigating how various policies affect preferences, demands, and the political involvement of citizens and organized interest groups (see Larsen 2018; Mettler and Soss 2004; Mettler 2015 as overviews).

2.3 Mechanisms Linking Policy and Political Engagement

A first pathway through which policies influence citizens' political attitudes and behavior is *resource effects*. Policy enforcement mechanisms include financial incentives, payments, fines and penalties, and changes in taxation that alter the distribution of goods and services but also burdens. Individuals who gain material profit from

---

[2] Similar conclusions were reached by Campbell (2003), who focused on age-related programs and senior citizens' political activation.

a policy that is implemented are expected to become interested in how to secure benefits beyond future policy change, and thus in political affairs more generally. For example, the introduction or expansion of a pension system increases the availability of financial resources for senior citizens. This, in turn, shapes their perception of government being responsive to their needs, while at the same time they become politically active when policy retrenchment may threaten their gains (Campbell 2012). As another example, training programs for the long-term unemployed may improve their work-related skills and chances of re-employment, which would then equip them with resources and improve their political efficacy and political engagement (Soss 1999). Following this line of reasoning, material resource effects have been related to citizens' political orientations and behavior (e.g., trust in institutions, satisfaction with government, and political participation), beliefs about social justice, and group-based attitudes about deservingness (Mettler and Soss 2004).

Second, laws and policies may affect citizens by conveying information and normative cues, usually labeled as *interpretive effects* (also symbolic or cognitive effects; Pierson 1994). Individuals who are exposed to policy content may not only become engaged because of material gain, but also because policies signal that government addresses citizens' concerns. Rather than material interest, here, the main driver is the experience of government being responsive and elevating one's status. This makes individuals more likely to respond with a sense of duty and obligation, which in turn facilitates political activation, political support, and civic engagement. As an example, Mettler (2002) shows that veterans receiving welfare benefits through the G.I. Bill[3] not only profit in terms of monetary resources, but also change the perception of their social status and how government values their contributions.

Third, policy content may operate at the aggregate level as *social norms* signaling socially acceptable and rewarding behaviors. For example, an indoor smoking ban can shape social norms with regard to smoking, and breaching this norm imposes not only a legal sanction (e.g., payment of a fine), but also brings social disapproval and reputational damage as social consequences. A potential pathway through which policies shape social norms is framing and messaging in public debates and mass media content. Politicians express their support for or opposition to specific policy measures on a regular basis. Moreover, symbolic language use identifies and defines public images of societal groups and their deservingness (Stewart 2012). Socially shared stereotypes and group-related stigmatization have real-life consequences for the self-perception of target groups and intergroup relations more generally. The social construction of target populations also affects the distribution of benefits and burdens in policy designs, as well as how members of social groups are treated by public officials (Schneider and Ingram 1993).

Fourth, *recent debates* on policy feedback mechanisms examine, for example, the degree to which policy designs motivate self-interest versus collective aims (Jacobs and Mettler 2018). As another example, studies on the link between policy and

---

[3] G.I. refers to soldiers in the United States Army. The G.I. bill (enacted as "Servicemen's Readjustment Act" of 1944) regulated a broad range of benefits for returning World War II veterans, such as compensational payments in the case of unemployment.

citizens' policy support have mostly focused on self-reinforcing feedback processes (i.e., positive feedback), where the resource or interpretive content of a policy fosters constituents' policy support, which in turn leads to policy stability or expansion. More recently, possible self-undermining effects in which policy support reduces over time have been studied more extensively (i.e., negative feedback). A basic explanation for negative feedback is that policies simply fulfill their intended purpose, which leads to eroding public demand.[4] Alternatively, burdens and unexpected losses or the salience of alternative policies may undermine public support for specific policies over time (Jacobs and Weaver 2015). Looking at a policy–opinion link as a function of time, short- and long-term effects may even diverge in such a way that the introduction of a policy stimulates opposition in the short term, which then turns into policy support in the long run (e.g., Naumann 2014).

2.4 Conditions of Policy Feedback

Policy effects might vary considerably across groups or political circumstances (Patashnik and Zelizer 2013). A broad approach to conditional policy feedback is to distinguish group-specific and mass public effects, depending on the scope and range of policy designs. *Group-specific effects* refer to a policy-induced change in orientations and behaviors of target groups. Depending on their degree of power and their social construction, Schneider and Ingram (1993) distinguish between four types of target groups. Groups that are advantaged have a relatively large amount of power and a positive social reputation; contenders are high in power, but negatively constructed; dependents are low in power, but rather positively constructed; and deviants have a low level of power and a negative public image. At the stage of policy design, the advantaged typically receive mostly benefits from a policy and few burdens. Contenders receive medium benefits and few (but highly visible) burdens. Dependents receive limited benefits and few but typically hidden burdens, while deviants receive limited to no benefits and a high share of burdens. According to this classification, the amount of power and the public image of a target group operate as conditions of policy feedback, which already determines the design and outputs of policy measures and thus yields consequences for potential policy feedback effects.

With regard to *mass public effects*, Soss and Schram (2007) categorize policies according to the dimensions of proximity and visibility in a four-field matrix.[5] Visibility describes the degree to which a policy is salient to the public. Political communication flows and mass media coverage are important determinants for making policy content more visible. Proximity refers to citizens' exposure to policy measures, either directly as personal experience, or indirectly via social networks or

---

[4] This argument is highlighted in public opinion research, such as on the thermostatic model of representation (Wlezien 1995), according to which public demands signal governments about desired policies. Policies meeting public demand then lead to an adjustment of public preferences, and would ultimately result in a long-term equilibrium of representation (e.g. Brooks and Manza 2006; Page and Shapiro 1983).

[5] There are many other ways to classify policies, for example according to their consequences (e.g., generating costs and benefits) or governance principles (see Knill and Tosun 2012, Chapter 2). Nonetheless, the dimensions of visibility and proximity appear to be particularly useful when studying policy effects on citizens.

Visibility high

Policy effects on citizens mediated by
mass media;
E.g., foreign policies with high media
coverage, immigration policy

Policy effects driven by media dis-
course and personal experience;
E.g., social security policies, family poli-
cies, education policies

Proximity low   *Proximity high*

Limited/no policy effects;
E.g., sector-specific work regulations,
distant foreign policies

Policy effects driven by personal
experience;
E.g., health and safety regulations,
consumer protection

Visibility low

Fig. 1 Proximity and visibility as conditions of policy effects on citizens. (Modified from Soss and Schram 2007)

social contexts in which individuals are embedded. The more individuals and groups of a society are (potentially) affected by policy measures, the higher is the proximity of a policy. If visibility and proximity are both low, there are likely to be no policy effects on citizens' behaviors and attitudes. Under high visibility and low proximity, policy effects are likely to occur through political communication in the mass media. If both visibility and proximity are high, effects on citizens' behaviors and attitudes are likely to be driven by both media discourse and personal experience, whereas low visibility and high proximity heighten the role of personal experience (see Hedegaard and Larsen 2014, p. 272). Fig. 1 summarizes how the conditions of proximity and visibility relate to feedback effects and lists some illustrative policy examples.

Additional conditions emphasized in empirical studies include attributes of persons (e. g., political interest, socioeconomic status, and political sophistication), individual performance evaluations of policy measures, and contextual factors (e. g., political accountability, political alternatives, institutions, administrative efficiency, and economic contexts; see Kumlin and Stadelmann-Steffen 2014). To what extent the single factors attenuate or amplify political attitudes and participation depends on the specific processes under study. For example, political interest may contribute to people's opinion formation through gathering information on enacted policies. Whether this results in positive or negative evaluations then hinges upon additional characteristics such as the benefits expected to ensue from a respective policy measure.

2.5 Methodological Issues

In an overview of the state of the policy feedback literature, Mettler (2015) identifies a number of methodological limitations. First, empirical studies on policy feedback largely focus on social welfare policies and should be extended to cover a wider range of public policies. Second, previous studies mainly look at the policy impact among groups targeted by policy design. However, the question of whether policies yield effects on other societal groups and the general public should receive more attention. For example, policies may induce unintended consequences by heightening perceptions of deprivation of groups that compete for benefits with a target group. Third, greater methodological sophistication should improve causal inference from empirical models testing policy feedback assumptions. One important issue in this regard is selection, meaning that people with specific political orientations opt into particular social security programs due to underlying and unobserved characteristics, which would obscure causal inference of program participation on political attitudes. What is more, the concept of policy feedback posits that policies generate constituencies of supporters that in turn affect subsequent courses of policy development. This implies processes of reciprocal causality, where policies and public opinion affect each other. This might even reflect the political reality better than a unidirectional model would. However, reverse causality potentially biases statistical results from analyses of observational data, especially if these data are cross-sectional. Specifically, regression analysis typically assumes exogeneity of predictor variables (Verbeek 2004). This assumption is violated through omitted variables or reciprocal causation leading to biased estimates (see also Schmidt-Catran et al. 2019). In terms of methodological advancement, an increasing number of studies address issues that are related to selection and endogeneity by using longitudinal and experimental designs (e. g., Bechtel et al. 2017; Häusermann et al. 2018; Ziller and Helbling 2017).

3 Review of Empirical Studies

The review of empirical studies focuses on social policies (Sect. 3.1) and immigrant integration policies (Sect. 3.2). Both policy areas are suitable for illustrating policy effects on citizens' political engagement because they are characterized by high visibility through media discourse and public debate.[6] Integration policies are symbolically charged due to high levels of immigration into Western societies. Moreover, welfare measures such as unemployment protection, healthcare, and pensions affect large segments of society, and thus have a high level of proximity. Within these areas of research, I have selected empirical studies that (i) link policy and individuals' political engagement,[7] (ii) apply a comparative (cross-national) research design, and (iii) have attracted scholarly interest as indicated by a comparatively large number of

---

[6] Larsen (2018) provides an empirical review of feedback effects and also discusses policies other than welfare and immigrant integration.

[7] This includes policy-specific attitudes, general political support, as well as forms of political behavior.

citations (per-year average). To illustrate different mechanisms of policy feedback, the selected studies cover policy effects on both target populations and the general population. Rather than providing a broad-brush overview of as many studies as possible, the aim is for the selected studies to conjointly inform central debates on how (social welfare and immigrant integration) policy measures are linked to political engagement.

3.1 Social Policies

Social (welfare) policies are designed and implemented to address social problems related to unemployment, poverty, family instability, inadequate healthcare, and the needs of the elderly. Depending on the range and scope of welfare systems, this includes the taxation and redistribution of wealth, labor market policies, policies and insurance related to health, family services, education, and pensions. A large number of studies investigating social policy effects have been conducted in the US context (e. g., Campbell 2003; Mettler 2002; Soss 1999) or single European countries (Svallfors 1997). As a focal point, these works look at how participation in social security programs relates to participants' political support and engagement. While single-country studies typically compare outcomes of individuals' participating in different programs, comparative cross-national studies use coarsened (and thus comparative) policy indicators such as social spending (see Larsen 2018 as an overview). Kumlin and Haugsgjerd's (2017) literature review on the welfare state and political trust reveals that many comparative studies examine macroeconomic country determinants (e. g., income inequality or economic growth) of political orientations, but only few investigate the role of specific social policies.

   The following review covers studies addressing implications of social policy design (e. g., means-tested versus universal) for target groups (3.1.1). Another central debate refers to how welfare state generosity affects political support among the general public, and what role is played by anticipated benefits and costs (3.1.2).

3.1.1 Social Policy and the Role of Policy Design

Social policies and social security programs typically concern the redistribution of resources and can be classified as means-tested or universal policies, depending on whether they aim at the entitlement of particular target groups or are widely accessible. Countries with largely means-tested or low-income-targeting welfare programs produce relatively small constituencies of beneficiaries compared to countries with universalistic measures. Since broad middle-class political support is necessary for large-scale income redistribution, countries with selective welfare measures tend to produce greater poverty and inequality than universalistic ones do—a phenomenon that has become known as the "paradox of redistribution" (Korpi and Palme 1998). Although the prevalence of this paradox has recently been contested (Brady and Bostic 2015), studies suggest that at the level of specific programs, participants in means-tested programs respond with political inactivity rather than activation (Mettler 2002; Soss 1999).

In addition to the distributional characteristic of whether programs are designed as universal or means-tested, another important dimension is the degree of conditionality. Conditional programs contain a direct or indirect prescription of how to behave, and a violation of rules affects the provision of welfare. In such programs, caseworkers become paternalistic supervisors that oversee the fulfillment of obligations and enforce sanctions in case of violation (e. g., benefit reduction or removal). Bruch et al. (2010) present longitudinal empirical evidence that the authority–client relations of social security programs affect the political participation of clients in the US context: The more paternalist the program, the stronger the negative effects on political participation. In contrast, in programs with a low degree of paternalism, low-income groups become more engaged in civic and political activities. These results suggest that it is authority–client relations within programs that are the critical triggers, and not the question of whether a program is means tested (e. g., it targets the poor) or universal. Focusing on underlying mechanisms, Watson (2015) examines longitudinal evidence on beneficiaries of conditional social security programs in the United Kingdom. She finds that conditional programs suppress political participation, political interest, and political efficacy, whereas rights-based programs tend to foster political activation. Watson also finds suggestive evidence that the negative effect of conditionality is mitigated for programs that are universal rather than means tested.

Shifting the focus from specific program structures to the institutional and societal context of policy implementation, Jordan (2013) examines institutional bases of welfare state support for the policy areas healthcare, pensions, and unemployment. Linking indicators of inclusiveness (e. g., proportion of the population covered) and survey measures of social policy support, the author finds a positive and robust association when comparing 17 Western countries. Highly inclusive welfare regimes thus appear to generate broader segments of supporters compared to redistributive or means-tested ones, which in turn is associated with higher levels of specific and general support for social policies.

Rehm et al. (2012) analyze how societal configurations affect welfare state support. Specifically, the authors argue that levels of welfare support are lower and more polarized in countries where low income and unemployment risks are strongly correlated (compared to contexts in which these factors are less closely related). The reason is that when risks are concentrated among the poor, this will shrink the circle of beneficiaries (and thus supporters) to a rather small societal segment. Using comparative macrodata for 13 European countries and a survey conducted in the US, the authors find lower levels of welfare support in contexts with highly interrelated disadvantage (i. e., low income) and risk (i. e., unemployment risk). Although not explicitly tested by Rehm et al., it is plausible to assume that policies addressing economic disadvantage or individual risks may contribute, at least to some extent, toward broadening the basis of welfare support among the general population.

3.1.2 Social Policy Effects Among the General Public

While universal social policies affect large segments of a society, means-tested programs are less relevant for the general population in terms of personal exposure.

Nevertheless, support for social policies related to health, age, and unemployment can also be expected to ensue from individuals who are currently not benefiting, because they anticipate relying on them in the future (Andersen 1992). In other words, there should be substantially high levels of welfare state support also among the general population, while at the same time cost–benefit considerations might be relevant, especially among those who shoulder a disproportionately large share of the costs. In a similar vein, Van Oorschot et al. (2012) examine the relationship between welfare state generosity (measured as social spending) and citizens' welfare state evaluations in Europe. The authors argue that the degree of welfare state generosity not only leads to perceptions of positive effects (e. g., mitigation of individual risks, reduction of poverty, etc.), but that it also increases awareness of possible negative consequences (e. g., higher tax burden). As research design, the authors use multilevel models relating national social spending to individual-level perceptions of positive social consequences (e. g., poverty reduction), negative economic consequences (e. g., place strain on economy), and negative moral consequences (e. g., make people lazy). The results show systematic positive associations between social expenditures and all three outcomes, with the effect being greater for positive than for negative consequences, meaning that positive consequences in high spending contexts appear to outweigh negative implications.

Giger (2012) investigates the link between social policy retrenchment and government support. Drawing on the policy feedback literature and assuming that social policies are largely popular, cutbacks should increase government unpopularity. Moreover, this should particularly be the case for individuals who regard welfare as a salient and important political issue. She tests these arguments by linking respondents' ratings of the performance of the incumbent government (obtained from the Comparative Study of Electoral Systems) with welfare cutbacks in pensions, unemployment benefits, and healthcare (obtained from data about welfare reforms from the International Social Security Association database). Using Bayesian multilevel regression models on 19 elections (18 countries of the Organisation for Economic Co-operation and Development, OECD), the author finds no systematic average relationships among the general population. However, when looking at respondents who state that welfare is an important issue, pension and healthcare retrenchment are systematically related to lower government ratings.

Kumlin (2014) focuses on degrees of generosity of unemployment benefits and their predictive capacity for citizens' satisfaction with democracy. Comparing 11 Western countries over five points in time, the author finds support for a positive link between generosity and satisfaction with democracy. Moreover, high levels of unemployment or high salience of unemployment as a political issue attenuates the relationship. This hints at the role played by visible costs as a conditioning factor of policy feedback, especially when looking at the general public. With regard to individual risk factors, Gingrich and Ansell (2012) focus on employment protection legislation as moderator of the link between individuals' experiences of labor market risks and welfare state support. Using survey data on 19 OECD countries, respondents' social policy support in the areas of unemployment protection, healthcare, and industrial aid serve as outcome variables. Education and occupation-related skills serve as individual-level indicators of unemployment risk. Results

from multilevel regression models show significant negative interactions between individual risk factors and employment protection policy. This indicates that social protection policies diminish the relevance of individuals' risk as a determinant of policy preferences. Alternatively, the results can also be interpreted in such a way that employment protection laws increase social policy support particularly for those who potentially profit from them (i. e., people with few skills or without higher education), while these laws are less conducive for those with low individual risks.

3.1.3 Summary of Core Results

The review of empirical studies focused on the role of social policy design and implications for target groups and the general public (see Table 1 as an overview). Apart from the distinction between means-tested and universal social policies, a number of studies emphasize the critical role of conditional versus rights-based elements of policy design as well as authority–client relations that shape recipients' political engagement. In terms of broader public policy support, the societal distribution of risks and the degree of inclusiveness (i. e., proximity) appear to operate as conditioning factors.

The reviewed studies on mass public effects included both institutional, structural, and individual factors as conditions of how social policy relates to political engagement. At the same time, social policies not only produce beneficiaries, they also imply costs for taxpayers. Applying the argument of self-interest, citizens should take into account (potential) benefits and anticipated costs. The studies reviewed tend to provide empirical support for the relevance of both cost evaluation (e. g., Van Oorschot et al. 2012; Jordan 2013; Kumlin 2014) and anticipated benefits in terms of societal and individual configurations and risk factors (e. g., Gingrich and Ansell 2012; Rehm et al. 2012). Furthermore, sociotropic versus altruistic orientations of citizens may reflect additional factors that condition social policy effects on political engagement (Jacobs and Mettler 2018).

In terms of methodology, the selected studies specify macro–micro hypotheses, and partly tackle potential methodological shortcomings. Studies comparing participants across different policy programs are potentially plagued by selection bias, meaning that, for example, people who are politically inactive opt into paternalistic welfare programs due to unobserved characteristics. The two reviewed studies on program participation applied longitudinal data methods in order to minimize this concern. For studies applying cross-national comparisons, unobserved heterogeneity between countries is a serious concern of potential bias. Even studies using multiple time points do not always control for time-constant differences using country fixed effects specifications. Another methodological issue that is rarely considered in the study of policy effects among the general public is degrees of salience or knowledge of policy measures (but see Giger 2012 and Kumlin 2014). Including measures of salience and knowledge would nonetheless improve on current practices of measurement (Campbell 2012; Ziller 2014).

Table 1 Summary of reviewed studies—social policy

Study	Countries covered	Time points covered	Macro indicators, moderator	Outcome/indicator of political engagement	Method	Main results
Bruch et al. (2010)	1 (20 US cities)	3	Type of social welfare programs	Voting, membership in political organizations, membership in civic organizations	Binary/ordered logistic regression	Participation in paternalist government assistance programs predicts lower levels of political engagement
Watson (2015)	1	19	Structure of social welfare program: Degree of conditionality	Political participation (voting, membership in civic organizations), political interest, political efficacy	Pooled OLS and fixed effects models on panel data	Recipients of conditional welfare benefits have lower levels of political participation, political interest, and political efficacy
Jordan (2013)	15–17	1	Inclusiveness of policy sector healthcare, pensions, and unemployment insurance	Support for healthcare, pensions, and unemployment insurance	Ordered probit regression with robust standard errors clustered at the country level	Degree of inclusiveness is positively associated with social policy support
Rehm et al. (2012)	13	1	Country-specific correlation between income and unemployment risk	Support for unemployment benefits, support for social policies	Generalized ordered logit models to generate macro indicators; OLS with macro-level indicators	Lower welfare support in contexts where low income and unemployment risk are closely related
Van Oorschot et al. (2012)	25	1	Social spending per capita	Negative economic consequences of social policy, negative moral consequences, positive social consequences	Linear multilevel regression (country random intercept, no random slopes)	Social expenditure is particularly related to higher rates of perceived positive social consequences of welfare activity

Table 1 (Continued)

Study	Countries covered	Time points covered	Macro indicators, moderator	Outcome/indicator of political engagement	Method	Main results
Giger (2012)	18	1	Social policy retrenchment (pension, unemployment, healthcare), author's coding using data about welfare reforms from the International Social Security Association database; Moderator: issue salience welfare domain	Rating of incumbent government performance	Bayesian multilevel regression (country random intercept, no random slopes)	Pension and healthcare retrenchment are systematically related to lower government ratings, but only for those with corresponding issue salience
Kumlin (2014)	11	5	Unemployment benefits	Satisfaction with democracy	Linear multilevel regression (country random intercept, no random slopes)	Generous unemployment benefits predict higher levels of satisfaction with democracy; negative moderation with salience and prevalence of country-specific unemployment
Gingrich and Ansell (2012)	19	1	OECD database indicator on employment protection legislation (e. g., regulation on temporary employment); Moderators: education; occupation-related skill	Attitudes towards government spending (unemployment protection, healthcare, industrial aid)	Logistic multilevel regression with country fixed effects and robust standard errors (country random intercept, no random slopes)	Negative significant interaction between skill and employment protection policy; that is, social protection policies make individuals' exposure to risk a less critical determinant of policy preferences

OLS Ordinary Least Squares, *OECD* Organisation for Economic Co-operation and Development

3.2 Immigrant Integration Policies

Immigrant integration is a process in which immigrants adopt language and cultural skills, participate in the educational system and labor market, develop social ties with members of the receiving society, participate politically, and increasingly identify with the society in which they live (Freeman 2004; see Careja 2019). Immigrant integration policies target immigrant populations by determining specific rights and obligations.[8] In their most general form, citizenship regimes can be classified as ethnic or civic, depending on whether access to citizenship is inherited by ethnic group membership, or whether other modes of citizenship acquisition exist (Brubaker 2012). Further conceptualizations incorporate a cultural dimension and distinguish between collectivistic-ethnic, collectivistic-civic, and individualistic-civic regimes (Greenfeld 1999). The breakdown into collectivistic and individualistic thereby reflects the degree to which immigrants are expected to assimilate into the receiving society and give up distinctive cultural characteristics.

Similarly, the degree to which ethnic and cultural differences are politically accepted (or even promoted) is at the center of assimilationist versus multicultural understandings of integration (Banting and Kymlicka 2013). While classic studies cluster countries categorically into integration regimes, recent approaches allow for a continuous policy classification along numerous sub-dimensions of integration (Helbling 2013). Typically, these indicators range from restrictive (or limited rights) to liberal (or permissive rights), and contain sub-dimensions such as individual and cultural rights (e. g., Index of Citizenship Rights for Immigrants, ICRI; Koopmans and Michalowski 2012), as well as political, cultural, labor, family, and education-related domains of integration (e. g., Migrant Integration Policy Index, MIPEX; Huddleston et al. 2015). In addition, anti-discrimination policies have a much broader scope, as they aim at combating unequal treatment based on group membership including gender, ethnic origin, religion or belief, disability, age, or sexual orientation. Similar to the distinction of assimilationist versus multiculturalist, anti-discrimination measures may also be distinguished into color-blind measures focused on individual rights of equal treatment and positive action measures focused on overcoming existing inequalities (Ziller 2017). Beyond the regulation of immigrant rights, further indicators on immigration policies (e. g., access to countries) exist that are beyond the scope of this review (see Goodman 2015).

In terms of covered debates, this review includes studies that examine how policy design (multicultural versus assimilationist policies) relates to immigrants' political engagement (Sect. 3.2.1). Other studies focus on majority members' responses to integration policies in terms of political support and satisfaction with democracy (Sect. 3.2.2).[9]

---

[8] Depending on their range and scope, integration policies also address ethnic minority members, non-citizens in general, and people discriminated against based on group membership.

[9] As an adjacent debate, a number of comparative studies look at the relationship between immigrant integration policies and public opinion towards immigrants and immigration (e. g., Careja and Andreß 2013; Schlueter et al. 2013).

3.2.1 Integration Policy and Political Engagement of Immigrants

A core dimension of immigrant integration is the regulation of citizenship acquisition. Becoming a citizen not only means a change in legal status and entitlement to participate in politics (e.g., voting in national elections), but may also trigger processes of social learning and identity formation that are presumably positively linked to political engagement. This implies that when people feel that they are fully fledged members of a political community, they respond by taking a greater interest and becoming engaged in political affairs, and may even internalize the democratic ideals of active citizenship. Just and Anderson (2012) examine these contentions using comparative data on 19 European countries. Results from multilevel models and models including country fixed effects show that foreign-born populations are less likely to take part in conventional and non-conventional forms of political engagement compared to natives, while this gap in participation is particularly wide for non-citizens. Essentially, the findings suggest that citizenship acquisition promotes political engagement. But also here, the possibility of selection bias is an issue. It is quite plausible that immigrants who seek citizenship (compared to those who do not) are more highly motivated to engage in politics. To address this source of endogeneity, Just and Anderson apply an instrumental variables approach and find that their results are robust under this additional specification.[10]

Looking at specific integration policies, a major debate emphasizes potential negative effects of multicultural policies (compared to assimilationist approaches) on the integration outcomes of immigrants and ethnic minorities. Multicultural policies are "specific government policies designed to positively recognize diversity and help minorities maintain cultural and religious practices while integrating them into public life" (Wright and Bloemraad 2012, p. 78). Critics of cultural rights for immigrants and ethnic minorities suggest that multicultural policies do not create sufficiently strong incentives for immigrants to acquire the language of the receiving society and develop interethnic contacts (Koopmans 2010; 2013). Empirical evidence on this contention is rather mixed. Koopmans (2010) finds higher unemployment rates among immigrants, higher rates of residential segregation, and higher rates of immigrants involved in crimes in countries characterized by implemented multicultural (rather than assimilationist) policies combined with a generous welfare state (especially in the Netherlands, Sweden, and Belgium). Wright and Bloemraad (2012) use comparative survey data from pooled European Social Survey waves 2002–2008 and the United States Citizenship, Involvement, Democracy Survey, merged with policy data from the Citizenship Policy Index (CPI; Howard 2005) and the Multi-

---

[10] Similarly, Hainmueller et al. (2015) aim to isolate the causal effect of citizenship acquisition on political integration by using data on close naturalization referendums in Switzerland. Their results show that obtaining citizenship has a positive impact on voter turnout among immigrants, as well as on political efficacy and political knowledge.

culturalism Policy Index (MCP; Banting and Kymlicka 2013).[11] The authors examine immigrants' political trust, satisfaction with government, political interest, and political participation across three country clusters: countries with *restrictive* citizenship rights and *low* levels of multiculturalism; countries with *permissive* citizenship rights and *low* levels of multiculturalism; and countries with *permissive* citizenship rights and *high* levels of multiculturalism. By and large, the results show that absolute levels of immigrant political engagement, as well as gaps between immigrants and natives, are similar across the three country clusters, indicating a limited influence of political context.

As another example, Goodman and Wright (2015) test whether and how mandatory integration requirements[12]—intended to promote the assimilation of immigrants—affect immigrants' political orientations. Using data from the European Social Survey merged with policy data from the Civic Integration Policy Index (CIVIX; Goodman 2010), regression analysis on immigrant samples from 14 Western European countries is used to compare averages in social and political attitudes between countries of low versus high mandatory requirements. The results show little evidence of systematic differences across political contexts, which suggests that implemented civic integration requirements have limited impact on immigrants' political engagement and social integration.

3.2.2 Integration Policy and Natives' Political Attitudes

Beyond the scope on immigrants as target populations of integration policies, scholars are also interested in whether and how integration policies shape political attitudes and political support of mass publics. The underlying rationale here is that people who are skeptical of immigration respond with political discontent when political actors implement permissive integration policies. In contrast, citizens with a positive stance towards immigrant rights might respond with increased political support when permissive policies are enacted. Hooghe and De Vroome (2015) analyze comparative survey data from the European Social Survey merged with three different integration policy indicators (MIPEX, MCP, and ICRI). Using multilevel regression models, the authors find no empirical evidence of a systematic average relationship between integration policy indicators and political trust or satisfaction with government. Instead, they observe significant positive interactions of individual education and policy in the way that low (highly) educated groups respond with decreasing (increasing) political support to the implementation of permissive integration policies.

---

[11] The CPI index covers information on whether or not a country grants *jus soli* citizenship; the minimum years of residence required for naturalization; and whether or not naturalized immigrants are allowed to hold dual citizenship. The MCP index entails information on a variety of dimensions, including official affirmation of multiculturalism; multiculturalism in the school curriculum; inclusion of ethnic representation/sensitivity in public media; exemptions from dress codes for minorities; funding of ethnic organizations to support cultural activities; funding of bilingual and mother-tongue instruction; and affirmative action for disadvantaged immigrant groups.

[12] Such requirements include the acquisition of political knowledge, language skills, and value orientations.

McLaren (2015) examines immigrant integration policy (measured as MIPEX country scores) as a moderator of the link between anti-immigrant sentiment and political trust in Europe. People with strong anti-immigrant sentiment may tend to blame political actors for failing to protect society from any negative effects of immigration, and respond with lower political support. Permissive immigrant integration policies will thus additionally contribute to reducing political support for societal segments characterized by high levels of anti-immigrant sentiment. Using multilevel models on data from the European Social Survey, the author finds a significant negative cross-level interaction, indicating that in permissive policy contexts, concerns about immigration are more strongly related to lower levels of political trust than they are in restrictive contexts. Interpreted in a symmetric way, immigrant integration policies are related to lower levels of political trust for people whose anti-immigrant sentiment runs high, compared to those who have a low level of anti-immigrant sentiment. In a similar way, Citrin et al. (2014) examine anti-immigrant sentiment as a moderator of the link between multicultural policy in Europe (measured as MCP country scores) and individuals' political support (measured as political trust, satisfaction with democracy, and satisfaction with government). Using multilevel models on data from the European Social Survey and the European Values Study, the authors find that the effect of multiculturalism on political support is particularly negative for people with a high level of anti-immigrant sentiment.

Kesler and Bloemraad (2010) test an interaction between country-level immigration (measured by the share of the foreign-born population) and multicultural policies (measured by MCP data). Using multilevel models with country fixed effects on data from repeated survey waves of the World Values Survey, the authors find systematic positive interactions, especially when considering income inequality as an additional moderator. In other words, strong multicultural policies, as well as low levels of income inequality, mitigate the potential negative consequences of immigration-related ethnic diversity on political engagement.

Focusing on anti-discrimination measures as a sub-dimension of integration policy, Ziller and Helbling (2017) examine whether and how anti-discrimination policies affect political support in Europe for target groups and the general public. While an expansion of anti-discrimination laws should yield relevance for political support of individuals who actually face discrimination, effects among the general public should occur particularly for individuals who hold egalitarian values (and thus corresponding issue preferences). The authors use multilevel models on repeated cross-sectional survey data waves (Eurobarometer and European Social Survey), merged with time-varying information on anti-discrimination laws (MIPEX) and policy knowledge (aggregated country-year survey responses). Country fixed effects tackle unobserved time-constant country differences and exploit over-time variance at the country-year level. The results show that individuals who are discriminated against express particularly high levels of political support in contexts where anti-discrimination laws are expanded. Moreover, an increase in country-level policy knowledge predicts greater political support, especially for individuals high in egalitarianism.

Table 2 Summary of reviewed studies—immigrant integration policy

Study	Countries covered	Time points covered	Macro indicators, moderators	Outcome/indicator of political engagement	Method	Main results
Just and Anderson (2012)	19	1	Access to citizenship via individual citizenship status	Political engagement (contacting politicians, working for a political party or organization, wearing a campaign badge, signing a petition, boycotting products, protesting)	Linear multilevel regression (country random intercept, random slopes for minority status); country fixed effects and IV regression as robustness tests	Immigrants without citizenship participate less than immigrants with citizenship
Wright and Bloemraad (2012)	18	4	CPI, MCP	Political trust, satisfaction with government, political interest, political participation	OLS/logistic regression with cluster-robust standard errors at the country level	No substantial differences in political engagement between natives and immigrants across policy regimes
Goodman and Wright (2015)	14	6	CIVIX	Political interest, political efficacy	OLS/logistic regression	No substantial differences in political engagement between natives and immigrants across policy regimes
Hooghe and De Vroome (2015)	9–20	1	MIPEX, MCP, and ICRI; Moderator: education	Political trust, satisfaction with government	Linear multilevel regression (country random intercept, random slope for education)	No significant average effect of integration policies on political support; significant interaction: Lower (higher) educated groups respond with decreasing (increasing) political support to the implementation of permissive integration policies
McLaren (2015)	16	1	MIPEX; Moderator: anti-immigrant sentiment	Political trust (legal system, parliament, politicians)	Linear multilevel regression (country random intercept, no random slopes)	Significant negative interaction between anti-immigrant sentiment and immigrant integration policy: The gap in political support between pro- and anti-immigrant publics is wider in contexts with permissive integration policies

Table 2 (Continued)

Study	Countries covered	Time points covered	Macro indicators, moderators	Outcome/indicator of political engagement	Method	Main results
Citrin et al. (2014)	16	2	MCP; Moderator: anti-immigrant sentiment	Political trust, satisfaction with democracy, satisfaction with government	Linear multilevel regression (country random intercept, random slope for anti-immigrant sentiment); OLS with country fixed effects and cluster-robust standard errors as robustness test	Similar to McLaren (2015): Significant negative interaction between anti-immigrant sentiment and multiculturalism: The gap in political support between pro- and anti-immigrant publics is wider in contexts with strong multicultural policies
Kesler and Bloemraad (2010)	17–19	2–3	MCP	Organizational membership, non-electoral political actions (signing a petition, joining a boycott, and attending a lawful demonstration)	Logistic multilevel regression (country-year random intercept, no random slopes) with country fixed effects	Multicultural policies mitigate potential negative consequences of ethnic diversity on political engagement: Low levels of income inequality additionally mitigate negative consequences on political engagement
Ziller and Helbling (2017)	21	2, 4	Anti-discrimination policies (sub-index of MIPEX); Moderators: egalitarian values, discrimination	Evaluation of public administration, satisfaction with democracy, political trust	Linear multilevel regression (country-year random intercept, random slopes for moderator variables) with country fixed effects	Anti-discrimination measures increase political support for people suffering from discrimination; significant average effect of policy knowledge on political support, which is additionally amplified for people who hold egalitarian values

OLS Ordinary Least Squares, *IV regression* Instrumental variables regression, *CPI* Citizenship Policy Index, *MCP* Multiculturalism Policy Index, *CIVIX* Civic Integration Policy Index, *ICRI* Index of Citizenship Rights for Immigrants, *MIPEX* Migrant Integration Policy Index

3.2.3 Summary

The reviewed empirical studies are summarized in Table 2. For citizenship acquisition, the evidence reviewed suggests a positive effect on immigrants' political engagement. With regard to the distinction between multicultural and assimilationist policies, multiculturalism appears to be less relevant to immigrants' political attitudes than is suggested by critics, and may even mitigate any negative consequences of immigration on political engagement for the general population. While the empirical evidence of a catalyst function of naturalization appears to be robust, additional research employing longitudinal or experimental research designs is needed in order to further determine the effects of multiculturalist versus assimilationist policies.

Studies on mass publics merely find average relationships between integration policy and political attitudes. Instead, individual characteristics appear to critically moderate how policy translates into political support. People who feel easily threatened by newcomers respond with lower rates of political support, while those who harbor few such concerns increase their support. Hence, education, sentiment towards immigrants, and egalitarian values are group characteristics that decisively trigger how policy relates to political attitudes among the general public.

In terms of methodology, most of the studies reviewed control for confounding variables and conduct a number of robustness checks to tackle risks of endogeneity bias, while others apply longitudinal designs at the country-year level and include country fixed effects.

4 Conclusion

Policy feedback investigates "whether policies render citizens more or less engaged in politics and how public programs shape citizens' beliefs, preferences, demands, and power" (Mettler and Soss 2004, p. 60). Drawing on this framework, this article reviewed macro–micro mechanisms on how policy content affects people's political engagement, including resource distribution and informational and normative effects. While such mechanisms and conditions may serve as heuristic for various policy areas and outcome variables, I focused on social policy and immigrant integration policy and their implication for political attitudes and behavior as illustrative examples. Both policy areas are highly visible through public debates, and social security policies in particular concern large segments of society.

The results of a review of selected comparative empirical studies suggest that policy effects on political engagement are highly conditional upon third factors, including the policy design and features of administrative implementation (e.g., client–participant relations), the broader institutional and economic context (e.g., degrees of policy inclusiveness, distribution of risks, unemployment), and individual characteristics (e.g., individual risks, salience of policy issues, education, anti-immigrant sentiment). What became less explicit in the reviewed studies are the potential mechanisms through which policies affect individual political attitudes and behavior, such as resource distribution, information, and norms. Another topic that should have received greater attention in the studies reviewed is how political

engagement in turn feeds back and affects political processes, and how policy feedback evolves over longer periods of time, something which also entails dealing with reciprocal causality in empirical models.

Taking up Mettler's (2015) recommendation that policy areas other than welfare should be investigated, this review examined policy effects in the domain of immigrant integration. In terms of future research, it might be worthwhile to focus on sub-aspects of multicultural or integration policy in order to capture important variations in policy design, such as labor market integration policies (Kogan 2016) or family reunification (Gundelach and Manatschal 2017). As another example, Ziller (2017) finds that the policy effects of equal treatment regulations depend on their range and scope. While color-blind institutional fairness is related to increasing gaps in trust between natives and immigrants, more comprehensive anti-discrimination measures tend to help close gaps in political trust and support (see also Ziller and Helbling 2017).

In methodological terms, studies on participants in welfare programs are largely focused on citizens in the U.S. context (Mettler and Soss 2004). More recently, a number of cross-national studies examined policy effects due to increasingly available comparative country indicators. Cross-national comparisons typically apply multilevel analysis in order to assess how policy differences relate to individual outcomes, while at the same time accounting for non-independence due to the hierarchical data structure in which respondents are nested within countries (Snijders and Bosker 2012). A number of the studies reviewed apply multilevel analysis to cross-sectional data. However, results from analyses of cross-sectional observational data are prone to produce biased inference due to omitted variables. Longitudinal designs using panel data at the level of individuals and/or countries are able to control for confounding of omitted time-constant variables, and thus increase the validity of the results. Beyond this, advanced methods such as cross-lagged autoregressive structural equation models including fixed effects would account for both reciprocal causal relationships and the potential omission of time-constant confounding variables (Allison et al. 2017; Hamaker et al. 2015).

Among the studies reviewed, only a small fraction actually capitalizes on modeling over-time variations whilst controlling for cross-country differences. One reason for the lack of cross-national longitudinal designs is related to the availability of comparative policy indicators that include a sufficient number of countries and vary over time. For example, the Migration Policy Institute has only recently published time-series data for the MIPEX indicator, and other scholars have launched projects collecting time-varying data on immigration and immigrant integration policies. The IMPALA (International Migration Law and Policy Analysis; Beine et al. 2016) and IMPIC (Immigration Policies in Comparison; Helbling et al. 2017) projects collect data on immigration policies (e.g., entry, settlement, and regulation laws) for numerous countries and over time. Similarly, the collection of data on welfare state policies as already pursued by Giger (2012) should be extended to time-series data in order to stimulate longitudinal research designs in future studies.

At the same time, it is important to emphasize that cross-country comparison using multilevel analysis reflects only one analytical strategy of linking the political context to individual attitudes and behaviors. To improve on the causal identification

of policy effects, recent studies employ (quasi-)experimental designs. For example, Hopkins and Parish (2018) use difference-in-differences models on survey data collected before and after the Medicaid expansion in the US, and find an increase in health policy support especially among low-income Americans. Häusermann et al. (2018) use conjoint experiments to evaluate conditions of social policy retrenchment, and find that compensations to the relevant opposition groups increase their willingness to accept cutbacks. Bechtel et al. (2017) use conjoint experiments in order to analyze public opinion towards different policy designs for international bailouts. Using difference-in-differences models on social media data from the US, Flores (2017) finds that the introduction of a restrictive state law directed towards immigrants negatively affects public opinion towards them. Contrasting observational and experimental studies, there is of course a tradeoff that needs to be carefully taken into account between the causal identification of policy effects and the generalizability of the findings. A triangulation of empirical results using evidence from different methods and research designs can thus also be considered as a golden standard of research on policy feedback effects.

References

Allison, Paul D., Richard Williams and Enrique Moral-Benito. 2017. Maximum likelihood for cross-lagged panel models with fixed effects. *Socius* 3:1–17.

Andersen, Jørgen G. 1992. Sources of welfare-state support in Denmark: Self-interest or way of life? *International Journal of Sociology* 22:25–48.

Banting, Keith, and Will Kymlicka. 2013. Is there really a retreat from multiculturalism policies? New evidence from the multiculturalism policy index. *Comparative European Politics* 11:577–598.

Barrett, Martyn, and Ian Brunton-Smith. 2014. Political and civic engagement and participation: Towards an integrative perspective. *Journal of Civil Society* 10:5–28.

Bechtel, Michael M., Jens Hainmueller and Yotam Margalit. 2017. Policy design and domestic support for international bailouts. *European Journal of Political Research* 56:864–886.

Beine, Michel, Anna Boucher, Brian Burgoon, Mary Crock, Justin Gest, Michael Hiscox, Patrick McGovern, Hillel Rapoport, Joep Schaper and Eiko Thielemann. 2016. Comparing immigration policies: An overview from the IMPALA database. *International Migration Review* 50:827–863.

Brady, Henry E. 1999. Political participation. In *Measures of political attitudes*, eds. John P. Robinson, Philip R. Shaver, L. S. Wrightsman, pp. 737–801. San Diego: Academic Press.

Brady, David, and Amie Bostic. 2015. Paradoxes of social policy: Welfare transfers, relative poverty, and redistribution preferences. *American Sociological Review* 80:268–298.

Brooks, Clem, and Jeff Manza. 2006. Social policy responsiveness in developed democracies. *American Sociological Review* 71:474–494.

Brubaker, Rogers. 2012. *Citizenship and nationhood in France and Germany*. Cambridge: Harvard University Press.

Bruch, Sarah K., Myra M. Ferree and Joe Soss. 2010. From policy to polity: Democracy, paternalism, and the incorporation of disadvantaged citizens. *American Sociological Review* 75:205–226.

Campbell, Andrea L. 2003. *How policies make citizens: Senior political activism and the American welfare state*. Princeton, NJ: Princeton University Press.

Campbell, Andrea L. 2012. Policy makes mass politics. *Annual Review of Political Science* 15:333–351.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American voter*. New York: John Wiley and Sons.

Careja, Romana. 2019. Immigrants' labor market outcomes: Contributions from multilevel studies. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchen-

hauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00597-8.

Careja, Romana, and Hans-Jürgen Andreß. 2013. Needed but not liked—The impact of labor market policies on natives' opinions about immigrants. *International Migration Review* 47:374–413.

Citrin, Jack, Morris Levy and Matthew Wright. 2014. Multicultural policy and political support in European democracies. *Comparative Political Studies* 47:1531–1557.

Dalton, Russell J. 2014. *Citizen politics: Public opinion and political parties in advanced industrial democracies*. 6th edition. Los Angeles: Sage, CQ Press.

Downs, Anthony. 1957. *An economic theory of democracy*. New York: Harper and Row.

Druckman, James N., and Arthur Lupia. 2000. Preference formation. *Annual Review of Political Science* 3:1–24.

Druckman, James N., and Arthur Lupia. 2016. Preference change in competitive political environments. *Annual Review of Political Science* 19:13–31.

Flores, René D. 2017. Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter data. *American Journal of Sociology* 123:333–384.

Freeman, Gary P. 2004. Immigrant incorporation in Western democracies. *International Migration Review* 38:945–969.

Giger, Nathalie. 2012. Is social policy retrenchment unpopular? How welfare reforms affect government popularity. *European Sociological Review* 28:691–700.

Gingrich, Jane, and Ben Ansell. 2012. Preferences in context: Micro preferences, macro contexts, and the demand for social policy. *Comparative Political Studies* 45:1624–1654.

Goodman, Sara W. 2010. Integration requirements for integration's sake? Identifying, categorising and comparing civic integration policies. *Journal of Ethnic and Migration Studies* 36:753–772.

Goodman, Sara W. 2015. Conceptualizing and measuring citizenship and integration policy: Past lessons and new approaches. *Comparative Political Studies* 48:1905–1941.

Goodman, Sara W., and Matthew Wright. 2015. Does mandatory integration matter? Effects of civic requirements on immigrant socio-economic and political outcomes. *Journal of Ethnic and Migration Studies* 41:1885–1908.

Greenfeld, Liah. 1999. Is nation unavoidable? Is nation unavoidable today? In *Nations and National Identity*, ed. Hans P. Kriesi et al., pp. 37–54. Chur: Ruegger.

Gundelach, Birte, and Anita Manatschal. 2017. Ethnic diversity, social trust and the moderating role of subnational integration policy. *Political Studies* 65:413–431.

Hainmueller, Jens, Dominik Hangartner and Giuseppe Pietrantuono. 2015. Naturalization fosters the long-term political integration of immigrants. *Proceedings of the National Academy of Sciences* 112:12651–12656.

Hamaker, Ellen L., Rebecca M. Kuiper and Raoul P.P.P. Grasman. 2015. A critique of the cross-lagged panel model. *Psychological Methods* 20:102–116.

Häusermann, Silja, Thomas Kurer and Denise Traber. 2018. The politics of trade-offs: studying the dynamics of welfare state reform with conjoint experiments. *Comparative Political Studies*. 1–37. https://doi.org/10.1177/0010414018797943.

Hedegaard, Troels F., and Christian A. Larsen. 2014. How proximate and visible policies shape self-interest, satisfaction, and spending support: The case of public service production. In *How welfare states shape the democratic public: Policy feedback, participation, voting, and attitudes*, eds. Staffan Kumlin and Isabelle Stadelmann-Steffen, 269–288. Cheltenham: Edward Elgar.

Helbling, Marc. 2013. Validating integration and citizenship policy indices. *Comparative European Politics* 11:555–576.

Helbling, Marc, Liv Bjerre, Friederike Römer and Malisa Zobel. 2017. Measuring immigration policies: The IMPIC database. *European Political Science* 16:79–98.

Hooghe, Marc, and Thomas de Vroome. 2015. How does the majority public react to multiculturalist policies? A comparative analysis of European countries. *American Behavioral Scientist* 59:747–768.

Hopkins, Daniel J., and Kalind Parish 2018. The medicaid expansion and attitudes toward the affordable care act. Working Paper. http://dx.doi.org/10.2139/ssrn.2990576.

Howard, Marc M. 2005. Variation in dual citizenship policies in the countries of the EU. *International Migration Review* 39:697–720.

Huddleston, Thomas, Bilgili, Ozge, Joki, Anne-Linde and Vankova, Zvezda. 2015. *Migrant integration policy index 2015*. Barcelona/Brussels: CIDOB and MPG (www.mipex.eu).

Inglehart, Ronald. 1997. *Modernization and postmodernization*. Princeton, NJ: Princeton University Press.

Jacobs, Alan M., and R. Kent Weaver. 2015. When policies undo themselves: Self-undermining feedback as a source of policy change. *Governance* 28: 441–457.

Jacobs, Lawrence R., and Suzanne Mettler. 2018. When and how new policy creates new politics: Examining the feedback effects of the affordable care act on public opinion. *Perspectives on Politics* 16:345–363.

Jordan, Jason. 2013. Policy feedback and support for the welfare state. *Journal of European Social Policy* 23:134–148.

Just, Aida, and Christopher J. Anderson. 2012. Immigrants, citizenship and political action in Europe. *British Journal of Political Science* 42:481–509.

Kesler, Christel, and Irene Bloemraad. 2010. Does immigration erode social capital? The conditional effects of immigration-generated diversity on trust, membership, and participation across 19 countries, 1981–2000. *Canadian Journal of Political Science* 43: 319–347.

Knill, Christoph, and Jale Tosun. 2012. *Public policy. A new introduction*. Basingstoke, New York: Palgrave Macmillan.

Kogan, Irena. 2016. Integration policies and immigrants' labor market outcomes in Europe. *Sociological Science* 3:335–358.

Koopmans, Ruud. 2010. Trade-offs between equality and difference: Immigrant integration, multiculturalism and the welfare state in cross-national perspective. *Journal of Ethnic and Migration Studies* 36:1–26.

Koopmans, Ruud. 2013. Multiculturalism and immigration: A contested field in cross-national comparison. *Annual Review of Sociology* 39:147–169.

Koopmans, Ruud, Ines Michalowski and Stine Waibel. 2012. Citizenship rights for immigrants: National political processes and cross-national convergence in Western Europe, 1980–2008. *American Journal of Sociology* 11:1202–1245.

Korpi, Walter, and Joakim Palme. 1998. The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the Western countries. *American Sociological Review* 63:661–687.

Kuehnle, Daniel, and Christoph Wunder. 2017. The effects of smoking bans on self-assessed health: Evidence from Germany. *Health Economics* 26:321–337.

Kumlin, Staffan. 2014. Policy feedback in political context: Unemployment benefits, election campaigns, and democratic satisfaction. In *How welfare states shape the democratic public: Policy feedback, participation, voting, and political attitudes,* eds. Staffan Kumlin and Isabelle Stadelmann-Steffen, 181–197. Cheltenham: Edward Elgar.

Kumlin, Staffan, and Atle Haugsgjerd 2017. The welfare state and political trust: Bringing performance back. In *Handbook on political trust*. Sonja Zmerli and Tom Van der Meer, 285–301. Cheltenham: Edward Elgar.

Kumlin, Staffan, and Isabelle Stadelmann-Steffen. 2014. How welfare states shape the democratic public: Borrowing strength across research communities. In *How welfare states shape the democratic public: Policy feedback, participation, voting, and political attitudes,* eds. Staffan Kumlin and Isabelle Stadelmann-Steffen, 311–325. Cheltenham: Edward Elgar.

Larsen, Erik G. 2018. Policy feedback effects on mass publics: a quantitative review. *Policy Studies Journal*. https://doi.org/10.1111/psj.12280.

Lasswell, Harold D. 1936. *Politics: Who gets what, when, how*. New York: McGraw Hill.

Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet. 1948. *The people's choice: How the voter makes up his mind in a presidential campaign.* New York: Columbia University Press.

McLaren, Lauren M. 2015. *Immigration and perceptions of national political systems in Europe.* Oxford: Oxford University Press.

Mettler, Suzanne. 2002. Bringing the state back in to civic engagement: Policy feedback effects of the GI Bill for World War II veterans. *American Political Science Review* 96:351–365.

Mettler, Suzanne. 2015. Twenty years on: Paul Pierson's dismantling the welfare state? *PS: Political Science & Politics* 48:270–273.

Mettler, Suzanne, and Joe Soss. 2004. The consequences of public policy for democratic citizenship: Bridging policy studies and mass politics. *Perspectives on Politics* 2:55–73.

Naumann, Elias. 2014. 11. Raising the retirement age: retrenchment, feedback and attitudes. In *How welfare states shape the democratic public: Policy feedback, participation, voting, and political attitudes*, eds. Staffan Kumlin and Isabelle Stadelmann-Steffen, 223–243. Cheltenham: Edward Elgar.

Nelson, Thomas E., and Donald R. Kinder. 1996. Issue frames and group-centrism in American public opinion. *The Journal of Politics* 58:1055–1078.

Norris, Pippa 2002. *Political activism*. Cambridge: Cambridge University Press.

Pacheco, Julianna. 2013. Attitudinal policy feedback and public opinion: The impact of smoking bans on attitudes towards smokers, secondhand smoke, and antismoking policies. *Public Opinion Quarterly* 77:714–734.

Page, Benjamin I., and Robert Y. Shapiro. 1983. Effects of public opinion on policy. *American Political Science Review* 77:175–190.

Patashnik, Eric M., and Julian E. Zelizer. 2013. The struggle to remake politics: Liberal reform and the limits of policy feedback in the contemporary American state. *Perspectives on Politics* 11:1071–1087.

Pierson, Paul. 1994. *Dismantling the welfare state? Reagan, Thatcher, and the politics of retrenchment*. Cambridge: Cambridge University Press.

Rehm, Philipp, Jacob S. Hacker and Mark Schlesinger. 2012. Insecure alliances: Risk, inequality, and support for the welfare state. *American Political Science Review* 106:386–406.

Schattschneider, Elmer E. 1935. *Politics, pressures and the tariff. A study of free private enterprise in pressure politics, as shown in the 1929–1930 revision of the tariff*. New York: Prentice-Hall.

Schlueter, Elmar, Bart Meuleman and Eldad Davidov. 2013. Immigrant integration policies and perceived group threat: A multilevel study of 27 Western and Eastern European countries. *Social Science Research* 42:670–682.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schmitt-Beck, Rüdiger. 2019. Political systems and electoral behavior: a review of internationally comparative multilevel research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00608-8.

Schneider, Anne, and Helen Ingram. 1993. Social construction of target populations: Implications for politics and policy. *American Political Science Review* 87:334–347.

Schröder, Martin. 2019. Varieties of capitalism and welfare regime theories: Assumptions, accomplishments, and the need for different Methods. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00609-7.

Sears, David O., and Carolyn L. Funk. 1991. The role of self-interest in social and political attitudes. *Advances in Experimental Social Psychology* 24:1–91.

Skocpol, Theda. 1992. *Protecting soldiers and mothers: The political origins of social policy in the United States.* Cambridge, MA: Belknap Press of Harvard University Press.

Snijders, Tom A. B., and Roel J. Bosker. 2012. *Multilevel analysis. An introduction to basic and advanced multilevel modeling.* 2nd ed. Los Angeles: Sage.

Soss, Joe. 1999. Lessons of welfare: Policy design, political learning, and political action. *American Political Science Review* 93:363–380.

Soss, Joe, and Sanford F. Schram. 2007. A public transformed? Welfare reform as policy feedback. *American Political Science Review* 101:111–127.

Spies, Dennis Christopher, and Simon T. Franzmann. 2019. Party competition and vote choice. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00611-z.

Stewart, Julie. Ed. 2012. Fiction over facts: How competing narrative forms explain policy in a new immigration destination. *Sociological Forum* 27: 591–616.

Svallfors, Stefan. 1997. Worlds of welfare and attitudes to redistribution: A comparison of eight western nations. *European Sociological Review* 13:283–304.

van Oorschot, Wim, Tim Reeskens and Bart Meuleman. 2012. Popular perceptions of welfare state consequences: A multilevel, cross-national analysis of 25 European countries. *Journal of European Social Policy* 22:181–197.

Verba, Sidney., Kay L. Schlozman and Henry E. Brady. 2002. *Voice and equality. Civic voluntarism in American politics*. Cambridge, London: Harvard University Press.

Verbeek, Marno. 2004. *A guide to modern econometrics*. Chichester: Wiley.

Watson, Sara. 2015. Does welfare conditionality reduce democratic participation? *Comparative Political Studies* 48:645–686.

Welzel, Christian, and Russell J. Dalton. 2014. From allegiant to assertive citizens. In *The civic culture transformed: From allegiant to assertive citizens*, edited by Russell J. Dalton and Christian Welzel, 282–306. Cambridge: Cambridge University Press.

Wlezien, Christopher. 1995. The public as thermostat: Dynamics of preferences for spending. *American Journal of Political Science* 39:981–1000.

Wright, Matthew, and Irene Bloemraad. 2012. Is there a trade-off between multiculturalism and socio-political integration? Policy regimes and immigrant incorporation in comparative perspective. *Perspectives on Politics* 10:77–95.

Zaller, John R. 1992. *The nature and origins of mass opinion*. Cambridge: Cambridge University Press.

Ziller, Conrad. 2014. Societal implications of antidiscrimination policy in Europe. *Research & Politics* 1:1–9.

Ziller, Conrad. 2017. Equal treatment regulations and ethnic minority social trust. *European Sociological Review* 33:563–575.

Ziller, Conrad, and Marc Helbling. 2017. Antidiscrimination laws, policy knowledge and political support. *British Journal of Political Science* First View: 1–18. https://doi.org/10.1017/S0007123417000163

Zukin, Cliff, Scott Keeter, Molly Andolina, Krista Jenkins and Michael X. Delli Carpini. 2006. *A New engagement? Political participation, civic life, and the changing American citizen*. New York: Oxford University Press.

Conrad Ziller 1983, Dr. rer. pol., Vertretungsprofessor Chair for Empirical Social and Economic Research, Institute of Sociology and Social Psychology, University of Cologne. Areas of research: political sociology, migration and integration, and empirical methods. Publications: Antidiscrimination laws, policy knowledge and political support. British Journal of Political Science, first view 2017 (with M. Helbling); Equal treatment regulations and ethnic minority social trust. European Sociological Review 33, 2017; Ethnic diversity, economic and cultural contexts, and social trust: Cross-sectional and longitudinal evidence from European regions, 2002–2010. Social Forces 93, 2015.

Kölner Zeitschrift für Soziologie und Sozialpsychologie

Conditions and Consequences of Unequal Educational Opportunities in the Life Course: Results from the Cross-National Comparative *eduLIFE* Project

Hans-Peter Blossfeld · Nevena Kulic · Jan Skopek · Moris Triventi · Elina Kilpi-Jakonen · Daniela Vono de Vilhena · Sandra Buchholz

Abstract Using longitudinal data, this chapter studies the development of educational inequalities over the life course in 12–17 different industrialized societies. By comparing highly-standardized country case studies in specific phases of the educational career, it provides evidence of major communalities in modern societies. First, the cross-national findings show that educational inequalities are created and perpetuated in family settings, early in a child's life, long before children start school. Children from less privileged families are the ones who are least likely to attend high-quality institutions, and if they do, their gains are only moderate and generally too small to effectively counteract the family influence. When children are in school, the comparative analyses demonstrate that socioeconomically-advantaged families

Extended author information available on the last page of the article.

H.-P. Blossfeld (✉)
Lehrstuhl für Soziologie I, Fakultät Sozial- und Wirtschaftswissenschaften, Otto-Friedrich-Universität Bamberg
Postfach 1549, 96045 Bamberg, Germany
E-Mail: hans-peter.blossfeld@uni-bamberg.de

N. Kulic
Via dei Roccettini 9, European University Institute
50014 San Domenico di Fiesole, Italy
E-Mail: Nevena.Kulic@eui.eu

J. Skopek
Department of Sociology, Trinity College Dublin
College Green 3, Dublin, Ireland
E-Mail: skopekj@tcd.ie

M. Triventi
Dipartimento di Sociologia e ricerca sociale, Università di Trento
Via Verdi 26, 20142 Trento, Italy
E-Mail: moris.triventi@unitn.it

🖄 Springer

manage to secure the "pole positions" in education for their children, regardless of the organizational specificities of the school system across different countries. They always succeed in strategically exploiting various opportunities provided by different school systems. Finally, the cross-national comparisons of adult learning over the life course show a strong cumulative advantage: Adult learning tends to reproduce and reinforce the outcomes of initial formal education in the later adult life course.

Bedingungen und Konsequenzen ungleicher Bildungschancen im Lebenslauf: Ergebnisse aus dem international vergleichenden eduLIFE-Projekt

Zusammenfassung Dieses Papier vergleicht für eine größere Anzahl moderner Gesellschaften deskriptive Befunde zur Entstehung und zu den Konsequenzen ungleicher Bildungschancen im Lebenslauf. Die international vergleichende Analyse basiert auf einer Reihe hochstandardisierter (und damit vergleichbarer) länderspezifischer Fallstudien, in denen jeweils spezifische Bildungsphasen (Kleinkindalter, Schulalter, Erwachsenenalter) mit Längsschnittdaten untersucht wurden. Ziel des Beitrags ist es, gemeinsame Mechanismen des Bildungsverlaufs in modernen Dienstleistungsgesellschaften herauszuarbeiten. Die Ergebnisse des internationalen Vergleichs zeigen zunächst, dass die Bildungsungleichheiten (insbesondere die Kompetenzunterschiede) bereits im frühen Kleinkindalter in den Familien entstehen, also lange bevor die Kinder überhaupt beginnen, zur Schule zu gehen. Die Grundlagen der Bildungsungleichheiten werden in modernen Gesellschaften damit weiterhin im Kleinkindalter von einer Generation auf die nächste übertragen. Kinder aus sozial privilegierten Familien gehen in den meisten modernen Gesellschaften danach auch häufiger in Kinderbetreuungseinrichtungen mit höherer Qualität. Generell sind die Kompetenzzuwächse für benachteiligte Kinder in den bestehenden Kinderbetreuungseinrichtungen aber zu moderat und zu klein, um die Kompetenzdifferenzen zwischen Kindern aus verschiedenen Herkunftsfamilien effektiv ausgleichen können. In der Schule zeigen die komparativen Analysen, dass sozio-ökonomisch privilegierte Familien dafür sorgen, dass ihre Kinder immer wieder die aussichtsreicheren „Pole-Positionen" erhalten. Mit anderen Worten: unabhängig davon, wie das Bildungssystem in einem Land organisiert ist (als Gesamtschule, in Form interner Schul- oder Fächerdifferenzierung oder als dreigliedriges Schulsystem), gelingt es privilegierten Familien in allen Schulsystemen immer wieder, die jeweiligen Chancen und Vorteile, die diese unterschiedlichen Systeme bieten, für ihre Kinder zu nutzen. Damit sind institutionellen Reformen des Schulsystems zur Erreichung von mehr Bildungsgleichheit enge Grenzen gesetzt. Schließlich zeigt der internationale Vergleich, dass das formale und non-formale Lernen von Erwachsenen im späteren Lebenslauf in modernen Gesellschaften einen starken kumulativen Charakter auf-

weist. Das Bildungsverhalten von Erwachsenen tendiert damit generell dazu, die Bildungsunterschiede der Erstausbildung im Lebenslauf noch weiter zu verstärken, anstatt auszugleichen.

Schlüsselwörter Bildungsungleichheit · Matthäus-Effekt · Institutionelle Konfigurationen von moderneren Bildungssystemen · International vergleichende Lebensverlaufsanalyse · Standardisierte Fallstudien · Frühe Kindheit · Bildungsstratifikation in der Schule · Gesamtschule · Erwachsenenbildung

1 Introduction

The study of life courses has become one of the most active research fields in the social sciences in the last four decades. Retrospective life course and prospective panel studies have become available during this period in most modern societies, especially in North America and in Western Europe. Most of these datasets are representative longitudinal studies that explicitly recognize the dynamic nature of social roles and circumstances as men and women move through their life paths, the interdependence of lives and life choices, the situational imperatives confronting actors in various countries, and the cumulation of advantages and disadvantages experienced by the individual within national settings (Elder et al. 2004).

Today, most of the life course analyses have been studies of, and in, single societies. Based on such limited work, some life course researchers have interpreted their findings by contrasting what they have learned about the country that they actually studied and what is known or believed to be true about some other country or countries. Melvin Kohn (1987) classified such interpretations and comparisons as *implicitly* cross-national. The increasing availability of life history and panel studies for many countries provides a promising opportunity for more *explicit* cross-national life course analysis. They allow (1) to *establish the generality of findings* about the life course found in one particular society, and (2) to s*tudy the specific impact of variations in institutional configurations and social structures*, historically developed and country-specific, on specific phases of the life course or the life course as a whole (see also Kroneberg 2019).

Cross-national comparative life course studies can therefore greatly extend the scope of sociological knowledge when they answer the question of *whether a specific life course mechanism* established in one country also *applies outside of this particular national context*. Research based on longitudinal data from diverse countries therefore provides a particularly promising way to generate, test, and further develop sociological theory. Longitudinal data offer a much better handle to execute "*internal analysis*"—the analysis of variations within each country in a cross-national comparative study (Janoski and Hicks 1994; see also Schmidt-Catran et al. 2019). Life course studies also tend to *deepen our understanding of cross-national differences* when we are able to give a convincing explanation of the impact of institutional and social structural conditions on the life courses in various nations. In other words, cross-national life course research helps us to escape cultural one-

sidedness or ethnocentrism, because we, as social researchers, often wear cultural blinders of some sort that are connected to the society in which we are socialized.

This chapter presents selected results from the *Education as a Lifelong Process* (*eduLIFE*) project, supported by an Advanced Grant from the European Research Council (2012–2016). The aim of this cross-national project was to study how individuals' educational careers unfold over the life course in different societies. The project concentrated on conditions and on (short- and long-term) consequences of unequal educational opportunities over longer spans of the life course. In this chapter, we limit ourselves to three educational life phases: (1) the age of "early childhood education and care" (ECEC) before starting school, (2) the phase when pupils in secondary school are confronted with different models of school differentiation, and (3) lifelong learning in adulthood. Our leading research question is: How does social inequality influence educational careers and their outcomes in these three life phases?[1]

There is a widespread consensus today that panel and life history data improve the opportunities to describe trajectories of growth and development over the life course and to study the patterns of causal relationships over longer time spans in different societies. The choice of the countries for our cross-national comparisons was determined by the availability of representative longitudinal data (secondary analysis), and by asking whether including a particular country sheds additional light on the theoretical issue being studied. We have included 12–17 countries in our cross-national comparisons in each of the three educational phases. Our design is based on identical or highly similar meanings of survey questions in each country. The measurements are functionally equivalent, which means that they may assume different institutional forms in various countries (e. g. different institutions of pre-school education, various forms of school differentiation, and distinctive organizational models of adult learning), but refer to the same conceptual framework.

The country case studies were carried out by national experts who are familiar with the data sets available within each country and are able to analyze them to the fullest advantage. The joint comparative perspective and method were developed in several international workshops. These workshops included specifying theory and hypotheses, the comparability of concepts across countries, the question of how countries can be compared over longer historical periods, the application of statistical controls, and the equivalent measurement of dependent and independent variables.

On the pages below, we first develop the guiding theoretical perspective of our life course research, and then summarize the empirical key findings of the three educational life phases studied in the eduLIFE project.

---

[1] Limited space has prevented us from considering the particular phase of school-to-work transitions and their country variations. We refer the reader to the comparative volumes of Shavit and Müller (1998) and Shavit et al. (2007) with regard to the links between vocational training and higher education institutions in different national contexts.

2 Education as a lifelong process

The particular theoretical orientation of our cross-national comparisons is the life course. This perspective aligns attention toward the *process dimension* of education, and *links the (changing) social structure to the unfolding of individual lives*. The life course provides a framework for studying education at the nexus of developmental trajectories and social pathways, as well as institutional and social change. Elder et al. (2004) have summarized the following five general principles of life course research.

2.1 The principle of lifespan development

The first principle requires a focus on long-term individual trajectories over the lifespan. The resulting emphasis in sociology is therefore placed on systematic pathways of development and educational career profiles over time. Major aspects of educational careers are nationally-varying contexts that foster or hinder learning, competence development, and educational progress. Elementary and secondary educational institutions are often strictly age graded in modern industrialized societies. Thus, education as a lifelong process is to a large extent age structured, because age and time often exert a formal influence on progression through educational institutions during early childhood, school age, youth, and early adulthood. In other words, the movements of individuals through the educational systems was a central object of the *eduLIFE* project, both as a phenomenon to be explained and as a determinant of subsequent economic and noneconomic outcomes throughout the life course.

Educational institutions have not only the task of social integration, but they often serve as *gate keepers* in the lifelong process of reproducing social inequality. Thus, educational systems in modern societies intentionally *sort students into differing positions*, whether *within* schools, *between* schools, or both. The *eduLIFE* project therefore focused not only on between-school tracking, but also on ability grouping, age grouping, and interest grouping as the most common within-school stratification mechanisms in modern societies. These mechanisms structure educational career lines by opening up some doors whilst closing others. The research carried out by the *eduLIFE* project that is reported here traced the trajectories of individuals from early childhood, to lower and upper secondary school, and across adult learning.

Life-course research shows that the events and states of earlier educational stages often have lasting consequences for subsequent educational processes and outcomes. Dannefer (1987) introduced the "*Matthew effect*" into the literature on the life course. The Matthew effect means that small initial educational inequalities become magnified over the lifespan. Thus, there seems to be a kind of logic in educational careers in the sense that those who have already received an education receive even more education, and those who have received a poor education become relatively poorer over the life course. The Matthew effect is sometimes also referred to as the *cumulative disadvantage/advantage hypothesis* (O'Rand and Henretta 1999). It offers a cumulative explanation of how intracohort inequality is engendered from early education, via attendance at school, to adult learning. The Matthew effect seems to be particularly interesting today with regard to (1) the *long-term consequences of*

different institutions of early educational investments at pre-school age, and (2) the *opportunities offered by various adult learning systems to compensate for disadvantages engendered in the school system*. Cross-national educational research, which is often based on cross-sectional data, has paid relatively little attention in the past to the challenges of describing and explaining long-term educational trajectories. The research carried out by the *eduLIFE* project contributes to the longitudinal analysis of long-term processes.

2.2 The principle of linked lives

The second principle of life-course research concerns the interdependence of lives over time, especially in the family, where *individuals are linked across generations* by bonds of kinship and processes of intergenerational transmission. The *eduLIFE* project examined long-term relationships between parents and their children, and studied how these relationships influence the educational careers of children, adolescents, and adults over the life course. The kin-based perspective on the life course focuses on families, and helps to understand how different societies reproduce inequalities across generations. The life-course perspective of "linked lives" also refers to important *relationships outside the family*. Kindergartens and schools are the first educational organizations that children experience, and they constitute important social networks for most children. It is these educational settings in which knowledge and competencies are constantly *tested, evaluated, and compared with other students*, and in which children develop a sense of their intellectual efficacy.

2.3 The principle of agency

The third principle guiding the *eduLIFE* analyses concerns agency in human development and the idea that planfulness and intention can affect life-course processes and outcomes. In sociology, the idea of agency has been developed in the theories of *methodological individualism* and *rational action theory*—that is, theories that the macrolevel aggregates of educational inequality have to be reconstructed via the educational and occupational choices that families and individuals make under certain constraints in the life course. Drawing on rational action theory, the *eduLIFE* project employs *models of educational decision-making at critical branching points* over the educational career. These micro-macro models provide important conceptual tools for understanding how individuals from different social origins might incorporate the risk of educational failure along with beliefs about what kind of choices are possible into a rational calculation of costs and benefits.

2.4 The principle of timing of events and transitions

The fourth principle, that of the life-course perspective, emphasizes that the consequences of life transitions and events vary according to their timing in a person's career. It recognizes that the impact of life events is contingent on when they occur in an individual's life. For example, there are *"vulnerable" phases in an educational career* in most societies such as (a) the timing of entry into the school system, (b) the

period of transition to secondary school, (c) the period of transition from secondary school to vocational training, university education, or the employment system, and (d) the appropriate timing for starting adult education.

2.5 The principle of time and place

The fifth principle, namely that of time and place, states that *individuals' educational careers are embedded and shaped by the highly-specific historical times and country-specific institutions*. During the last decades, cross-national life-course research has demonstrated the necessity of nesting individual lives in social and historical contexts. The life-course researchers of the *eduLIFE* project have therefore considered a set of mechanisms related to period and cohort effects in terms of institutional and social change. The *multilevel design* of the *eduLIFE* project also allowed researchers to specify the *complexities of time and environments* in different societies more accurately for educational processes.

3 Social inequality and early childhood education and care (ECEC)

Early childhood is a decisive developmental phase that sets the stage for a broad range of later life course outcomes: Children's early educational experiences create developmental foundations often translating into long-term path dependencies in educational and occupational careers (DiPrete and Eirich 2006). Research in the United States in particular has demonstrated the efficacy of preschool investment in improving socio-economic outcomes for children facing adverse environmental conditions in very disadvantaged parental homes (Heckman 2006). The most influential empirical study from the United States is the often cited Perry Preschool experiment in Ypsilanti, USA, which started in the 1960s (see Schweinhart 2013). In this study, 120 Afro-American children with relatively low IQs (around 80) from very disadvantaged families (headed mostly by lone, uneducated, and often unemployed mothers) where randomly assigned to two groups at the age of 3–4: (i) a treatment group of about 60 children were sent to a high-quality preschool, and their families received additional support from professionals at home, and (ii) a control group (of about 60 children), where children and their families did not receive any additional support. The individuals of both groups were then interviewed and tested several times over their life course (Becker and Zangger 2015). The interesting finding was that the treatment group behaved differently from the control group even up to age 50. The members of the treatment group were more likely to be employed and achieve higher earnings, and were less dependent on social welfare (Heckman 2006). The Perry Preschool experiment established a remarkable long-term effect for a very specific study population (children from extremely poor families in the USA). Of course, it is desirable to understand and generalize the results of such a study as broadly as possible (see Schubert and Becker 2010). It is however doubtful whether the observed effect can be simply generalized to other social origin groups and countries.

Thus, a first aim of the cross-national *eduLIFE* study (Blossfeld et al. 2017) was to *focus on different social origin groups* and to analyze the *relevance of institutional contexts in modern societies*. In order to understand how social inequality in early educational opportunities is produced, it was at first crucial for the *eduLI*FE project to consider the *role of national institutions in early childhood*. Indeed, not only do the *availability and quality of childcare provision* vary across countries, but there are also major cross-national differences in the *variety of childcare options and services* (Gambaro et al. 2014). For instance, while early childhood education varies widely in quality in the United States (e. g. Vandell and Corasaniti 1990) and is strongly market based (Kamerman and Waldfogel 2005), early childhood programs in Europe are usually much more standardized by state regulations, more homogeneous in service, and more universally provided (Spiess et al. 2003). These differences also explain the quest for early educational interventions and programs in the United States that are targeted towards children from highly disadvantaged backgrounds. Nonetheless, organizational features of childcare systems differ vastly within Europe as well because of the plethora of country-specific social and educational policies. Hence, by taking a broader *cross-national perspective on early childhood education and care (ECEC)*, the *eduLIFE* project tried to enlarge the scope of the somewhat Anglophone-centric empirical literature, and includes other regions such as Northern, Southern, and Central European countries, and even Russia.

The aim of the *eduLIFE* project has been to understand *how (educational) inequalities emerge in early childhood*, and *what can be done to combat them*. *Parental care* is the first option available to families, and *is typically predominant* as exclusive care in the early months of children's lives. The first research question therefore focused on *the role of parental involvement and care* when it comes to causing social disparities in cognitive and non-cognitive outcomes in infancy and earliest childhood. Parental involvement refers to physical care, the stimulation of intellectual capacity and social behavior or, more broadly, time and material investment in children. Cognitive outcomes are considered in developmental and ability tests, while non-cognitive skills cover child attention and activity, as well as different soft skills.

A second research question specifically addressed the *relation between care arrangements* in infancy and early childhood up to preschool age *and children's social background*. In addition to parental care, the *eduLIFE* project considered two further main kinds of childcare: *informal childcare and formal childcare*. Informal care includes a variety of actors taking care of the child, such as grandparents, other relatives, friends, neighbors, and baby-sitters. The third form is formal childcare, which refers to institution-based forms of care, such as public or private nurseries. We were particularly interested in understanding whether and to what degree the decisions about *the various forms and timings of childcare arrangements in different countries* are influenced by a mother's education, household wealth and income, as well as parental social class.

A third research question concerned *different types of childcare arrangements* and how these mitigate or exacerbate social inequalities in early and later educational achievement. It is worth noting that, when studying this issue, the *eduLIFE* project considered not only exposure to formal childcare in the early years of life pure and simple, but also the important characteristics of the type of childcare attended

Table 1 Overview of (country-specific) studies on early childhood education and care (ECEC). Authors' own work

Authors	Data	Country
Weinert, Attig, and Roßbach	German National Educational Panel Study (NEPS), Starting Cohort 1	Germany
Barnett and Frede	Abbott Pre-K program evaluation	United States
Brilli, Kulic, and Triventi	ISTAT—Italian Survey on Births (2002, 2005 and 2012)	Italy
Dämmrich and Esping-Andersen	PIRLS/PISA	Cross-national analysis
Karhula, Erola, and Kilpi-Jakonen	Register data from Statistics Finland	Finland
Kosyakova and Yastrebov	Russian Longitudinal Monitoring Survey (RLMS)	Russia
Leseman, Mulder, Verhagen, Broekhuizen, van Schaik, and Slot	The national pre-COOL2–5 cohort study	The Netherlands
McGinnity, McMullin, Murray, and Russell	The Growing Up in Ireland longitudinal study (GUI)	Ireland
Del Boca, Piazzalunga, and Pronzato	Millennium Cohort Study	United Kingdom
Skopek	German National Educational Panel Study (NEPS), Starting Cohort 2	Eastern and Western Germany
Viklund and Duvander	Administrative register data from the Swedish Social Insurance Agency	Sweden
Wahler, Buchholz, and Breinholt	The Danish Longitudinal Survey of Children (DALSC)	Denmark
Zachrisson, Dearing, Blömeke, and Moser	Behavior Outlook Norwegian Developmental Study (BONDS)	Norway

such as its *quality, duration, and frequency*. An overview of the individual country-specific contributions within each topic is found in Table 1.

The cross-national research strategy therefore involved 12 in-depth country-specific studies in Europe and the US using longitudinal datasets and a standardized comparative study of 14 OECD countries. These studies were conducted by expert scholars in the field of early childhood who are familiar with the respective country contexts under study. Studies are grouped according to the three topics of the project: Some focus on the role of home environments and parental involvement in social inequalities in childhood; some deal with the stratification patterns in various childcare systems; while the majority investigates the link between social background, institutions of early childcare and education, and short- and long-term educational outcomes. The studies were not standardized in terms of methodology, yet each handled the questions with the most suitable methods. Applied methods range from traditional multivariate analyses (linear regression, binomial or multinomial logistic regression), which were employed in the majority of the studies, to experimental and simulation-based designs (e.g. US and UK study).

This research design is complex but provides one of the first cross-national investigations of the factors that drive *achievement gaps in cognitive and non-cognitive development in early childhood*, the goal being to understand the potential to combat these early inequalities through educational policies. Although the project embraces

insights from various disciplines, it largely emphasizes a sociological perspective. Attention is devoted to the role played by *country-specific early education institutions* in reducing educational inequality among children from different social backgrounds. Three research questions guide the presentation of our research results: (1) What is the role of early parental involvement and care in educational success? (2) How do families of different origins choose modes of childcare? (3) What are the consequences of early childcare and education for inequality of educational opportunity?

The first question was studied by looking at the *earliest mother/child interaction and parental involvement* in the course of early childhood. Using unique data from the National Educational Panel Study (NEPS), for example, the German study shows that a child's interactive behavior such as 'attention to objects' and 'activity level' is positively influenced by a higher social background and mediated by the mother's interactive behavior. At the early age of six to eight months, there is however only weak measurable evidence with regard to children's social disparities in developmental status, learning resources, or motor skills. Much early development seems to take place in the brain, and is hard to observe with conventional methods. When the children become older, for example the Irish study is able to demonstrate a direct relation between a better home-learning environment and more advanced vocabulary skills at different preschool ages. Based on our cross-national comparison, we can state that *educational inequalities are created and perpetuated in the family very early in a child's life*—long before school age. Resources, activities, and mother/child interaction in the family, shape children's early conditions and opportunities for learning at home differently by social background. Although initially rather minor and hard to measure, these early differences seem to be important harbingers of future social inequalities in educational achievement. Even if they are hard to detect shortly after birth, disparities among children of different social backgrounds tend to grow substantially through the early childhood years. This means that skill formation is a path-dependent and cumulative process (Phillips and Shonkoff 2000; Cunha and Heckman 2007): *learning begets learning*—tiny differences in infancy tend to grow, probably most strongly in toddlerhood; and if certain skills and competencies are not mastered at crucial ages, it will be hard for children to catch up when they are older (see also German National Academy of Sciences Leopoldina 2014).

Therefore, *parents' choices regarding early environments of education and care are critical*, given that these may enable or hinder opportunities for learning and acquiring relevant competencies. Almost all of the case studies in the *eduLIFE* project present evidence of the *social selectivity of childcare arrangements* where children from higher social backgrounds are more likely to attend institutional childcare and education and tend to take advantage of intensive participation in high-quality care. In contrast, informal care arrangements seem to be dominant for children from less privileged social backgrounds, especially in countries with rationing and low affordability of formal childcare services. The general findings show that *families with a lower social position rely more strongly on parental care*, while those with a higher social position more often resort to formal or informal childcare. This holds true for different measures of social background, be it the mother's education

or the father's occupation. Yet, the *eduLIFE* study specifically compared measures of social background, and concluded that a *mother's education has a primary role in explaining the early decisions made about childcare*.

In the Swedish system of universal access to childcare and very high participation rates, the mechanism of *differentiation in childcare take-up* is visible in the *age of entry into day care*—however, with only a small difference. Further cross-national analyses were also able to show that the decisions taken by *highly-educated parents* are particularly susceptible to policy changes and depend on the perceived quality of the childcare services: They *enroll their children earlier when the quality of service is high, and opt to wait when the quality is low*.

Several country-specific case studies were able to assess *shorter and longer-term consequences of children's experiences of early formal care*. Heckman (2006) argued that early education and care have the power to enhance children's cognitive and non-cognitive abilities and compensate for the initial disadvantage faced by less privileged children. When it comes to the first part of this hypothesis, confirmation was found in several studies of the *eduLIFE* project. For example, the German study on preschool children born in 2005/06 provided evidence that *earlier enrolment in center- or group-based care* (day care centers, crèches, or play groups) *is related to higher linguistic competencies* at age five for all children. Similarly, the British study showed that children who attended formal childcare (center- and group-based, non-parental care settings) when they were one and a half years old perform significantly better on average when it comes to school readiness at the age of three.

The Finnish study found that having attended day care below the age of three, as compared to being cared for at home, is associated with positive long-term outcomes such as attendance at upper secondary qualification, as well as entry into higher education. The Abbott Pre-K program in the United States that was designed to support disadvantaged children in 31 low-wealth school districts identified a positive impact of the preschool program on various achievements (such as math, language, and literacy), and this impact persisted through the first six grades of school. However, *the long-term positive effects of preschool are not found in the Danish case study*. Thus, with the exception of the Danish study, the eduLIFE project was able to *demonstrate that center-based care has a higher positive impact on the achievements of children from a lower social background, although the effects are limited in scope*.

In addition, several country-specific studies tested for the *role of preschool center quality or the quality of particular programs* in *compensating for the early disadvantages of children from a lower social background*. We learned from their results about the important role of centers with a structured curriculum in compensating for early disadvantage in achievements. The previously-mentioned case study on the United States gave additional insights into how to design and maintain high-quality early education programs to combat inequality among children. In addition, it could be shown that the government programs in the Netherlands target children of a low social background as early as the age of two by placing them into childcare centers of the highest quality. This study found a *strong correlation between the quality of ECEC and the growth of cognitive and non-cognitive skills* among disadvantaged children, even though the catch-up effects are found to be relatively minor. The
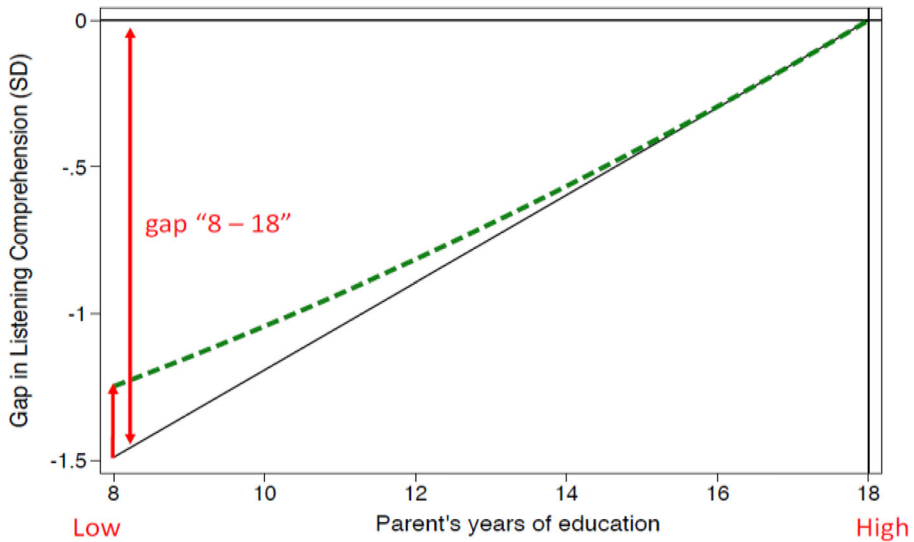
Fig. 1 Achievement gaps in listening comprehension by parental education at age 5 with and without attendance at institutions of early childhood education and care (ECEC) (*Dotted line:* with attendance at ECEC, *solid line:* without attendance at ECEC). (Source: Skopek 2016)

majority of Nordic countries with universal childcare systems in this project report relatively minor socio-economic disparities in access to early education.

In summary, based on our cross-national analysis, we have learned that the *highest returns on early childhood education and care programs*, particularly below the age of three, *could be harvested by children coming from the most disadvantaged backgrounds*. However, one has to bear in mind that (1) these children are also the ones who are least likely to attend, and (2) that their gains are only moderate and generally too small to achieve effective compensation. Several of the country case studies demonstrate that *the gains that children from disadvantaged families can achieve through participation in early childcare institutions are rather minor in relation to the extant achievement gaps for children coming from privileged families*. Using NEPS data, Skopek (2016) for example demonstrates this in Fig. 1. It shows that, at age five in Germany, the achievement gaps in listening comprehension of children from parents with only eight years of education compared to children whose parents received 18 years of education are only reduced to a small degree through attendance at early childcare and education institutions.

Thus, the most important theoretical conclusions of the cross-national comparative *eduLIFE* study are (1) that *all children can profit from early childcare (general elevator effect)*, (2) that *disadvantaged children certainly profit more than children from advantaged parental homes* (interaction effect), and (3) that *the social inequalities in achievement between children from unequal social backgrounds can only be marginally reduced by their attending early education and childcare institutions*. In other words, the effect of the Perry Preschool experiment that was demonstrated for disadvantaged children cannot be simply generalized to other social groups and other countries, and that *the interpretation of the Perry Preschool Program seems to be far*

too optimistic with regard to the possible reduction of social inequalities. From the cross-national evidence accumulated in the *eduLIFE* project, one can say that *greater exposure to early education and care programs of children from disadvantaged backgrounds* in Europe and beyond *would probably only partially narrow down the early achievement gaps between children from unequal social groups*. However, it is worth underlining that one should not only offer access to formal childcare at an earlier age, but also make sure that the *service is of a high quality*. Otherwise, attending formal childcare for children from low social backgrounds might not lead to any compensation of their early disadvantage in cognitive abilities, or might even widen these achievement gaps. Finally, the cross-national analyses of the *eduLIFE* project also provided evidence that the gains brought about by early *educational investments are only sustainable if they are supported during school age*. Several case studies in the *eduLIFE* project clearly demonstrated that *early educational intervention effects might fade away* over the years unless they are supported in school with a set of policy efforts that secure the early gains.

4 Social inequality and different models of educational differentiation in secondary education

Our country-specific studies on early childhood education and care (ECEC) made it clear that students' academic performance when starting school is closely related to their family background. Among the various goals of educational systems, two are prominent in contemporary societies: (1) to provide all students with a common foundation of competencies for full participation in civic and socioeconomic life (*social integration*) and (2) to *sort and select students* according to their abilities and diverse life-course goals (Van de Werfhorst and Mijs 2010). Consequently, sooner or later in the school career, students will enroll in different tracks, school types, ability groups, curricula, or subject courses (Dupriez et al. 2008). *Tracking between schools versus comprehensive schooling* stand out in the literature as the two ideal-typical approaches of sorting in secondary school. However, *these two basic school types have been converging in organizational terms* in many countries in recent years, so that the long-standing differences between them have become increasingly blurred.

On the one side, countries with a traditionally rigid system of early tracking have been introducing reforms aiming to make their education systems more flexible. Besides raising the compulsory school age, countries such as Germany or Switzerland have *increased the permeability of tracks, facilitated mobility between types of school, or have promoted inclusive school types* in addition to track schools (Benavot and Resnik 2006). In addition, *performance hurdles for the prestigious academic track have been reduced*. For example, several of the *Länder* in Germany have *abolished* the obligatory *achievement-based teacher's recommendation for upper secondary schools*. At the same time, *admission to higher education has been opened up more and more for students in non-traditional academic routes*, which is creating new opportunities, particularly for students starting off on vocational routes. All these reforms have been intended to make the early track allocation

less rigid and consequential in the tracking system, especially for children from less advantaged social origins. In addition, several reforms were introduced in the tracking systems in order to increase students' participation in higher education. For example, by creating specialized secondary schools (so-called "*Fachoberschulen*" and "*Berufsoberschulen*"), and introducing the less demanding professional colleges ("*Fachhochschulen*") next to the traditional Universities (Blossfeld et al. 2015).

On the other side, nations with c*omprehensive school systems* have been fanning out curricula programs by introducing new educational options (e. g. types of school, curricula, and subjects) which *lead to an unprecedented differentiation of the educational landscape in these schools*. Many of these transformations have been sponsored by a neoliberal stance on "school choice" which became increasingly dominant (Ascher et al. 1996). These neoliberal arguments not only underscore the centrality of parents' freedom when it comes to choosing the education that they would like their children to have, but also emphasize the autonomy of schools acting as agents in a quasi-market of educational supply and demand. While such market models of schooling based on principles of freedom of choice may contribute to the overall effectiveness and efficiency of a school system, they might harm equality of opportunity for children from lower social backgrounds (Ascher et al. 1996). Thus, *social inequality in access to the more prestigious and advantageous educational pathways might have become increasingly exacerbated in comprehensive systems*.

These transformations in secondary education systems call for a more fine-tuned approach towards analyzing social inequalities in education. Here we summarize some of the key findings from the cross-national *eduLIFE* project (Blossfeld et al. 2016b). It was aimed at overcoming the simple dichotomy between formally-tracked and untracked systems that are typically used in cross-national studies of educational inequalities, by also studying other more hidden ways of tracking and by adopting a longitudinal design to unravel the ways in which students have travelled through the education system in recent birth cohorts of students.

Two strands of cross-national research are particularly relevant for this work. First, the social stratification literature has examined the role of social background for individuals' educational transitions and educational attainment, and their changes across birth cohorts in various modern societies (e. g. Shavit and Blossfeld 1993; Shavit et al. 2007; Breen et al. 2009; Jackson 2013; Blossfeld et al. 2016a). These works provided important empirical evidence on cross-national differences in the strength and trends of inequalities of educational attainment, but *rely on relatively old cohorts of individuals* (usually born not later than the 1960–70s). Furthermore, they were not able to incorporate in the analyses detailed information on the specific type of secondary education attended, thereby failing to study a potentially important source of stratification of educational opportunities.

Second, there is also a *broad literature on educational inequalities using international school-based surveys and large-scale assessments* such as the "Programme for International Student Assessment" (PISA), or the "Trends in International Mathematics and Science Study" (TIMSS) (e. g. Duru-Bellat and Suchaut 2005; Marks 2005; Horn 2009; Becker and Schulze 2013). Without a doubt, several of the studies drawing on such data have significantly improved our understanding of educational differentiation in secondary school and its consequences for inequalities in student

achievement. However, most of this cross-national research has been carried out based on a very narrow definition of formal tracking (e.g. Hanushek and Wössmann 2006), while neglecting less visible forms of educational differentiation working in the background—such as placement in high-ability groups or specific course-taking patterns—which can be highly relevant in the social stratification of the student body—not only between but also within schools (Lucas 1999). Furthermore, these *studies have relied solely on snapshots of students' and schools' characteristics measured at a specific student age* (about age ten in TIMSS and age 15 in PISA). Thus, lacking a longitudinal design, these cross-national studies have been unable to assess the development of students' performance within different forms of secondary education, let alone the long-term consequences of educational differentiation for subsequent school transitions.

Last but not least, lacking information on students' prior educational experiences before allocation to different forms of secondary education, previous comparative research was very limited when it came to drawing conclusions on the consequences of differentiation for inequalities of educational opportunities. A *failure to incorporate prior achievement measures into statistical models of educational inequality makes it impossible to disentangle the "added value" of different types of differentiation* from mere selection effects arising from sorting students into tracks or ability groups according to ability (Morgan 2001).

The cross-national *eduLIFE* project aimed to overcome these kinds of drawbacks of the previous research. We conducted in-depth country-specific case studies in 17 countries characterized by various models of secondary education (see Table 2). The studies were unified by a common analytical scheme, addressed the same research questions within the context of a particular country's education system, and were conducted by reputable scholars in the field who are experts on the respective school systems and country contexts.

Figure 2 illustrates the overall longitudinal framework of our comparative study. Children's development of various skills in general, and students' academic performance in primary school in particular, prove to be strongly related to their family background in terms of cultural and socioeconomic resources. This creates differential opportunities for them to thrive. When students and their families face educational transitions—that is, when decisions have to be made on which types of education to pursue in lower and upper secondary school—both *social background and early school performance operate jointly in producing allocation outcomes*. These outcomes may, in turn, have consequences for students' subsequent achievement, given that the various forms of education provide different curricula-specific learning input and heterogeneous learning opportunities. Various path dependencies may lead such "school factors" to have profound effects on the subsequent educational trajectories, as well as on final educational attainment.

In our international workshops we identified *two major dimensions for classifying various aspects of differentiation in secondary education*. The first dimension distinguishes between *external* and *internal*. External differentiation refers to differences between schools, whereas internal differentiation refers to heterogeneity within schools such as differences across school classes or courses. The second dimension distinguishes between *formal* and *informal*. Formal differentiation refers

Table 2 Overview of the country-specific case studies included in the cross-national comparison of secondary school differentiation. Authors' own work

Authors	Data	Country
Early tracking model		
Buchholz, Skopek, Zielonka, Ditton, Wohlkinger, and Schier	NEPS Starting Cohorts 3/4/6, BiKS	Germany I
Lauterbach and Fend	LifE study	Germany II
Buchmann, Kriesi, Koomen, Imdorf, and Basler	COCON, TREE	Switzerland
Horn, Keller, and Róbert	NABC, HLCS	Hungary
Dronkers and Korthals	COOL, VOCL89	Netherlands
Nordic inclusive model		
Rudolphi and Erikson	CILS4EU, Register data	Sweden
Kilpi-Jakonen, Erola, and Karhula	Register data	Finland
Wahler, Buchholz, and Møllegaard	Register data	Denmark
Individual choice model		
McMullin and Kulic	LSYPE	England
Klein, Iannelli, and Smyth	Scottish and Irish School Leaver Surveys	Scotland and Ireland
Schührer, Carbonaro, and Grodsky	NELS	United States
Chesters and Haynes	LSAY	Australia
Mixed tracking model		
Farges, Tenret, Brinbaum, Guégnard, and Murdoch	Panel 1995 of French Ministry of Education	France
Contini and Triventi	INVALSI-SNV, IARD, ISTAT	Italy
Kosyakova, Yastrebov, Yanbarisova, and Kurakin	TrEC	Russia
Täht, Saar, and Kazjulja	ESS and FFS on Estonia	Estonia
Blank, Shavit, and Yaish	Register data	Israel

to regulated forms of diversity that are recognized by law and manifested in school certificates and qualifications. Informal differentiation refers to differences between types of education that are not recognized formally but can impact on the quality of instruction and levels of students' learning. Table 3 provides a classification of the main forms of differentiation based on their location along these two theoretical dimensions.

Although the education systems in the 17 countries studied in the cross-national project incorporate very different models of secondary education, we found that *allocation to different types of secondary education can be regarded as a general mechanism for the intergenerational reproduction of social inequalities* in contemporary societies. In all countries under study, *social background is associated positively with attendance at more prestigious types of secondary education* that provide students with higher-quality scholastic preparation, improve their performance, and increase their chances of entering more promising educational programs later on in their educational careers. Differential allocation by social background emerges largely from
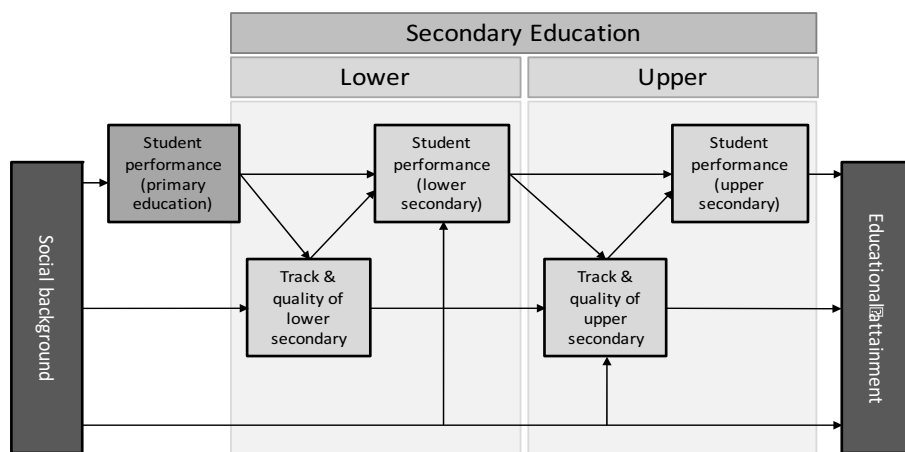
Fig. 2 The eduLIFE framework for analyzing individual trajectories through lower and upper secondary education. Source: Blossfeld et al. 2016b

differences in students' previous performance. Ability differences between social strata, though, do not fully explain these patterns.

Institutional forms of secondary education clearly vary across contemporary societies. Inequality of educational opportunity can also emerge from less obvious forms of differentiation such as school sector (e.g. public vs. private, religious vs. nonreligious), region, placement in ability groups, or choice of subjects within systems possessing flexible curricula (see also Blossfeld et al. 2016b). These *"hidden" forms of differentiation* can occur together with more *established and "visible" forms of tracking*, but they might also be manifest before children are formally allocated to different types of secondary education. In France, for instance, students with a higher social background are more likely to choose German as a first foreign language or Latin or Ancient Greek as an option. These choices, in turn, are related to greater chances of succeeding in an academic track. Also in Russia, the socially stratified pre-tracking allocation into top-tier (Lyceum and Gymnasium) and ordinary schools is likely to affect subsequent placement in upper secondary education and chances of success in higher education. Even in 'comprehensive systems' such as those found in Sweden, teaching based on *subject-specific ability grouping* occurs

Table 3 Classification of various forms of differentiation in secondary education. Source: Blossfeld et al. 2016b

	External (between schools)	Internal (within schools)
Formal	Formal school tracks School maintainer (public vs. private) School specialization (e. g. generalist vs. denominational school)	Specializations Subjects on advanced level
Informal	School reputation (e. g. ranking) School resources Student composition at the school level	Ability grouping Classroom composition Teachers' characteristics in different school classes

before students are streamed into upper secondary, tracks which contribute toward diverging the subsequent educational trajectories of students from different social backgrounds. Thus, our findings clearly demonstrate that in order to gain an appropriate understanding of the reproduction of educational inequalities in contemporary education systems, we need to take into account "hidden" forms of differentiation within secondary education (see Blossfeld et al. 2016b).

We also observed that individuals' secondary schooling careers are less fixed than the state of research suggests. The first allocation to different types of secondary education nowadays predicts subsequent educational pathways to a much lesser degree. Moreover, *mobility between types of secondary education is not as rare as previously thought in many countries*. Nonetheless, the flip side of the coin is that mobility patterns are strongly stratified by social background—beyond the importance of students' academic performance in directing their mobility between tracks. The few upward movements occur disproportionally frequently among students with highly-educated parents, whereas downward mobility is much more common among students with a low social background. As a corollary, one could argue that increasing flexibility in movements between tracks effectively leads to increasing rather than decreasing social inequalities in education.

Furthermore, the type of secondary education has lasting effects on students' subsequent educational careers, and on educational outcomes (such as competencies and skills) in later life. Notably, this result seems to be of a universal character, given that it holds true for all 17 countries studied in the *eduLIFE* project irrespective of the kind of educational differentiation in secondary school. This underscores the importance of integrating cross-national analyses of test score inequality with an extended perspective on further educational transitions and final educational attainment.

Comparative findings obtained from in-depth analyses in the *eduLIFE* project show us that *socioeconomically-advantaged families manage to secure the "pole positions" in education for their children regardless of the specificities of a school system*. They succeed in *strategically exploiting various opportunities provided by school systems in order to harvest the most favorable outcomes for their children*. As an unintended consequence, therefore, strategic behavior of socioeconomically-advantaged families appears as a strong social force offsetting the desired impacts of many educational reforms that aim to reduce social inequalities of educational opportunities. Thus, *the strategic behavior of families clearly limits the impact of educational reforms aiming to reduce inequalities of educational opportunities in school systems*.

5 Adult learning and social inequalities

Today, when individuals have entered the labor market, adult learning is increasingly shaping their social and economic opportunities. Especially in globalized and aging societies, it seems that the *generational replacement of older workers with obsolete skills by younger workers who have up-to-date qualifications has become a less efficient mechanism* to adapt the workforce to the rapidly-changing demands made by jobs and labor markets (Janossy 1966; Blossfeld and Stockmann 1999).

The focus is shifting towards keeping workers' skills continuously up-to-date during their working lives. *It is no longer appropriate that education takes place solely at the beginning of the life course*. In most modern societies, the primary policy focus with regard to adult learning has therefore been to increase participation rates. For example, the target of Europe's 2020 Agenda is to raise average participation rates to 15% for adults aged 25–64. Adult learning has also received considerable attention as a strategy to enable older workers to stay employed longer, thereby also reducing the pension burden of welfare states (D'Addio et al. 2010; Organisation for Economic Co-operation and Development [OECD] 2004). There are good reasons for being equally interested in the social inequalities of participation in adult learning and for focusing on the (educational) selectivity of participants.

The aim of the *eduLIFE* project was to examine adult learning by e*xploring cross-national patterns of participation in different adult learning activities* and *their consequences on individuals' labor market trajectories* using a life-course approach. We assessed the extent to which cross-national commonalities and differences exist in the mechanisms of social inequality in two different types of adult learning, namely formal and non-formal job-related adult learning. The research design included 13 country-specific case studies and two cross-national comparative studies (Blossfeld et al. 2014). The countries analyzed and the data used in these analyses are summarized in Table 4. The best available data were used in each country, and the analyses were guided by the same theoretical ideas, and the same statistical models were often used. As far as is possible, these studies used longitudinal data and statistical modeling that enables the analysis of theoretically important mechanisms over the life course. The data were modeled with multivariate statistical models, often using either event history analysis or methods of panel data analysis (such as random and fixed effects models). In the majority of analyses, the results that are included refer to models where the impact of other individual-level characteristics (such as age and labor force status) was taken into account. Moreover, the analyses were frequently run separately for women and men in order to take gender-specific life courses into account. Although previous research on these issues exists, these often only focus on either participation or on outcomes, and usually examine only one type of adult learning or include all types together. As far as we are aware, the *eduLIFE* project was the first one to combine the *study of both inequalities in participation and outcomes*, and to *examine whether the processes of social inequality are similar across the two different types of adult learning*.

Adult learning can be divided into formal, non-formal and informal learning. In our longitudinal analysis, we examined only the first two, and we further focused on learning related to the labor market due to the centrality of employment in modern societies. We view *formal adult education* as learning that leads to recognized certificates that can also be obtained along the typical educational career; it often takes place in formal educational institutions. In contrast, *non-formal adult learning* consists of (often) shorter training courses, and is frequently at least partly sponsored by employers. Nevertheless, non-formal adult learning may also be certified, but these certificates are not widely recognized qualifications in the same way as those obtained from formal education are. Finally, *informal adult learning* differs from these two by being less institutionalized; it is often self-directed. It should also be noted

Table 4 Overview of the (country-specific) case studies of adult learning. Authors' own work

Authors	Data	Country
Dämmrich, Vono de Vilhena, and Reichart	Adult Education Survey (AES)	Cross-national comparison or participation in adult learning
Triventi and Barone	International Adult Literacy Survey (IALS)	Cross-national comparison of returns to adult learning
Buchler, Chesters, Higginson and Haynes	Household Income and Labour Dynamics in Australia (HILDA)	Australia
Hamplová and Simonová	AES 2008, Labour Force Survey 2011 (LFS), Social Cohesion Survey 2005/2006 (SCS)	Czech Republic
Wahler, Buchholz, Myrup Jensen and Unfried	Integrated Database for Labor Market Research (IDA)	Denmark
Saar, Unt and Roosmaa	AES 2007, Family and Fertility Survey 2004/2005 (FFS)	Estonia
Kilpi-Jakonen, Sirniö and Martikainen	Register data from Statistics Finland	Finland
Buchholz, Unfried and Blossfeld	National Educational Panel Study (NEPS)	Germany
Csanádi, Csizmady and Róbert	AES 2007, Hungarian House-hold Panel Study (HHP)	Hungary
Barbieri, Cutuli, Lugo and Scherer	Indagine Longitudinale sulle Famiglie Italiane (ILFI)	Italy
Kosyakova	Russia Longitudinal Monitoring Survey (RLMS-HSE)	Russia
Vono de Vilhena and Miret Gamundi	Panel Survey on Inequalities in Catalonia (PAD)	Spain
Kilpi-Jakonen and Stenberg	Register data from Statistics Sweden (LISA)	Sweden
Elman and Weiss	National Longitudinal Study of Youth (NLSY79)	USA
McMullin and Kilpi-Jakonen	British Household Panel Study (BHPS)	UK

that countries differ in the way that adult learning is organized, which means that there is some variation in how each country in the *eduLIFE* project operationalized the two types of learning (functional equivalence of concepts).

Participation in adult learning often displays a pattern of cumulative advantage whereby those who are already better endowed also receive more (*Matthew effect;* see Merton 1968; Dannefer 1987). With regard to non-formal learning, this pattern has been explained by employers' incentives since it is employers who play a major role in sponsoring learning after labor market entry. It has been argued that the *higher educated are more trainable*, which means that each unit of training produces a greater enhancement in the productivity of highly-educated workers compared to those with lower educational attainment (Boeren et al. 2010; Dieckhoff 2007; Oosterbeek 1998). In addition, the *occupations in which the highly educated tend to work are likely to require more training* due to being knowledge intensive and requiring knowledge and skills to be kept constantly up to date, whereas low-skilled

jobs may remain more stable in their required tasks but have a greater risk of becoming obsolete in the long run due to technological innovations (OECD 2013).

On the other hand, individuals' incentives and barriers are likely to be more relevant for explaining educational selectivity for participation in formal education. In particular, *low prior educational attainment can be a barrier to entry despite the expansion of tertiary education* and new possibilities for individuals who do not satisfy traditional entry requirements. Moreover, psychological reasons such as prior schooling experiences may also act as an indirect barrier (Field 2000; Illeris 2003; Rubenson and Desjardins 2009). In addition, there are likely to be few incentives for individuals who are already highly educated to enter time-consuming formal education. Finally, there may also be ceiling effects so that at some point individuals are no longer able to climb the (formal) educational ladder. All in all, it is likely that the *benefits of participation in formal education are highest and the barriers impeding it are lowest for individuals with medium levels of education*.

It can generally be expected that adult learning is linked to positive labor market returns due to increased productivity as a consequence of the accumulation of human capital. In our research, we focused on what we term '*career progress*': Depending on the country in question, this can be defined as *upward mobility in terms of occupational prestige, social class or earnings*, or as (changes in) the level of these measures. However, *formal adult education* does not always take place as an upgrade from the previously held level of education, but *can also constitute an educational step sideways*, particularly when individuals want to change careers. In these cases, occupational status or earnings may not be any higher than they were before participation. On the other hand, participation in formal adult education (because it is time consuming) is also likely to act as a signal of higher motivation to employers, which should also increase employment outcomes. Furthermore, entry into many higher-status occupations tends to be restricted to individuals with the requisite qualifications, thus opening up access to individuals who acquire those (often tertiary-level) qualifications.

Non-formal learning is more likely to lead to productivity increases, particularly *when sponsored by the employer*, though these increases are likely to be smaller in size due to the shorter duration of training courses. On the other hand, the productivity-enhancing effects of training may be overstated if participants are already selected on the basis of higher productivity (or productivity potential). However, a review of studies that have been able to measure both wage and productivity growth concludes that individuals are able to capture only between one-fifth and a half of the financial returns to training, and the rest of the benefit goes to employers (Hansson 2008). Due to our focus on the benefits to the individual, our results are likely to reflect only a part of the overall benefits of training participation. Finally, it has also been suggested that not all types of non-formal learning lead to productivity increases, either because they are not designed to do so (e.g. because they are related to statutory requirements, such as health and safety courses, Field 2000), or because the compulsion to attend specific courses leads to low motivation and poor learning outcomes (e.g. some active labor market programs, Illeris 2003).

On the whole, *we expect the main mechanisms behind educational selectivity and the effect had by adult learning on career progress to be relatively similar across*

countries. However, it should also be recognized that the countries included in our analysis differ substantially in their *institutional configurations*, some of which are expected to affect different aspects of adult learning (e. g. Brunello 2001; Dieckhoff 2007; Wolbers 2005; see also Dämmrich et al. 2014; Triventi and Barone 2014). Our purpose here is not to assess the effect of specific institutions on particular aspects of adult learning, but rather to build a broader picture of adult learning across the different aspects and the two types that are analyzed.

If we summarize the results of the *eduLIFE* analysis of adult education in modern societies, the overwhelming conclusion is that, despite a wide variation in participation rates, *the main mechanisms of adult learning tend to be relatively similar across countries. Particularly when it is non-formal, adult learning displays a pattern of cumulative advantage and improves participants' career progress*. Nevertheless, some further observations can be made, in particular relating to the variation that is found in selectivity into formal adult learning.

One macro-level factor that seems to be related to less selectivity in formal adult learning is the education level of a country's adult population. A possible explanation for this may be that, in countries with high proportions of people with tertiary education, there is a greater perceived need among lower-educated adults to acquire additional education. This seems to be most keenly felt in Russia, where the social safety net of the state is also relatively sparse and therefore success in the labor market is paramount. In Estonia, higher participation rates of women compared to men and women's greater returns to new formal qualifications have also been explained by the expansion of higher education, accompanied by rising qualification requirements for certain (female-dominated) occupations, which in turn create pressures on women to gain new qualifications in order to be competitive, whereas men may rely more on their accumulated labor market experience. On the other hand, country differences in educational and occupational systems can also shape the form that these educational 'needs' take: In Germany, workers without occupational certificates have a strong incentive to obtain them due to their importance in the labor market.

Although a number of countries were not found to support the expectation that the medium-educated would be most likely to participate in formal adult learning, additional results from most of these countries suggest that the participation pattern is not one of cumulative advantage purely and simply. *It is often the case that disadvantages in the labor market also increase propensities towards participation*. This is the case in Australia, for example, where the conclusion is drawn that it is individuals in an intermediate position in the labor market who are most likely to re-enter formal education as adults. In Sweden, the long-term income trajectories of adult learners show that they have steadily fallen behind those among their peers who had similar levels of education. *Labor market disadvantages can also increase participation in non-formal learning*, although normally this is only the case for *non-formal learning that is not sponsored by employers* (but which is nevertheless related to the labor market). For example in Spain, workers in stable jobs are less likely to have participated in this type of learning than are those in precarious jobs or the unemployed—but this type of learning does not tend to improve one's labor market position.

With regard to how selectivity and career progress following adult learning combine to form broader patterns of social inequality, it is relatively clear that *no country is truly able at present to reduce social inequalities through adult learning*. Even in countries where some of the more disadvantaged individuals are able to obtain more adult learning, it is often those who obtain higher levels of (formal) adult learning who benefit more (such as in the Russian Federation and in Finland). The situation is worst in countries such as Hungary, where the disadvantages of low-status groups are exacerbated because they face considerable barriers when it comes to gaining access to any type of adult learning, and even when they do, they can only get into lower levels of adult education, which are less beneficial in the labor market. The same situation tends to hold for employer-sponsored non-formal learning, particularly in countries with strongly-segmented insider-outsider labor markets, such as in Italy, where marginal workers (not to mention individuals outside employment) are de facto excluded from beneficial training opportunities.

In this sense, *many countries seem to display a trade-off between equality and labor market rewards for adult learning*, which has been found for initial education (Bol and van de Werfhorst 2013). This is particularly the case for non-formal learning, but also to some extent for formal learning. *The comparative study also found that (short-term) wage returns of formal adult education correlate with wage returns on years of (initial) schooling*, suggesting that the institutional mechanisms driving the two are similar. On the other hand, there are suggestions in the results that positive combinations are possible: Lower selectivity and relatively widespread beneficial effects are seen in some countries, though only for formal learning, which tends to be much less widespread than non-formal learning. Nevertheless, positive cycles with greater investment in adult learning, the greater participation of less privileged individuals and gains in the labor market are possible, though not inevitable.

The most uniform pattern found in our analysis is one of cumulative advantage: Those members of society who are better off are better able to access adult education, and tend to see greater benefits ensuing from such learning. More generally, adult learning tends to reproduce and reinforce the outcomes of initial education. However, *there are substantial differences between the two types of adult learning that we have analyzed*: *Whereas the processes of cumulative advantage* and the trade-off between equality and labor market rewards *is clear for non-formal learning, this is less often the case for formal learning*. One of the reasons behind this difference may be that non-formal adult learning is more often sponsored by employers than is formal learning, particularly when the non-formal learning is job-related, which is what the *eduLIFE* project has analyzed. This means that *adult learning policies in modern societies need to explicitly target older, less-skilled workers as well as immigrants and the unemployed* because these groups tend to be overlooked in market-based systems (see also OECD 2013). *Age-based learning policies* are one step in this direction (Schuller and Watson 2009), but a broader conception of different life-cycle-based needs is also necessary (Billet 2010). Moreover, attention needs to be paid to the *content of the courses* in order to ensure that the participants also benefit from their participation in the labor market.

Taken as a whole, since much of the policy discourse related to adult learning emphasizes helping individuals keep their skills up-to-date and constantly develop, our message is that *countries need to shoulder greater responsibility when it comes to distributing opportunities for learning equitably* and *promoting the learning of individuals* who are not intrinsically motivated.

6 Summary and conclusions

Our cross-national research on the development of educational inequalities over the life course demonstrates that cross-national similarities greatly extend the scope of our knowledge about theoretical mechanisms in modern societies (Kohn 1987). Based on comparisons of diverse societies that vary widely in important institutional characteristics, our interpretations have gained considerable generality. By using an explicit cross-national comparison, it has been shown that there is empirical evidence of more universal sociological regularities.

Based on the cross-national analyses of the eduLIFE project, we have learned that *educational inequalities are created and perpetuated in the family early in a child's life*, long before children enter school. *Resources, activities, and mother-child interaction in the family, shape children's early conditions and opportunities for learning at home differently by social background*. Heckman (2006) claims that institutions of early childhood education and care are able to increase children's cognitive and non-cognitive abilities, and might compensate for the initial disadvantage faced by less privileged children. Our cross-national analysis confirms the first part of this hypothesis. *All children can profit from early childhood education and care* (elevator effect), and *children coming from the most disadvantaged backgrounds gain the greatest returns from these programs*. However, one has to bear in mind that these children are also the ones who are least likely to attend high-quality institutions, and if they do, their gains are only moderate and generally too small to effectively counteract the family influence.

When children are in school, the findings that we obtained from the comparative analyses demonstrate that *socioeconomically-advantaged families always manage to secure the "pole positions" in education for their children, regardless of the organizational specificities of the school system (this is the third time the authors have said this!)*. They always succeed in strategically exploiting various opportunities provided by different school systems, and thus obtain the most favorable outcomes for their children. *This strategic behavior of families clearly limits the success of educational reforms aiming to reduce inequalities of educational opportunities within school systems*.

Finally, we found a *uniform cumulative advantage in adult education*: The better off members of society are better able to access adult learning, and tend to see greater benefits from learning. More generally, adult learning tends to reproduce and reinforce the outcomes of initial education in the life course.

All *these regularities are* of course *far from being sociological laws*. They can only be generalized to the countries actually studied. Nevertheless, our theoretical explanations can focus on more general life course mechanisms common to them

(Kohn 1987, p. 719). Indeed, apparent similarities can always mask profound societal differences, but this danger is reduced significantly when the studies in one particular country are replicated by competent social scientists from other countries using comparable measurements and concepts, as well as systematic techniques of longitudinal analyses with extensive time-related statistical controls.

However, we should also mention that our examples of cross-national comparative life course studies also produced many *interesting cross-national differences*. When observed relationships differ from country to country, these inconsistencies have to be interpreted in terms of how the country-specific case studies or the countries differ. If we can rule out methodological differences between case studies as an explanation, we must take into account what is idiosyncratic about the particular countries for our interpretation (see Goerres et al. 2019). From an analytical point of view, it would be great if cross-national life course differences could be interpreted as instances of lawful regularities. This however requires a more explicit theoretical consideration of cultural and institutional conditions and further replications of the analyses in further countries.

References

Ascher, Carol, Norm Fruchter and Robert Berne. 1996. *Hard lessons: Public schools and privatization*. New York, NY: Twentieth Century School Fund.

Becker, Rolf, and Alexander Schulze. Eds. 2013. *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* [Educational contexts. Structural preconditions and causes of inequal educational opportunities]. Wiesbaden: VS Verlag für Sozialwissenschaften.

Becker, Rolf, and Christoph Zangger. 2015. Experimentelle Bildungsforschung. Beiträge experimenteller Forschung zur Bildungssoziologie und -ökonomie [Experimental research into education. Contributions of experimental research towards the sociology and economics of education]. In *Experimente in den Sozialwissenschaften, Soziale Welt*, Sonderband 22, ed. by Tobias Wolbring and Marc Keuschnigg, pp. 105–131. Bern: University Bern.

Benavot, Aaron, and Julia Resnik. 2006. Lessons from the past: A comparative socio-historical analysis of primary and secondary education. In *Global educational expansion historical legacies and political obstacles, edited by* Aaron Benavot, Julia Resnik, and Javier Corrales, pp. 1–89. Cambridge (MA): American Academy of Arts and Sciences.

Billet, Stephen. 2010. The perils of confusing lifelong learning with lifelong education. *International Journal of Lifelong Education* 29:401–413.

Blossfeld, Hans-Peter, and Reinhard Stockmann. 1999. The German dual system in comparative perspective. *International Journal of Sociology* 28:3–28.

Blossfeld, Hans-Peter, Sandra Buchholz, Jan Skopek and Moris Trivent. Eds. 2016b. *Secondary education models and social inequality. An international comparison*, Cheltenham, UK: Edward Elgar.

Blossfeld, Hans-Peter, Elina Kilpi-Jakonen, Daniela Vono de Vilhena and Sandra Buchholz. Eds. 2014. *Adult learning in modern societies: Patterns and consequences of participation from a life-course perspective*. Cheltenham, UK: Edward Elgar.

Blossfeld, Hans-Peter, Nevena Kulic, Jan Skopek and Moris Triventi. Eds. 2017. *Childcare, early education and social inequality. An international perspective*. Cheltenham, UK: Edward Elgar.

Blossfeld, Pia Nicoletta, Gwendolin Blossfeld and Hans-Peter Blossfeld. 2016a. Changes in educational inequality in cross-national perspective. In *Handbook of the Life Course,* edited by Michael J. Shanahan, Jeylan T. Mortimer and Monica Kirkpatrick Johnson, pp. 223–247. New York, NY: Springer.

Blossfeld, Pia Nicoletta, Gwendolin Josephine Blossfeld and Hans-Peter Blossfeld. 2015. Educational expansion and inequality in educational opportunity: Long term changes for East and West Germany. *European Sociological Review* 31:144–160.

Boeren, Ellen, Ides Nicaise and Herman Baert. 2010. Theoretical models of participation in adult education: The need for an integrated model. *International Journal of Lifelong Education* 29:45–61.

Bol, Thijs, and Herman G. van de Werfhorst. 2013. Educational systems and the trade-off between labor market allocation and equality of educational opportunity. *Comparative Education Review* 57:285–308.

Breen, Richard, Ruud Luijkx, Walter Müller and Reinhard Pollak. 2009. Nonpersistent inequality in educational attainment: Evidence from eight European countries. *American Journal of Sociology* 114:1475–1521.

Brunello, Giorgio. 2001. *On the complementarity between education and training in Europe*. IZA Discussion Paper No. 309. Bonn: Institute for the Study of Labor.

Cunha, Flavio, and James Heckman. 2007. *The technology of skill formation* (No. w12840). National Bureau of Economic Research.

D'Addio, Anna Cristina, Mark Keese and Edward Whitehouse. 2010. Population ageing and labour markets. *Oxford Review of Economic Policy* 26:613–635.

Dämmrich, Johanna, Daniela Vono de Vilhena and Elisabeth Reichart. 2014. Participation in adult learning in Europe: The impact of country-level and individual characteristics. In *Adult learning in modern societies: Patterns and consequences of participation from a life-course perspective*, ed. by Hans-Peter Blossfeld, Elina. Kilpi-Jakonen, Daniela Vono de Vilhena and Sandara Buchholz, pp. 29–55. Cheltenham, UK: Edward Elgar.

Dannefer, Dale. 1987. Aging as intercohort differentiation: Accentuation, the Matthew Effect, and the life course. *Sociological Forum* 2:211–236.

Dieckhoff, Martina. 2007. Does it work? The effect of continuing training on labour market outcomes: a comparative study of Germany, Denmark, and the United Kingdom. *European Sociological Review* 23:295–308.

DiPrete, Thomas A., and Gregory M. Eirich. 2006. Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments, *Annual Review of Sociology* 32:271–297.

Dupriez, Vincent, Xavier Dumay and Anne Vause. 2008. How do school systems manage pupils' heterogeneity?. *Comparative Education Review* 52:245–273.

Duru-Bellat, Marie, and Bruno Suchaut. 2005. Organisation and context, efficiency and equity of educational systems: what PISA tells us. *European Educational Research Journal* 4:181–194.

Elder, Glen Holl, Monica Kirkpatrick Johnson and Robert Crosnoe. 2004. The emergence and development of the life course. In *Handbook of the life course* edited by Jeylan T. Mortimer and Michael J. Shanahan. New York: Plenum.

Field, John. 2000. *Lifelong learning and the new educational order*. Stoke-on-Trent and Sterling: Trentham Books.

Gambaro, Ludovica, Kitty Stewart and Jane Waldfogel. Eds. 2014. *An equal start? Providing quality early education and care for disadvantaged children*. Bristol: Policy Press.

German National Academy of Sciences Leopoldina. 2014. *Socialisation in early childhood. Biological, psychological, linguistic, sociological and economic perspective*s, Statement, Deutsche Akademie der Naturforscher Leopoldina e. V. – Nationale Akademie der Wissenschaften. Halle (Saale): German National Academy of Sciences Leopoldina.

Goerres, Achim, Markus B. Siewert and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Hansson, Bo. 2008. *Job-related training and benefits for individuals: A review of evidence and explanations*. OECD Education Working Papers No. 19. Paris: OECD.

Hanushek, Eric. A., and Ludger Wössmann. 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal* 116:63–76.

Heckman, James. 2006. Skill formation and the economics of investing in disadvantaged children. *Science* 312:1900–1902.

Horn, Daniel. 2009. Age of selection counts: a cross-country analysis of educational institutions. *Educational Research and Evaluation* 15:343–366.

Illeris, Knud. 2003. Towards a contemporary and comprehensive theory of learning. *International Journal of Lifelong Education* 22:396–406.

Jackson, Michelle. 2013. Social background and educational transitions in England. In *Determined to Succeed?: Performance versus choice in educational attainment,* ed. by Michelle Jackson. pp. 253–278. Stanford: Stanford University Press.

Janoski, Thomas, and Alexander M. Hicks. Eds. 1994. *The comparative political economy of the welfare state*. New York, NY: Cambridge University Press.

Janossy, Franz. 1966. *Das Ende der Wirtschaftswunder* [The end of the Economic Miracle]. Frankfurt a. M.: Campus.

Kamerman, Sheila B., and Jane Waldfogel. 2005. Market and non-market institutions in early childhood education and care. In *Market and non-market institutions*, edited by Richard Nelson, pp. 185–212. New York: Russell Sage Foundation.

Kohn, Melvin L. 1987. Cross-national research as an analytic strategy. *American Sociological Review* 52:713–731.

Kroneberg, Clemens. 2019. Theory development in comparative social research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00604-y.

Lucas, Samuel R. 1999. *Tracking inequality. Stratification and mobility in American high schools*. New York: Teachers College Press.

Marks, Gary N. 2005. Cross-national differences and accounting for social class inequalities in education. *International Sociology* 20:483–505.

Merton, Robert K. 1968. The Matthew effect in science. *Science* 159:56–63.

Morgan, Stephen L. 2001. Counterfactuals, causal effect heterogeneity, and the Catholic school effect on learning. *Sociology of Education* 74:341–374.

OECD. 2004. *Completing the foundation for lifelong learning: An OECD survey for upper secondary schools*. Paris: OECD Publishing.

OECD. 2013. *OECD skills outlook 2013. First Results from the Survey of Adult Skills*. Paris: OECD Publishing.

Oosterbeek, Hessel. 1998. Unravelling supply and demand factors in work-related training. *Oxford Economic Papers* 50:266–283.

O'Rand, Angela M., and John C. Henretta. 1999. *Age and inequality: Diverse pathways through later life*. Boulder, CO: Westview.

Phillips, Jack P., and Deborah A. Shonkoff. Eds. 2000. *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.

Rubenson, Kjell, and Richard Desjardins. 2009. The impact of welfare state regimes on barriers to participation in adult education: A bounded agency model. *Adult Education Quarterly* 59:187–207.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schubert, Frank, and Rolf Becker. 2010. Social inequality of reading literacy. A longitudinal analysis with cross-sectional data of PIRLS 2001 and PISA 2000 utilizing the pairwise matching procedure. *Research in Social Stratification and Mobility* 28:109–133.

Schuller, Tom, and David Watson. 2009. *Learning through life: Inquiry into the future for lifelong learning (IFLL)*. Leicester: National Institute of Adult Continuing Education.

Schweinhart, Lawrence J. 2013. Long-term follow-up of a preschool experiment. J*ournal of Experimental Criminology* 9:389–409.

Shavit, Yossi, and Hans-Peter Blossfeld. Eds. 1993. *Persistent inequality: Changing educational attainment in thirteen countrie*s. Social Inequality Series. Boulder: Westview.

Shavit, Yossi, and Walter Müller. Eds. 1998. *From school to work. A comparative study of educational qualifications and occupational destinations*. Oxford: Oxford University Press.

Shavit, Yossi, Richard Arum and Adam Gamoran. Eds. 2007. *Stratification in higher education. A comparative study*. Stanford: Stanford University Press.

Skopek, Jan. 2016. Childcare as an 'equalizer'? *Early institutional care and educational outcomes of children—new evidence from the NEPS*, Unpublished Working Paper, Florence (Italy): European University Institute.

Spiess, Katharina C., Felix Büchel and Gert G. Wagner. 2003. Children's school placement in Germany: Does kindergarten attendance matter? *Early Childhood Research Quarterly* 18:255–270.

Triventi, Moris, and Carlo Barone. 2014. Returns to adult learning in comparative perspective. In *Adult learning in modern societies: Patterns and consequences of participation from a life-course perspective*, ed. by Hans-Peter Blossfeld, Elina Kilpi-Jakonen, Daniela Vono de Vilhena and Sandra Buchholz, pp. 56–75. Cheltenham, UK: Edward Elgar.

Vandell, Deborah L., and Mary Anne Corasaniti. 1990. Variations in early child care: Do they predict subsequent social, emotional, and cognitive differences? *Early Childhood Research Quarterly* 5:555–572.

Van de Werfhorst, Herman G., and Jonathan B. Mijs. 2010. Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology* 36:407–428.

Wolbers, Maarten H. J. 2005. Initial and further education: substitutes or complements? Differences in education and training over the life-course of European workers. *International Review of Education* 51:458–478.

Hans-Peter Blossfeld Prof., Dr. Dr. h. c., has held the Chair of Sociology I at the Otto Friedrich University Bamberg, Germany, since 2002 (he was on leave from the post from 2012–2017). Between 2012 and 2017, he was Professor of Sociology at the European University Institute (EUI) in Florence, Italy, and PI of the Education as a Lifelong Process (eduLIFE) project, supported by an Advanced Grant from the European Research Council (ERC). He has published 44 books and more than 320 articles on life course research, educational sociology, youth, family, labor market research, demography, social stratification and mobility, methods of quantitative social research, and statistical methods for longitudinal data analysis. These have been cited more than 28,000 times (H-Index = 71; I10-Index = 231; see Google Scholar, September 2018).

Nevena Kulic Dr., is a Research Fellow at the European University Institute (EUI) in Florence, Italy. She was a Senior Research Scientist in the eduLIFE project in the Comparative Lifecourse & Inequality Research Centre (CLIC) at the EUI. She obtained her PhD in Sociology and Social Research at the University of Trento, Italy, and has an MPhil in Economics from Glasgow University, United Kingdom. Her current research interests extend to the areas of family sociology, gender, inequality, and comparative longitudinal analysis, and cover a range of topics such as intra-household sharing, family forms and family formation, women in the labor market, and education.

Jan Skopek Dr., is an Assistant Professor at the Department of Sociology at Trinity College Dublin. He previously held positions as a post-doctoral researcher at the State Institute of Family Research at the University of Bamberg, Germany, and as a post-doctoral Senior Researcher in the eduLIFE project at the Comparative Life Course & Inequality Research Centre (CLIC), housed in the Department of Political and Social Sciences of the EUI, Florence, Italy. His research relates to family and social stratification, social demography, the life course, cross-national comparison, and quantitative methodology. He has co-edited several books, and publishes in internationally-renowned journals.

Moris Triventi Dr., is a Senior Assistant Professor in Quantitative Sociology at the University of Trento and a member of the CSIS (Center for Social Inequality Studies). He was a Senior Research Scientist of the eduLIFE project in the Comparative Lifecourse & Inequality Research Centre (CLIC) at the EUI, Florence, Italy, and a Lecturer in the Department of Political and Social Sciences (SPS) at the EUI. He received his PhD in Applied Sociology and Methodology of Social Research at the University of Milan-Bicocca in 2010, and received additional training on quantitative methods at the Universities of Essex, Oxford, Bamberg, Trento, and the EUI. His research interests extend to social stratification, educational inequalities, higher education, and quantitative methods of data analysis.

Elina Kilpi-Jakonen DPhil, is an Adjunct Professor and Academy Research Fellow of the Academy of Finland, based at the University of Turku, Finland. She was a Senior Research Scientist of the "eduLIFE" projects at the EUI, Florence, Italy, a Postdoctoral Research Fellow in the Oxford Network for Social Inequality Research, based at the Department of Sociology, University of Oxford, United Kingdom, and a non-stipendiary Junior Research Fellow at New College, Oxford. She also completed her DPhil and MSc in Sociology at Oxford, as well as her BA in Philosophy, Politics and Economics. Her research interests include educational and labour market inequalities, ethnicity, family, and international comparisons.

Daniela Vono de Vilhena Dr., is a Scientific Coordinator and Research Scientist at the Population Europe Secretariat, Max Planck Institute for Demographic Research in Berlin. She holds a Ph.D. in Demography, and was a Senior Research Scientist in the eduLIFE project at the Comparative Life Course & Inequality Research Centre, EUI, Florence, Italy, and a Research Fellow at the Centre for Demographic Studies, Autonomous University of Barcelona, Spain, where she completed her Ph.D. Her research topics include international comparisons of education systems, social inequality, life course research, labor market research and quantitative methods.

Sandra Buchholz Prof., Dr., has headed the research area "Educational Careers and Graduate Employment" of the German Centre for Higher Education Research and Science Studies (DZHW) in Hanover since 2018. She completed her Doctorate in 2008 and her Habilitation in 2016, both at Bamberg University. From 2012 to 2017, she held a temporary appointment as Professor of Sociology at the University of Bamberg. From 2017 to 2018, she was interim Director of the Leibniz Institute for Educational Trajectories and Professor for Longitudinal Educational Research in Bamberg. Her research focuses on international comparisons, education systems, social inequality, life course research, labor market research, and quantitative methods.

E. Kilpi-Jakonen
Unit of Sociology, Department of Social Research, University of Turku
20014 Turku, Finland
E-Mail: elina.kilpi-jakonen@utu.fi

D. Vono de Vilhena
Population Europe, WissenschaftsForum, Max Planck Institute for Demographic Research
Markgrafenstraße 37, 10117 Berlin, Germany
E-Mail: vono@demogr.mpg.de

S. Buchholz
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung
Lange Laube 12, 30159 Hannover, Germany
E-Mail: buchholz@dzhw.eu

ABHANDLUNGEN

How to Obtain Comparable Measures for Cross-National Comparisons

Jan Cieciuch · Eldad Davidov · Peter Schmidt · René Algesheimer

Abstract Comparisons of means or associations between theoretical constructs of interest in cross-national comparative research assume measurement invariance, that is, that the same constructs are measured in the same way across the various nations under study. While it is intuitive, this assumption needs to be statistically tested. An increasing number of sociological and social psychological studies have been published in the last decade in which the cross-national comparability of various scales such as human values, national identity, attitudes toward democracy, or religiosity, to name but a few, were tested. Many of these studies did not manage to fully achieve measurement invariance. In this study we review, in a nontechnical manner, the methodological literature on measurement invariance testing. We explain what it

J. Cieciuch
Institute of Psychology, Cardinal Wyszyński University in Warsaw
Wóycickiego 1/3 bud. 14, 01-938 Warsaw, Poland
E-Mail: jancieciuch@gmail.com

E. Davidov (✉)
Institut für Soziologie und Sozialpsychologie, Universität zu Köln
Albertus-Magnus-Platz, 50923 Cologne, Germany
E-Mail: e.davidov@uni-koeln.de

Department of Sociology and University Research Priority Program Social Networks,
University of Zurich
Andreasstrasse 15, 8050 Zurich, Switzerland

P. Schmidt
Center for International Development and Environmental Research (ZEU), University of Giessen
Senckenbergstr. 3, 35390 Giessen, Germany
E-Mail: peter.schmidt@sowi.uni-giessen.de

J. Cieciuch · R. Algesheimer
Department of Business Administration and University Research Priority Program Social Networks,
University of Zurich
Andreasstrasse 15, 8050 Zurich, Switzerland
E-Mail: rene.algesheimer@business.uzh.ch

⚫ Springer

is, how to test for it, and what to do when measurement invariance across countries is not given in the data. Several approaches have been recently proposed in the literature on how to deal with measurement noninvariance. We illustrate one of these approaches with a large dataset of seven rounds from the European Social Survey (2002–2015) by estimating the most trustworthy means of human values, even when strict measurement invariance is not given in the data. We conclude with a summary and some critical remarks.

Wie kann man invariante Messungen in international vergleichender Forschung erhalten?

Zusammenfassung Vergleiche von Mittelwerten und von Beziehungen zwischen theoretischen Konstrukten, die im Rahmen international vergleichender Forschung untersucht werden, gehen davon aus, dass diese Konstrukte messinvariant sind, d. h., dass sie in den verschiedenen Ländern identisch gemessen werden. Obwohl diese Annahme plausibel sein kann, muss sie jedoch statistisch getestet werden. Im letzten Jahrzehnt wurde eine zunehmende Zahl von soziologischen, politikwissenschaftlichen und sozialpsychologischen Studien veröffentlicht, in denen die internationale Vergleichbarkeit von verschiedenen Skalen zur Messung von z. B. menschlichen Werten, nationaler Identität, Einstellungen zu Demokratie oder Religiosität überprüft wurde. In vielen dieser Studien konnte Messinvarianz nicht völlig nachgewiesen werden. Die folgende Studie bietet in einer nicht technischen Art und Weise einen Überblick über die methodologische Literatur zur Messinvarianz. Es wird erklärt, was Messinvarianz ist, wie man sie überprüft und was man tun kann, wenn sie in den Daten nicht gegeben ist. In der Literatur wurden in der letzten Zeit verschiedene Ansätze vorgeschlagen, wie man fehlende Messinvarianz behandeln kann. Die Autoren illustrieren eine dieser Herangehensweisen (Alignment) mit einem großen Datensatz, der 7 Befragungsrunden des European Social Survey (2002–2015) beinhaltet, und schätzen den vertrauenswürdigsten Durchschnitt menschlicher Werte, auch wenn strikte Messinvarianz in den Daten nicht vorhanden ist. Abschließend folgen eine Zusammenfassung und einige kritische Anmerkungen.

1 Introduction

Comparisons in cross-national comparative research (CNCR) assume that the same constructs are measured in the same way across the various nations under study. It is a very basic and intuitive assumption. However, this assumption becomes quite problematic in applied research because fulfilling this assumption is often not easy. Indeed, the methodological literature suggests that even the strict application of the

same procedures of data collection and utilizing excellent translations of measurement instruments may not guarantee that the measurements are comparable and that cross-country comparisons based on these measurements would lead to meaningful results. Thus, while the paper of Goerres et al. (2019) in this issue discusses research designs and case selection for CNCR, Meuleman (2019) presents analytical techniques to analyze such data, and Schmidt-Catran et al. (2019) review techniques to deal with small and nonrandom country samples as well as unobserved heterogeneity in CNCR, our study focuses on how to make sure that our variables analyzed across the countries in CNCR are comparable. Three main questions arise: (1) What does it mean to measure the same construct in groups that differ in terms of language, history, culture, etc., or to measure the same construct over time?; (2) How can we be sure that we are measuring the same construct?; and (3) How can one obtain comparable scores across different groups under study? Answers to these questions are usually given in the framework of factor analysis and measurement models, where question items are used as reflective indicators for measuring latent variables that represent the theoretical constructs of interest that we want to compare. Before any cross-national comparisons are conducted, it is necessary to ensure that the same latent variables are measured in different countries, that respondents understand the items in a similar manner, and that they use the response scales in the same way. Meeting these three conditions allows us to speak of measurement invariance (Davidov et al. 2014).

We will present the topic of measurement invariance below, before going on to explain why it is important, what a test of measurement invariance requires and the logic behind it, how it can be tested across countries, and how and under which conditions its requirements may be released. We will demonstrate its procedure by examining the measurement invariance properties across countries and time points of human values measurements in the European Social Survey (ESS) between 2002 and 2015. We will then close with a summary and some concluding and critical remarks on the significance and indispensability of measurement invariance testing in cross-national research.

2 Measurement Invariance: What Is It, How Do We Test for It, and How Do We Deal with Noninvariance?

2.1 Measurement Invariance

Measurement invariance (or measurement equivalence) implies that using the same questionnaire in different groups (such as countries or at various points in time, or under different conditions) does measure the same construct in the same way (Chen 2008; Davidov et al. 2014; Horn and McArdle 1992; Millsap 2011). When measurement invariance is not established, comparisons between groups may not be meaningful because there is then no way to correctly determine whether observed differences across the groups are "true" or are only a methodological artifact (Chen 2008; Davidov et al. 2014). If a measurement is noninvariant, then findings on similarities or differences across groups may be misleading. It could well be the

case that differences that are found between groups do not correspond to real differences, or that observed similarities would not reflect real similarities. For example, social desirability response bias may be stronger in one country than in another, a question may be unclear to many respondents in a certain country due to the cultural specificity of a question, or a construct might have a different meaning across different nations. Factors such as these might lead to a different understanding of survey questions or to a different use of the response scale, thus rendering responses noncomparable across nations (Davidov et al. 2014).

2.2 How to Test for Measurement Invariance

Measurement invariance can be tested empirically. The most commonly used method to test it is multigroup confirmatory factor analysis (MGCFA: Jöreskog 1971; see also Cieciuch and Davidov 2012, 2015). The test requires using latent variables with multiple measures (indicators). MGCFA assesses whether (1) the same measurement model is used in all groups, (2) factor loadings are the same across groups, (3) measurement intercepts are the same across the groups to be compared, and (4) residual variances are fixed to be equal across groups. The first condition is known as configural invariance, and satisfying it still precludes comparisons. The second condition is referred to as metric invariance. It implies that the scale intervals are the same across groups because the loadings are the same in each group. Satisfying the second condition allows comparing unstandardized regression coefficients and/or covariances across groups. The third condition is known as scalar invariance. Meeting it allows also comparing the latent means across groups meaningfully. Scalar invariance is the most restrictive model, since it requires both factor loadings and intercepts to be the same in all groups. However, when it is fulfilled, it implies that the researcher may carry on any comparison across the groups in question with confidence. The fourth condition is dubbed full uniqueness measurement invariance. This basically means that the explained variance for every item is the same across groups, in other words that the latent construct is measured identically across groups. When the error variances are not equal and complete uniqueness is not given, it implies that the items are measured with different amounts of error in different groups, but that one can still compare unstandardized regression coefficients and latent means across groups. Therefore, typical multigroup factor analysis generally only applies the first three steps.

Technically speaking, the test of measurement invariance in this framework involves setting cross-group constraints on parameters (loadings or loadings and intercepts) and comparing hierarchically more constrained models with less constrained ones (Davidov et al. 2014; Vandenberg and Lance 2000). These models are hierarchical in the sense that, at the configural level, all loadings and intercepts are freely estimated, while at the metric level loadings are constrained to be equal across groups, and at the scalar level both loadings and intercepts are constrained to be equal across groups. If the fit of the more highly constrained model does not dete-
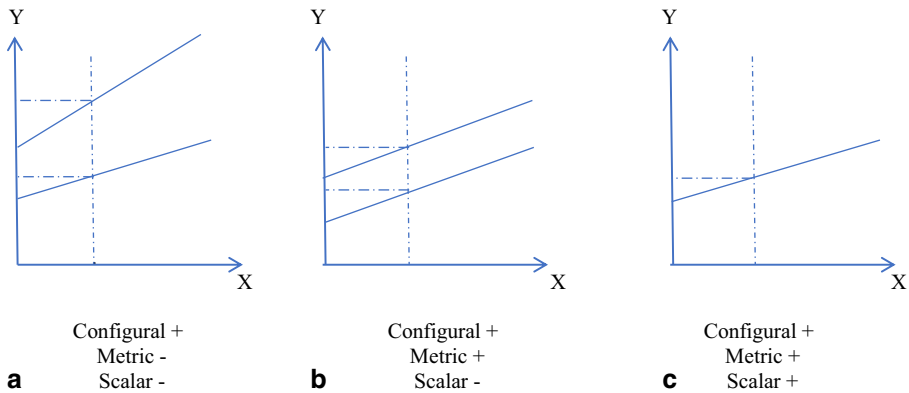
Fig. 1 Illustration of configural invariance, metric invariance, and scalar invariance across two countries. The X axis represents the latent variable mean; the Y axis represents the response to a survey question item measuring the latent variable. The diagonal represents the functional relation between the latent variable and the response to the survey question item in two countries (in unstandardized terms). **a** Configural invariance, **b** metric invariance, **c** scalar invariance. (Author's own work)

riorate considerably, then one can assume that it is supported by the data and that the corresponding measurement invariance level is established.[1]

To illustrate, imagine respondents answering a survey question in two countries, Austria (A) and Belgium (B). Figure 1 describes three scenarios of associations between the latent construct of interest, on the X axis, and the response to a survey question measuring this construct of interest, on the Y axis. Each scenario describes the relation between the latent variable (X) and the question item (Y). Each line describes the association between the latent variable and the item in one country. The first scenario (a) illustrates configural invariance (and metric and scalar *non*invariance). In this scenario, both the factor loadings and the intercepts are different across countries, as evidenced by the different slopes and intersections with the Y axis. The second scenario (b) describes metric invariance (and scalar *non*invariance). In this scenario, the slopes are identical, and thus reflect the fact that factor loadings are the same in the two countries. However, the intercepts are not. The third scenario (c) presents metric and scalar invariance. In this case, both the loadings and the intercepts are equal across countries (which is why one can observe only a single line). The figure makes it clear that, in the two scenarios (a) and (b), observed scores are different, even when the true score (on the X axis) is

---

[1] Researchers examine global fit measures and perform chi-square difference tests to determine whether a more highly restricted model is supported by the data, that is if a higher level of invariance is given. However, based on a Monte Carlo study, Chen (2007) proposed an alternative to the chi-square difference test, which leads too easily to a rejection of measurement invariance. He proposed that metric noninvariance is indicated by a change smaller than 0.01 in the comparative fit index (CFI), supplemented by a change smaller than 0.015 in the root mean square error of approximation (RMSEA), or a change smaller than 0.03 in the standardized root mean square residual (SRMR) compared with the configural invariance model. To guarantee scalar invariance, Chen (2007) proposed to inspect whether the change in CFI is smaller than 0.01, the change in RMSEA is smaller than 0.015, or the change in SRMR is smaller than 0.01, when moving from a metric to a scalar invariance model for sample sizes larger than 300 per group.

the same in both countries. Only in scenario (c) is it possible to observe the same score when the true score is identical in both countries.

2.3 Dealing with Noninvariance

While some studies were successful in establishing high levels of measurement invariance across groups (e. g. Davidov and Siegers 2010), applied research in the last decade has unfortunately shown that many scales fail to display comparability across countries or time. For example, Aleman and Woods (2016) and Sokolov (2018) recently showed that the Inglehart value scales are not comparable across all countries (but see Welzel and Ingelhart 2016); Ariely and Davidov (2010) found that public support for democracy cannot be compared across countries in the World Value Survey (WVS); Lomazzi (2018) demonstrated that gender-role attitudes are not comparable across all WVS countries; Davidov et al. (2008) showed that means of human values in the ESS may not be compared across all countries; Rudnev et al. (2018a) found that the means of Seeman's alienation scale are not comparable cross-nationally; Davidov (2009) discovered that the scale means of nationalism and patriotism are not comparable across countries participating in the International Social Survey Program (ISSP); and Coromina and Davidov (2013) concluded that social and political trust are not always comparable across countries and/or time points (see also Marsh et al. 2017). To tackle this problem, several solutions were proposed including technical procedures and theoretical approaches. What most of these approaches have in common is that they suggest releasing some of the strict constraints required by scalar invariance models while not compromising on the scales' comparability. Figure 2 presents an overview of the main recommendations in the form of a decision tree. We will discuss each of these approaches below.

Figure 2 depicts several analytical procedures that can be applied when non-invariance is obtained. First, several researchers suggested that instead of finding a model that meets measurement invariance requirements, noninvariance may be used as a unique opportunity and a useful source of information about cross-country differences (assuming that the theoretical model and measurement instrument are sound). They suggested that multilevel structural equation modeling (MLSEM) may be employed to explain why measurement invariance is not given (Davidov et al. 2012, 2016; Jak et al. 2013). MLSEM differentiates between the measurement model on the individual level and the measurement model on the country level. Researchers can use contextual variables such as the economic conditions in a country, policies, or the Human Development Index to try to explain, in a theoretically driven way, the lack of invariance of specific measurement items.

When the goal is, however, to identify a measurement model where measurement invariance holds to be able to conduct a meaningful cross-country comparison, then finding the reasons for noninvariance might not be sufficient. MLSEM may explain noninvariance, but does not necessarily ratify or correct for it. Thus, other researchers have proposed to release the strict constraints of measurement invariance testing so that the measurement invariance model fits the data better. They suggested that less strict models would be more realistic but still good enough for conducting meaningful comparisons. Some of them suggested testing for partial (metric or scalar)
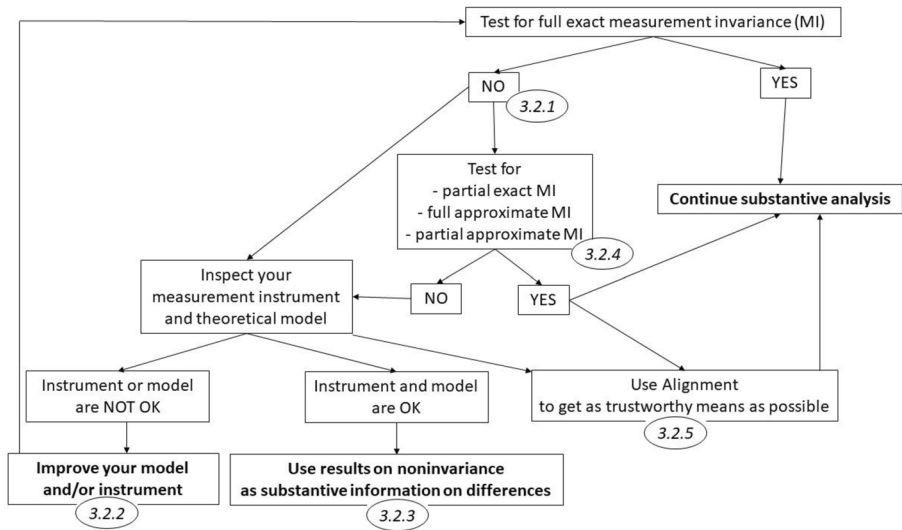
Fig. 2 Overview of the procedures and decisions in measurement invariance testing when exact measurement invariance is not supported by the data. *MI* measurement invariance; the numbers in the *ellipses* refer to subsections where we describe examples of a given procedure. (Author's own work)

measurement invariance, rather than for full (metric or scalar) invariance. These researchers argued that partial invariance may be sufficient for meaningful comparisons (Byrne et al. 1989; Steenkamp and Baumgartner 1998). Partial invariance is established when the parameters of at least two indicators (loadings at the metric level and loadings plus intercepts at the scalar level of the measurement) are equal across groups. In other words, the researcher identifies those items which are very different across groups, and releases them while ensuring that at least two items per scale have equal loadings and intercepts. Whereas this approach has been applied quite frequently by substantive researchers, it has also been criticized as insufficient to guarantee meaningful comparisons (see e.g. De Beuckelaer and Swinnen 2018; Steinmetz 2018). As a result, methodologists have recently come up with newer proposals on how to deal with the lack of measurement invariance.

Testing for full or partial measurement invariance, which we described above, is considered an *exact* approach because testing for either full or partial measurement invariance assumes that at least some of the parameters (loadings and/or intercepts) are *exactly* equal across groups. Recently, Muthén and Asparouhov (2013) suggested replacing the requirement of exact equality of measurement parameters with the requirement of an *approximate* equality of measurement parameters. They argued that approximate invariance may be sufficient for meaningful country comparisons (Muthén and Asparouhov 2013; van de Schoot et al. 2013). In other words, they suggested that it is sufficient when the parameters (factor loadings or intercepts) are more or less (rather than exactly) equal.

Tests for approximate measurement invariance can be performed in the Bayesian framework, where loadings and intercepts are treated as variables with a specific distribution. The parameters of this distribution (means and variance) are known

as *priors*, and can be defined by the researcher based on previous knowledge or assumptions (Muthén and Asparouhov 2013). In the exact measurement invariance approach, loadings and intercepts are constrained to be exactly equal, and consequently the *differences* between them are assumed to be zero by definition, with a zero variance. In the approximate approach, one assumes that the cross-country mean of the *differences* between loadings or intercepts equals zero, but that small variations for these mean differences are allowed. The amount of the variation that is allowed is indicated by the variance of the cross-country parameter differences that can be imposed on the model. Several simulation studies have shown that variances equal to 0.01 or 0.05 in the distribution of the cross-country differences in loadings or intercepts probably do not bias substantive conclusions for comparative research (Muthén and Asparouhov 2013; van de Schoot et al. 2013). More complete research on the tolerated level of variability of the differences between parameters is still missing, so it is still not clear in the literature which parameter differences may be tolerated and which may be too large to guarantee meaningful comparisons.

The test for approximate measurement invariance (implemented in the Mplus structural equation modeling software package: Muthén and Muthén 1998–2014; or in lavaan: Merkle and Rosseel 2016) provides researchers with two types of measures with which to assess the quality of the models. The first type measures the global fit of the model and includes the posterior predictive *p*-value (ppp) and the credibility interval (CI) for the difference between the observed and replicated chi-square values (but see Marsh et al. 2017 for the newly developed pppp measure, which is yet to be implemented in a commercial software package). According to Muthén and Asparouhov (2013) and van de Schoot et al. (2013), the Bayesian model fits the data when the ppp is not significant (larger than zero) and the CI contains a zero.

The second type of measure obtained in approximate measurement invariance testing is more detailed. It includes a list of all parameters that are noninvariant in each group and the significance of the deviation. Using this list, researchers can identify items that are particularly noninvariant, that is, deviate from invariance in many groups. In the next step, researchers may decide whether to drop these items, drop countries that have a particularly high number of noninvariant items, or increase the tolerated variance of parameter differences in the model (see e. g. Davidov et al. 2015). When following this strategy it is indeed, as indicated above, not clear to what extent this variance may be increased under different conditions without placing the meaningfulness of cross-country comparisons at risk.

Another approximate approach which has recently been proposed in the literature is alignment optimization (Asparouhov and Muthén 2014). This approach allows, under certain conditions, an unbiased comparison of means using an optimization process, even in the presence of noninvariance. The optimization process computes the most trustworthy latent means (Asparouhov and Muthén 2014; Cieciuch et al. 2018; Muthén and Asparouhov 2014, 2017) *without* constraining loadings and intercepts to be equal across groups (i. e. by using a configural invariance model only) (Asparouhov and Muthén 2014; Muthén and Asparouhov 2014). Thus, the means are estimated while taking into account real differences in loadings and intercepts. The alignment optimization method discovers the most optimal measurement invariance

pattern, in which a relatively small number of large noninvariant parameters—and many approximately invariant parameters—are present, rather than imposing exact equality constraints on all parameters. Asparouhov and Muthén (2014) compare this procedure to a rotation in exploratory factor analysis which simplifies the loading matrix without modifying the model fit. The alignment optimization procedure identifies the noninvariant parameters (loadings and intercepts). It can be performed on any multiple group model, and also when measurement properties are significantly different across groups.

Muthén and Asparouhov (2014) proposed a rule of thumb for determining when it is safe to continue performing the mean comparisons. They suggested that such a comparison is meaningful when up to about 25% of the parameters (factor loadings and intercepts) are noninvariant. However, this recommendation is based on a very limited set of simulation studies. It is still unclear whether this rule of thumb is too strict or too liberal. It is also not clear whether 25% of all measurement parameters (factor loadings and intercepts) may be noninvariant, or whether this rule of thumb should be applied separately for each set of parameters. Further simulation studies are required in order to determine how many parameters may be noninvariant without risking the meaningfulness of the mean comparisons across countries when using alignment. Notwithstanding these limitations, the alignment approach is particularly useful for substantive research because it is easier to apply than other tests of measurement invariance, especially when the number of countries to be compared is large.

In the next section we will illustrate the procedure on the human values measurements in the ESS. We will perform a large-scale test of the measurement invariance properties of human values across all countries which participated in seven rounds of the ESS. This is, to the best of our knowledge, the largest measurement invariance test applied to survey data using the alignment procedure (for a similarly large study using Alignment, see Munck et al. 2017). Human values scores obtained using this procedure may be potentially relevant when researchers are interested in comparing value scores across ESS countries *and* time points.

3 Measurement (Non)Invariance of Human Values as Measured in the ESS

Value preferences are considered in many sociological and social-psychological studies to be a dimension of major importance to describe persons, groups, and societies, and to explain attitudes, intentions, and behavior (Durkheim 1964 [1897]; Hitlin and Piliavin 2004; Hofstede 2000; Inglehart and Baker 2000; Kluckhohn 1951; Rokeach 1973; Schwartz 1992; Weber 1958 [1905]). Although several researchers developed different value theories and proposed various scales to measure them, the circular model of values proposed by Schwartz (1992; Schwartz et al. 2012; Steinmetz et al. 2012) is probably the one most frequently applied in the social sciences. Measures of the Schwartz human values scale are included in all seven rounds of the ESS

(Schwartz 2003)[2]. The model has received empirical support in many studies worldwide (Bilsky et al. 2011; Steinmetz et al. 2012; Rudnev et al. 2018b), in samples of adults and children (Cieciuch et al. 2016; Döring et al. 2015) using both cross-sectional and longitudinal data. There is no doubt that including value measurements in the ESS offers a unique opportunity to analyze value priorities and change in many European countries. However, a methodological precondition for performing such analyses meaningfully is to ensure that the value measurements in the ESS are invariant across countries and points in time, and therefore comparable. Obtaining the most trustworthy value means in the ESS is thus of paramount importance for substantive research. Below, we first present the value model to be examined and briefly review previous results on measurement invariance testing of the ESS value scale. Next, we build on previous results and apply the alignment optimization to the values in order to obtain the most trustworthy value means in the ESS.

3.1 Basic Human Values—the Circular Model and the Value Measurement in the ESS

In the circular model proposed by Schwartz (1992; Schwartz et al. 2012), values are defined as broad, transsituational goals that vary in importance and serve as guiding principles in the life of a person or group. The number of basic values according to the theory is limited. People or groups differ in their *hierarchy* of values rather than in the set of values they consider important. The main claim of Schwartz' (1992; Schwartz et al. 2012) circular model is that all values can be located on the circle according to the motivation they express. Neighboring values are based on a similar motivation and can be pursued in the same action. Values located on opposite sides of the circle express conflicting motivations in the sense that they cannot be pursued concurrently when performing the same behavior.

The value circle represents a circular motivational continuum. Thus, the circle can be divided in many ways according to the research goals and the measurement instrument used. Traditionally, the value circle was divided into ten values that form four higher-order values, as presented in Fig. 3 (the first two internal circles) (Schwartz 1992). The four higher-order values describe self-transcendence vs. self-enhancement values and conservation vs. openness to change values. The refined model (Schwartz et al. 2012) enables a division into 19 more narrowly defined values. At the same time, the value circle can also be divided into more general and broadly defined values, which are represented by the third and the fourth circles in Fig. 3. One division differentiates between anxiety-free growth values which oppose anxiety avoidance self-protection values. Another division differentiates between socially focused values vs. values with a personal focus.

Schwartz developed several methods to measure the ten values and the four higher-order values. The Portrait Value Questionnaire (PVQ) is the scale that is used most frequently (for a review, see Schwartz and Cieciuch 2016). The basic

---

[2] Measures of the Schwartz values are also included in other international surveys such as the World Values Survey or the U.S. General Social Survey (for further details, see http://www.worldvaluessurvey.org/wvs.jsp; http://www.norc.org/Research/Projects/Pages/general-social-survey.aspx).

Fig. 3 The circular motivational continuum of values (Source: Cieciuch et al. 2015)
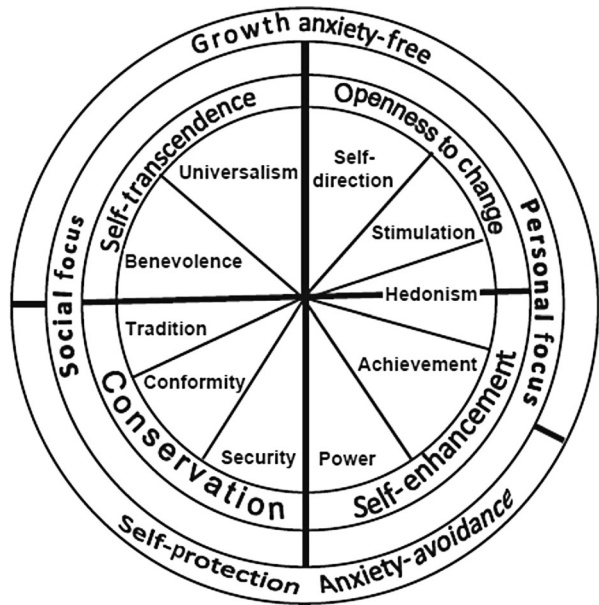
version of the questionnaire consists of 40 items (Schwartz et al. 2001) that describe other people. Respondents have to assess the similarity between themselves and the people described. A shortened version of the questionnaire consisting of only 21 items (PVQ-21: Schwartz 2003) was included in the ESS from the very beginning. Each item in the PVQ-21 is composed of two sentences which describe a portrait from a male or female perspective. The portraits contain goals, aspirations or desires that point to the importance of a value. For each item, the respondents in the ESS answer the question "How much like you is this person?", with a response scale ranging from 1 (not like me at all) to 6 (very much like me). The items included in the ESS to measure the ten values and the four higher-order dimensions are presented in Table 1.

3.2 Measurement Invariance of Values in the ESS and Dealing with Noninvariance: a Review and an Illustration

Values were measured in the ESS using the same questionnaire translated into different languages and by applying mostly face-to-face interviews (for exceptions and documentation of the data collection procedures, see the ESS website at www.europeansocialsurvey.org). This careful procedure is however not sufficient to ensure the comparability of values in the ESS, and their measurement invariance properties have to be examined before using their ESS measures in comparative research meaningfully. Below we will follow the steps depicted in Fig. 2 that guide the assessment of measurement invariance and how to deal with noninvariance.

Table 1 The ten basic human values, four higher-order values, and the Portrait Value Questionnaire (PVQ)-21 items in the European Social Survey (ESS; female version) to measure these values with their labels. (The number next to each question item refers to the placement of that item in the PVQ-21 questionnaire.). (Author's own work)

Item label	Items
1. Self-enhancement—Achievement	
Ipshabt	4. It's important to her to show her abilities. She wants people to admire what she does
Ipsuces	13. Being very successful is important to her. She hopes people will recognize her achievements
2. Self-enhancement—Power	
Imprich	2. It is important to her to be rich. She wants to have a lot of money and expensive things
Iprspot	17. It is important to her to get respect from others. She wants people to do what she says
3. Self-transcendence—Benevolence	
Iphlppl	12. It's very important to her to help the people around her. She wants to care for their well-being
Iplylfr	18. It is important to her to be loyal to her friends. She wants to devote herself to people close to her
4. Self-transcendence—Universalism	
Ipeqopt	3. She thinks it is important that every person in the world should be treated equally. She believes everyone should have equal opportunities in life
Ipudrst	8. It is important to her to listen to people who are different from her. Even when she disagrees with them, she still wants to understand them
Impenv	19. She strongly believes that people should care for nature. Looking after the environment is important to her
5. Conservation—Conformity	
Ipfrule	7. She believes that people should do what they're told. She thinks people should follow rules at all times, even when no-one is watching
Ipbhprp	16. It is important to her always to behave properly. She wants to avoid doing anything people would say is wrong
6. Conservation—Tradition	
Ipmodst	9. It is important to her to be humble and modest. She tries not to draw attention to herself
Imptrad	20. Tradition is important to her. She tries to follow the customs handed down by her religion or her family
7. Conservation—Security	
Impsafe	5. It is important to her to live in secure surroundings. She avoids anything that might endanger her safety
Ipstrgv	14. It is important to her that the government ensures her safety against all threats. She wants the state to be strong so it can defend its citizens
8. Openness—Self-direction	
Ipcrtiv	1. Thinking up new ideas and being creative is important to her. She likes to do things in her own original way
Impfree	11. It is important to her to make her own decisions about what she does. She likes to be free and not depend on others
9. Openness—Stimulation	
Impdiff	6. She likes surprises and is always looking for new things to do. She thinks it is important to do lots of different things in life
Ipadvnt	15. She looks for adventures and likes to take risks. She wants to have an exciting life

Table 1 (Continued)

Item label	Items
10. Openness—Hedonism	
Ipgdtim	10. Having a good time is important to her. She likes to "spoil" herself
Impfun	21. She seeks every chance she can to have fun. It is important to her to do things that give her pleasure

3.2.1 Lack of Exact Scalar Measurement Invariance

First, in 2008, Davidov et al. published a seminal paper testing for measurement invariance of values across 20 countries using data from the 1st round of the ESS. The findings were also replicated for the 2nd and 3rd rounds of the ESS (Davidov 2008, 2010). The two main results obtained by Davidov and colleagues are as follows: (1) Only seven values can be differentiated in most of the countries. Specifically, three pairs of adjacent values need to be unified: power with achievement, benevolence with universalism, and conformity with tradition. (2) Only metric invariance is established for the seven values. Scalar invariance across countries was not supported by the data. Thus, the findings suggested that cross-country mean comparisons of the values as measured in the ESS may be problematic. However, Davidov (2008, 2010) demonstrated that values displayed scalar invariance over time, suggesting that their means may be used for longitudinal comparisons in the countries of interest. Figure 2 presented various procedures to deal with measurement noninvariance. In the next steps, these procedures were applied in value research in a number of different ways that we briefly describe below.

3.2.2 Refining the Theory

In 2012, Schwartz and colleagues used information about noninvariance to improve the model and its measurement instrument. They refined the value theory to include 19 instead of ten values, and developed a new 57-item scale that was better suited to measuring the single values (Knoppen and Saris 2009; Beierlein et al. 2012). This refined theory did not contradict its older version. After all, the original theory suggested that one may divide the value circle into more or less specific values depending on the measurement instruments one has (Cieciuch et al. 2013). Cieciuch et al. (2014b) and Cieciuch et al. (2014a) demonstrated that this version of the theory and its measurement instrument can distinguish between all single values, and the refined instrument possesses much better cross-country measurement invariance properties than the ESS scale does.

3.2.3 Using Noninvariance as a Source of Information on Cross-cultural Measurement Differences

Another approach in Fig. 2 proposed using findings of cross-country noninvariance of the ESS value scale as an important source of information on country differences. Applying multilevel structural equation modeling, Davidov et al. (2012) showed

how to explain the noninvariance of a specific item measuring universalism which tapped into the importance of protecting the environment. This item was particularly noninvariant across European countries. The authors found that it was endorsed more strongly by residents of European countries that had a lower Human Development Index score. These populations apparently considered clean water and air to be a matter of survival and health, whereas these aspects appear to be taken for granted by individuals residing in more developed countries.

3.2.4 Applying Approximate Measurement Invariance Testing Techniques

Finally, the third approach as presented in Fig. 2 questioned the procedure of testing for exact measurement invariance of human values. Referring to new developments in Bayesian analysis, it raised the possibility that previous measurement invariance tests on the ESS value scale may have been too strict. After all, as discussed earlier, a small degree of noninvariance may not necessarily bias substantive conclusions in comparative studies. Thus, instead of relying on exact measurement invariance, Cieciuch et al. (2017) and Zercher et al. (2015) proposed that the ESS value scale be subjected to tests of approximate measurement invariance. They applied approximate measurement invariance procedures based on Muthén and Asparouhov (2013) and van de Schoot et al. (2013), who suggested that factor loadings and intercepts need not necessarily be *exactly* equal across countries. Instead, they may be almost equal.

Zercher et al. (2015) applied, for the first time, the approximate measurement invariance approach to values measured in the ESS data. They focused on only one value—universalism—and ran the approximate measurement invariance test across 15 countries which participated in all six available ESS rounds simultaneously, resulting in a comparison of ($15 \times 6 =$) 90 groups. They showed that, whereas scalar measurement invariance in the (traditional) exact approach was established across 37 groups, approximate measurement invariance could be established across no fewer than 73 groups, thus challenging previous findings on lack of invariance. Hence, whereas the universalism value was not measurement invariant across all country/time-point combinations, it was comparable across most of them.

Cieciuch et al. (2017) went one step further and subjected the full ESS value scale to an approximate invariance test for all values. However, due to the highly complex nature of the model, they did not run a simultaneous test across all 15 countries and six measurement time points of the ESS. Instead, they tested for approximate measurement invariance across the 15 countries *within* each ESS round. Furthermore, given that it was not possible to measure ten values, Cieciuch et al. conducted the approximate measurement invariance test *separately* on each higher-order value using the magnifying glass strategy (Cieciuch and Schwartz 2012) as illustrated in Fig. 4. This approach helped to avoid introducing cross-loadings which are inherent in the theory when all values are used simultaneously in a single model. Cieciuch et al. (2017) established approximate measurement invariance successfully for the
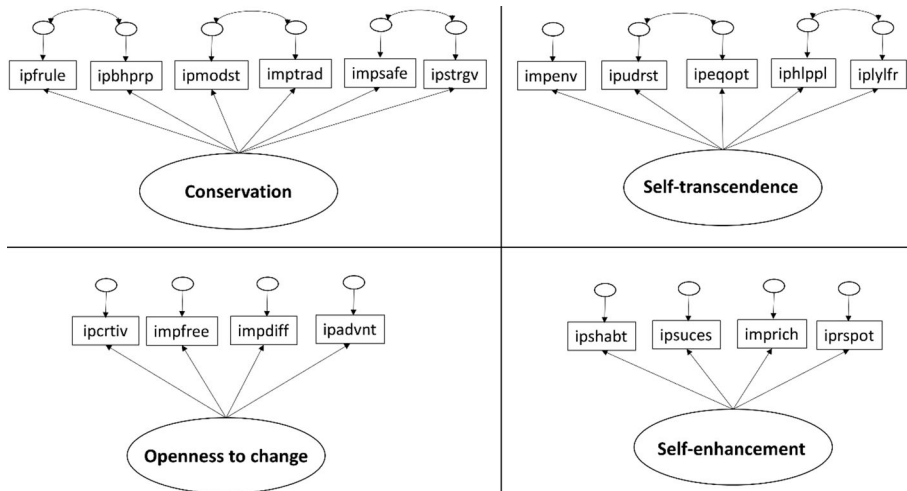
Fig. 4 Measurement models for four higher-order values. Item abbreviations are presented in Table 1. Error correlations are allowed between items that were originally designed to measure the same single value in order to take into account their common variance. (Author's own work)

higher-order values in most countries.[3] While these tests resulted in higher levels of measurement invariance than previous tests did, it is yet to be examined whether the tests were too liberal in the sense that they allowed too much variability of measurement parameters across countries. Indeed, researchers still need to determine how much variability may be allowed in approximate measurement invariance testing.

3.2.5 Applying the Alignment Optimization

We wish to illustrate below another technique presented in Fig. 2 (building on Cieciuch et al. 2017), namely the alignment optimization. This is also a method which allows for approximate rather than exact measurement invariance. However, instead of imposing exact or approximate equality constraints on the measurement parameters, it searches for the pattern of loadings and intercepts that minimize noninvariance. As a result, it produces the most trustworthy value means from the data[4]. For the illustration we used the same human values models as those

---

[3] The authors excluded the value 'hedonism' from this model. According to the theory, this value is located between openness to change and self-enhancement. Including this value in either of the models resulted in a significant reduction in model fit.

[4] The scale of latent variables is unknown, and hence their variance is also unknown (by definition, these variables are unobserved). Therefore, in order to identify the model, researchers need to apply some restriction for the estimation: either restricting the variance of the latent variable to an arbitrary value (typically it is then restricted to 1 in all groups), or fixing the scale of the latent variable by restricting the factor loading of one of the items (the so-called anchor item) to 1 in all groups. When doing so, it is important to guarantee that such a restriction fits the data at hand. In the former case, the restriction implies an implicit assumption that the latent variance is equal across groups. In the latter case, the restriction implies that the factor loading of the anchor item is indeed equal across groups. In both cases, researchers need to make sure that the assumption holds, for example by inspecting which of these parameters (factor loading of

Table 2 Number of respondents included in the analysis for each round and country. (Author's own work)

	1st Round 2002–2003	2nd Round 2004–2005	3rd Round 2006–2007	4th Round 2008–2009	5th Round 2010–2011	6th Round 2012–2013	7th Round 2014–2015
Belgium (BE)	1819	1734	1767	1704	1674	1809	1720
Denmark (DK)	1457	1457	1451	1554	1548	1610	1475
Finland (FI)	1758	1692	1645	1898	1638	2142	2044
Germany (DE)	2785	2800	2828	2697	2943	2910	2982
Hungary (HU)	1564	1407	1409	1388	1404	1919	1460
Ireland (IE)	1838	1139	1582	1682	2295	2498	2288
Netherlands (NL)	2301	1824	1814	1693	1754	1788	1802
Norway (NO)	1806	1543	1533	1374	1518	1598	1408
Poland (PL)	1982	1621	1629	1544	1675	1818	1550
Portugal (PT)	1417	1987	2117	2220	2035	2062	1209
Slovenia (SI)	1390	1297	1329	1172	1238	1159	1113
Spain (ES)	1638	1544	1802	2520	1862	1820	1857
Sweden (SE)	1677	1663	1585	1539	1457	1799	1755
Switzerland (CH)	2009	2084	1758	1764	1467	1453	1489
United Kingdom (GB)	1748	1806	2301	2230	2315	2212	2176
Total	25,441	23,792	24,249	24,749	24,508	26,385	26,328

Only countries that participated in all seven European Social Survey (ESS) rounds are included in the analysis

presented in Fig. 4, and analyzed 15 countries which participated in the first seven ESS rounds (2002/2003, 2004/2005, 2006/2007, 2008/2009, 2010/2011, 2012/2013, and 2014/2015). We utilized the Mplus software package Version 7.3 (Muthén and Muthén 1998–2014). The syntax of the analysis is available from the first author upon request.

Table 2 presents the number of respondents in each round and country included in the analysis. We followed the recommendations provided on the ESS website, and only considered respondents with no more than five missing values and no more than 16 identical responses for the 21 value items. As a result, the analysis included a total of 175,452 respondents. The remaining item nonresponse was dealt with by using the full information maximum likelihood (FIML) procedure (see Schafer and Graham 2002). The ESS website (www.europeansocialsurvey.org) provides documentation about the data collection procedure and permits the data and accompanying material to be downloaded. Table 3 presents the global fit measures for the configural invariance model of each higher-order value across 105 groups (15 countries × 7 rounds). The global fit measures suggested that the four models had a good fit to the data. All 1995 factor loadings of the higher-order values as depicted in Fig. 3 (six for conservation, five for self-transcendence, and four for openness to change and self-enhancement, respectively, in 105 country/time combinations) were significant, and

one of the items or the latent variable variance) are indeed most similar across groups, and choose the restriction which best corresponds with the data at hand (see also Brown 2015, p. 271).

Table 3 Model fit indices of the multiple-group confirmatory factor analyses across all countries and waves for each higher-order value (configural invariance model). (Author's own work)

	χ^2	df	CFI	RMSEA	SRMR
Self-enhancement	741.5	210	0.996	0.038 [0.035–0.041]	0.011
Self-transcendence	1907.0	315	0.989	0.053 [0.051–0.055]	0.015
Conservation	2758.2	630	0.988	0.043 [0.042–0.045]	0.017
Openness to change	4278.3	210	0.961	0.104 [0.101–0.107]	0.030

df degrees of freedom, *CFI* Comparative Fit Index, *RMSEA* Root Mean Square Error of Approximations; *SRMR* Standardized Root Mean Square Residuals

Table 4 The number (and percentage) of noninvariant loadings and intercepts identified in the alignment optimization for each higher-order value. (Author's own work)

	Loadings, %	Intercepts, %	Average, %
Conservation	12 (73/630)	47 (294/630)	29 (367/1260)
Self-transcendence	12 (61/525)	45 (234/525)	28 (295/1050)
Openness (without hedonism)	14 (60/420)	54 (228/420)	34 (289/840)
Self-enhancement	6 (26/420)	59 (247/420)	33 (273/840)

almost all of them were higher than 0.4 (Brown 2015). Only very few loadings (i.e. ten, which corresponded to about 0.5% of the total number of loadings) were slightly lower than 0.3. We nevertheless retained these items because these cases were very few in number, and we wanted our measurement models to correspond to the theory. Further information on the factor loadings and the measurement models may be obtained from the first author upon request.

Table 4 presents the number of noninvariant loadings and intercepts reported in the Mplus output (Muthén and Muthén 1998–2014). This information is important to determine whether the means computed by the alignment procedure are trustworthy. As previously indicated, Muthén and Asparouhov (2014) suggested that a cutoff criterion of about 25% of noninvariant parameters (factor loadings and intercepts) may not be exceeded. The average amount of noninvariance was lowest for self-transcendence (28%), followed by conservation (29%) and self-enhancement (33%), and was highest for openness to change (34%), and therefore slightly above the 25% cutoff criteria suggested by Muthén and Asparouhov (2014). As displayed in Table 4, noninvariance was particularly evident for intercepts. About half of the intercepts for the four higher-order values were not invariant according to the output. This finding corresponds to previous findings testing for measurement invariance of the

---

[5] Many studies evidenced that intercepts were not equal across groups, and that it was easier to guarantee equal factor loadings than equal intercepts when comparing different countries (see e.g. Davidov et al. 2014). In other words, it was often easier to establish metric invariance than scalar invariance. Different intercepts may also reflect different country-specific survey strategies, which in turn may result in different response patterns across countries.

ESS values in which factor loadings were rather invariant but intercepts were not.[5] Therefore, although alignment provides the most trustworthy means possible with the data at hand, the findings for the means, and particularly those for openness to change and self-enhancement, where the number of noninvariant intercepts exceeded 50%, should be treated with great caution.

Tables 5, 6, 7 and 8 in the Appendix present the means for each higher-order value in each combination of countries/rounds. Latent means are meaningless per se and should be interpreted in comparison to other country means. Countries with more than 25% noninvariant parameters (factor loadings and intercepts) are marked ([a]).[6] Furthermore, the tables present the country rankings within each round. This makes it possible to conclude from each table how, at some measurement time point, a country compares to any other country at the same or at any other time of measurement. For example, Table 5 shows that Poland displays the highest conservation value means in the first two rounds. In Round 7, Poland displays a higher mean compared to its measures in the first two rounds. However, it now ranks second (rather than first) in this last round because Slovenia is even more conservative in this round. Slovenia displays a large increase in its conservation scores between the 1st and 7th rounds, increasing from –0.877 to –0.089. As another example, turning to Table 6, in which self-transcendence scores are presented, we see that Switzerland considerably increased its self-transcendence scores when moving from the 1st round (–0.149) to the 7th round (0.187). However, even though Switzerland ranked highest in self-transcendence in the 1st round, it comes in the third place in the 7th round. The reason is that other countries displayed even more pronounced increases in self-transcendence during this period. Researchers who are interested in specific countries and time points may examine the scores for specific values in order to draw conclusions about value change and value development in countries. Thus, this illustration demonstrates that researchers can quite easily examine measurement invariance properties in complex and large cross-national datasets—such as in the large-scale investigation of human values considered here—and can estimate how the countries' mean scores compare with one another. In the final section, we will first reflect on the different approaches that can be implemented to test for measurement invariance discussed so far, and then consider the extent to which we may rely on mean scores when exact measurement invariance is not given by the data.

4 Summary and Discussion

The last decade has witnessed an increase in the number of published studies that included the testing of the measurement invariance of various scales. This is an important development in the literature. After all, in cross-national comparative research,

---

[6] Tables that display more highly specific information about the (non)invariance pattern for each higher-order value may be obtained from the first author on request. They present the number of noninvariant loadings and intercepts for each item and country. One way to estimate the amount of bias, discussed in Oberski (2014), is to perform a sensitivity analysis.

meaningful comparisons of means or associations between theoretical constructs that are of interest can only be performed when the same constructs are measured in the same way across the various nations under study. This assumption is also referred to as measurement invariance (or measurement equivalence). It must be satisfied in order to draw any meaningful conclusions in CNCR settings, either for direct comparisons of means or associations, or for multilevel analysis which implicitly assumes that scores are comparable across the units of analysis (e.g. countries).[7]

There are various statistical procedures to perform measurement invariance tests, but the most common is the multiple group confirmatory factor analysis. In this study we reviewed, in a nontechnical manner, the methodological literature on measurement invariance testing. We explained what it is, how to test for it, and what to do when measurement invariance across countries is not given in the data.

Indeed, in many studies where measurement invariance was examined, it was not possible to achieve sufficient levels of invariance. Failing to reach measurement invariance may threaten the meaningful interpretation of comparisons across nations. Lack of measurement invariance could result in methodological artifacts that may be responsible for misleading differences observed between country scores. At the same time, similarities observed across nations when measurement invariance is not present might also be misleading because they could mask true differences that cannot be observed due to a lack of comparability.

Several approaches on how to deal with measurement noninvariance, including testing for partial (rather than full) invariance, approximate (rather than exact) invariance and alignment, have been recently proposed in the literature. Alignment may be particularly interesting for applied researchers because it allows the most trustworthy means to be estimated in the data, even when measurement invariance is not present, since it provides researchers with the tools to assess whether and to what extent the scores are nevertheless comparable. It is also relatively easy to apply when the number of groups is very large.

We illustrated the use of the alignment approach with a large dataset consisting of seven rounds from the ESS (2002–2015) by estimating the most trustworthy means of higher-order human values in Schwartz' (1992) model across all available ESS rounds, even when strict measurement invariance was not given in the data. We included in the illustration the 15 countries that participated in seven ESS rounds (i.e. 105 groups). Unfortunately, it was not possible to establish measurement invariance of the values across all ESS countries and points in time. The alignment procedure nevertheless allowed us to estimate the means—even in the absence of measurement invariance. The scores revealed a significant variability of the human values scores both across countries and over time. Some changes in country rank order were found across rounds for all values. We also observed that in many cases, countries ranking highest or lowest on specific values tended to remain in these positions

---

[7] Methodologists also discuss the topic of isomorphism, which refers to equivalent construct meaning across levels of analysis. In other words, it refers to the presence or absence of measurement invariance across levels, for example across individuals and countries. However, examining isomorphism in cross-national data settings is beyond the scope of the present study (for a further discussion, see e.g. Guenole 2016; Muthén 1994; or Ruelens et al. 2016).

across some of the rounds. However, there were several exceptions. The findings therefore suggest that while values are rather stable (in terms of country rankings), they did change slowly over time. Researchers interested in investigating specific values in specific countries and points in time may repeat the analysis for their specific countries of interest.

Our illustration underlines several limitations in measurement invariance testing in general, and in the alignment procedure in particular. While it is obvious that measurement invariance is a necessary condition for meaningful cross-national comparison, empirical studies frequently demonstrate that full invariance cannot be reached. As a result, based on simulation studies, recent developments have proposed relaxing some of the parameter equality requirements in measurement invariance testing by allowing for partial or approximate (rather than full or exact) invariance. These approaches suggest freeing some of the measurement parameters (factor loadings and/or intercepts) or only requiring the measurement parameters to be approximately (rather than exactly) equal. However, it is still to be studied to what extent such relaxations of the requirements in the statistical test are legitimate. It is not yet clear how many equality constraints may be freed and how many measurement parameters must remain equal across countries in the partial invariance test. It is also not yet fully clear how much variability in the differences between the measurement parameters may be allowed in the approximate invariance test. Finally, it is not yet known whether one may allow for more or less than 25% of the measurement parameters to differ in the alignment optimization procedure. After all, if excessively generous tolerances are applied in the tests, one may run the risk of rendering the scores nonequivalent. Indeed, whereas alignment optimization provides the most trustworthy means, even when several measurement parameters are not invariant, it still does not guarantee that all means are comparable. Thus, it is desirable to complement such more liberal tests with robustness tests in order to inquire whether freeing certain parameters leads to invalid conclusions (Oberski 2014). Notwithstanding these limitations, applied sociologists and social psychologists are encouraged to perform the analyses that we presented above. Researchers can be confident that their scores are comparable cross-nationally only after measurement invariance tests have been performed.

Measurement invariance is important not only for cross-country comparisons, but also for multilevel modeling. A multilevel analysis relies on the assumption that country-level effects (of, e. g., composite scores of trust or threat due to immigration) are comparable when estimating a random slope. It also assumes that means are comparable when one estimates a random intercept. Thus, at least partial metric invariance is necessary for estimating random slopes, and at least partial scalar invariance is needed for estimating random intercepts. Indeed, nearly all multilevel analyses that use composite scores of multiple indicators assume but do not test the assumption of measurement invariance of these scores. To the best of our knowledge, there is no simulation study which estimates whether and to what extent multilevel analysis findings are biased if measurement invariance across groups is absent. Evaluating the implications of the lack of measurement invariance for estimating multilevel models using simulation studies is an important direction for future research.

Two final words of caution: First, one should take into account that even if metric and scalar invariance are given, there might be differences in meaning which cannot all be detected by quantitative techniques (Meitinger 2017). In such cases, one could consider applying mixed-methods approaches to explain instances of measurement noninvariance by combining measurement invariance tests with different qualitative techniques. Second, it might be important to check the heterogeneity of all the national population samples, as the intracountry mean differences might be greater than the intercountry differences (Magun et al. 2016; Schmidt-Catran et al. 2019). We hope that our review and illustration can provide researchers with tools to address the methodological challenges of comparability when using diverse scores in cross-national comparative research settings.

Appendix

Table 5 Importance of conservation values: Country rankings across all European Social Survey (ESS) rounds (The value means estimated by the alignment optimization are in parentheses.). (Author's own work)

Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7
PL[a] (−0.413)	PL[a] (−0.455)	HU[a] (−0.47)	ES[a] (−0.314)	SI[a] (−0.447)	SI[a] (0)	SI[a] (−0.089)
ES[a] (−0.488)	HU[a] (−0.489)	ES[a] (−0.493)	HU[a] (−0.615)	HU[a] (−0.477)	PL[a] (−0.28)	PL[a] (−0.274)
HU[a] (−0.568)	IE[a] (−0.741)	PL[a] (−0.571)	PL[a] (−0.711)	PL[a] (−0.51)	ES[a] (−0.531)	HU[a] (−0.517)
IE[a] (−0.813)	ES[a] (−0.82)	SI[a] (−0.938)	SI[a] (−0.768)	ES[a] (−0.58)	HU[a] (−0.642)	ES[a] (−0.526)
SI[a] (−0.877)	SI[a] (−0.999)	IE[a] (−0.962)	IE[a] (−0.849)	IE[a] (−1.017)	IE[a] (−0.792)	IE[a] (−0.812)
PT[a] (−1.042)	BE (−1.164)	PT[a] (−1.016)	BE (−1.207)	GB (−1.081)	GB (−0.863)	CH (−1.086)
BE (−1.317)	GB (−1.184)	BE (−1.217)	GB (−1.302)	CH (−1.138)	CH (−1.008)	GB (−1.1)
FI (−1.345)	PT[a] (−1.266)	GB (−1.265)	PT[a] (−1.367)	DE[a] (−1.157)	BE (−1.052)	BE (−1.144)
GB (−1.348)	FI (−1.419)	DE[a] (−1.441)	DE[a] (−1.449)	PT[a] (−1.186)	PT[a] (−1.162)	PT[a] (−1.297)
DE[a] (−1.364)	DE[a] (−1.445)	FI (−1.463)	CH (−1.473)	BE (−1.318)	DE[a] (−1.17)	DE[a] (−1.303)
CH (−1.594)	CH (−1.494)	CH (−1.465)	FI (−1.503)	FI (−1.432)	FI (−1.426)	FI (−1.477)
NL (−1.631)	NL (−1.627)	NL (−1.721)	NL (−1.734)	DK (−1.601)	NL (−1.594)	DK (−1.644)
DK (−1.788)	DK (−1.845)	NO (−1.875)	NO (−1.998)	NL (−1.683)	DK (−1.699)	NO (−1.661)
NO (−2.154)	NO (−1.897)	DK (−2.009)	DK (−2.015)	NO (−1.942)	NO (−1.722)	NL (−1.688)
SE (−2.392)	SE (−2.312)	SE (−2.311)	SE (−2.31)	SE (−2.196)	SE (−1.928)	SE (−2.096)

Country abbreviations listed in Table 2
[a]Country/round combinations where the number of noninvariant parameters exceeds 25%. Consequently, their means must be analyzed with caution

Table 6 Importance of self-transcendence values: Country rankings across all European Social Survey (ESS) rounds (The value means estimated by the alignment optimization are in parentheses). (Author's own work)

Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7
CH[a] (−0.149)	CH[a] (0)	ES[a] (0.113)	ES[a] (0.223)	CH[a] (0.192)	ES[a] (0.318)	SI[a] (0.267)
ES[a] (−0.152)	ES[a] (−0.216)	CH[a] (0.082)	CH[a] (0.144)	ES[a] (0.185)	SI[a] (0.25)	ES[a] (0.205)
BE (−0.283)	IE (−0.255)	BE (−0.229)	IE (−0.13)	SI[a] (0.024)	CH[a] (0.127)	CH[a] (0.187)
DE[a] (−0.37)	BE (−0.289)	FI[a] (−0.404)	BE (−0.221)	DE[a] (−0.081)	DE[a] (0.035)	DE[a] (0.101)
FI[a] (−0.409)	HU[a] (−0.391)	DK[a] (−0.411)	DK[a] (−0.246)	DK[a] (−0.258)	FI[a] (−0.129)	SE (−0.064)
DK[a] (−0.433)	PL[a] (−0.454)	SI[a] (−0.42)	DE[a] (−0.247)	BE (−0.274)	SE (−0.137)	FI[a] (−0.095)
IE (−0.482)	DE[a] (−0.474)	PL[a] (−0.444)	SI[a] (−0.311)	HU[a] (−0.306)	DK[a] (−0.185)	DK[a] (−0.226)
PL[a] (−0.614)	FI[a] (−0.474)	GB (−0.449)	FI[a] (−0.355)	SE (−0.353)	BE (−0.188)	BE (−0.243)
SI[a] (−0.65)	DK[a] (−0.484)	DE[a] (−0.459)	GB (−0.431)	GB (−0.393)	GB (−0.2)	PL[a] (−0.261)
NL (−0.658)	GB (−0.488)	IE (−0.464)	NL (−0.513)	PL[a] (−0.399)	PL[a] (−0.206)	GB (−0.29)
HU[a] (−0.7)	NL (−0.58)	NL (−0.529)	PL[a] (−0.537)	NL (−0.484)	HU[a] (−0.276)	HU[a] (−0.372)
GB (−0.737)	SI[a] (−0.598)	HU[a] (−0.549)	HU[a] (−0.609)	FI[a] (−0.486)	IE (−0.284)	IE (−0.436)
PT[a] (−0.824)	NO (−0.905)	NO (−0.753)	SE (−0.698)	NO (−0.496)	NL (−0.436)	NO (−0.502)
SE (−0.99)	SE (−0.917)	PT[a] (−0.784)	NO (−0.782)	IE (−0.599)	NO (−0.441)	PT[a] (−0.568)
NO (−0.996)	PT[a] (−1.204)	SE (−0.865)	PT[a] (−1.136)	PT[a] (−0.976)	PT[a] (−0.935)	NL (−0.592)

Country abbreviations listed in Table 2

[a]Country/round combinations where the number of noninvariant parameters exceeds 25%. Consequently, their means must be analyzed with caution

Table 7 Importance of self-enhancement values: Country rankings across all European Social Survey (ESS) (The value means estimated by the alignment optimization are in parentheses). (Author's own work)

Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7
SI[a] (−0.101)	HU[a] (−0.114)	SI[a] (−0.025)	SI[a] (0.085)	SI[a] (0.257)	HU[a] (0.298)	SI[a] (0.337)
HU[a] (−0.163)	PL[a] (−0.186)	PL[a] (−0.177)	HU[a] (−0.113)	HU[a] (0.045)	SI[a] (0.286)	HU[a] (0.204)
PL[a] (−0.321)	SI[a] (−0.209)	HU[a] (−0.214)	PL[a] (−0.202)	PL[a] (−0.067)	PL[a] (0)	PL[a] (−0.191)
ES (−0.344)	PT[a] (−0.226)	PT[a] (−0.342)	BE (−0.41)	IE[a] (−0.202)	IE[a] (−0.149)	IE[a] (−0.218)
PT[a] (−0.454)	BE (−0.562)	BE (−0.469)	PT[a] (−0.42)	PT[a] (−0.238)	CH[a] (−0.17)	CH[a] (−0.273)
IE[a] (−0.503)	IE[a] (−0.562)	CH[a] (−0.548)	IE[a] (−0.425)	CH[a] (−0.345)	PT[a] (−0.22)	BE (−0.461)
BE (−0.603)	DE[a] (−0.604)	IE[a] (−0.563)	CH[a] (−0.449)	BE (−0.538)	BE (−0.379)	PT[a] (−0.503)
DE[a] (−0.621)	ES (−0.609)	DE[a] (−0.581)	DE[a] (−0.63)	GB (−0.553)	GB (−0.48)	DK[a] (−0.526)
GB (−0.669)	GB (−0.621)	NL[a] (−0.655)	NL[a] (−0.646)	DE[a] (−0.625)	DK[a] (−0.517)	NL[a] (−0.684)
CH[a] (−0.674)	CH[a] (−0.667)	GB (−0.658)	DK[a] (−0.697)	NL[a] (−0.63)	NL[a] (−0.557)	GB (−0.784)
DK[a] (−0.741)	DK[a] (−0.734)	ES (−0.716)	GB (−0.739)	DK[a] (−0.684)	DE[a] (−0.589)	DE[a] (−0.852)
NL[a] (−0.763)	NL[a] (−0.79)	DK[a] (−0.722)	NO[a] (−0.78)	ES (−0.801)	ES (−0.777)	NO[a] (−0.857)
NO[a] (−0.941)	NO[a] (−0.838)	NO[a] (−0.854)	ES (−0.787)	NO[a] (−0.887)	NO[a] (−0.777)	ES (−0.904)
SE (−1.012)	SE (−0.991)	SE (−1.022)	SE (−0.933)	FI[a] (−0.978)	SE (−0.81)	SE (−1.019)
FI[a] (−1.081)	FI[a] (−1.094)	FI[a] (−1.137)	FI[a] (−1.18)	SE (−1.085)	FI[a] (−1.152)	FI[a] (−1.216)

Country abbreviations listed in Table 2

[a]Country/round combinations where the number of noninvariant parameters exceeds 25%. Consequently, their means must be analyzed with caution

Table 8 Importance of openness to change values (without hedonism): Country rankings across all European Social Survey (ESS) rounds (The value means estimated by the alignment optimization are in parentheses). (Author's own work)

Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7
CH[a] (−0.33)	HU[a] (−0.364)	SI[a] (−0.337)	SI[a] (−0.333)	SI[a] (−0.147)	SI[a] (−0.189)	SI[a] (0)
SI[a] (−0.416)	CH[a] (−0.391)	NL[a] (−0.43)	IE[a] (−0.357)	CH[a] (−0.354)	CH[a] (−0.231)	CH[a] (−0.236)
DK[a] (−0.424)	SI[a] (−0.517)	CH[a] (−0.5)	CH[a] (−0.425)	HU[a] (−0.365)	NL[a] (−0.342)	DK[a] (−0.27)
HU[a] (−0.5)	NL[a] (−0.542)	HU[a] (−0.507)	NL[a] (−0.444)	NL[a] (−0.369)	HU[a] (−0.436)	DE[a] (−0.502)
ES[a] (−0.527)	BE (−0.613)	IE[a] (−0.643)	DK[a] (−0.567)	DK[a] (−0.488)	ES[a] (−0.452)	NL[a] (−0.505)
BE (−0.562)	DE[a] (−0.683)	DK[a] (−0.655)	HU[a] (−0.594)	DE[a] (−0.525)	IE[a] (−0.471)	HU[a] (−0.535)
NL[a] (−0.566)	FI (−0.683)	GB (−0.659)	BE (−0.648)	FI (−0.527)	DE[a] (−0.483)	FI (−0.561)
GB (−0.591)	DK[a] (−0.684)	DE[a] (−0.671)	GB (−0.659)	ES[a] (−0.533)	BE (−0.493)	BE (−0.576)
FI (−0.598)	IE[a] (−0.72)	BE (−0.674)	DE[a] (−0.662)	IE[a] (−0.56)	DK[a] (−0.499)	IE[a] (−0.59)
IE[a] (−0.62)	GB (−0.725)	ES[a] (−0.688)	ES[a] (−0.674)	GB (−0.662)	SE[a] (−0.52)	SE[a] (−0.6)
DE[a] (−0.633)	ES[a] (−0.772)	FI (−0.709)	FI (−0.71)	BE (−0.691)	GB (−0.534)	ES[a] (−0.628)
PL[a] (−0.752)	PL[a] (−0.774)	PL[a] (−0.778)	SE[a] (−0.791)	PL[a] (−0.754)	FI (−0.556)	GB (−0.678)
SE[a] (−1.017)	SE[a] (−0.961)	SE[a] (−0.94)	PL[a] (−0.804)	SE[a] (−0.82)	PL[a] (−0.667)	NO[a] (−0.753)
NO[a] (−1.021)	NO[a] (−0.996)	NO[a] (−0.991)	NO[a] (−0.879)	NO[a] (−0.872)	NO[a] (−0.73)	PL[a] (−0.84)
PT[a] (−1.027)	PT[a] (−1.511)	PT[a] (−1.198)	PT[a] (−1.23)	PT[a] (−1.011)	PT[a] (−1.016)	PT[a] (−0.921)

Country abbreviations listed in Table 2

[a]Country/round combinations where the number of noninvariant parameters exceeds 25%. Consequently, their means must be analyzed with caution

References

Aleman, Jose, and Dwayne Woods. 2016. Value orientations from the World Value Survey: How comparable are they cross-nationally? *Comparative Political Studies* 49:1039–1067. https://doi.org/10.1177/0010414015600458

Ariely, Gal, and Eldad Davidov. 2010. Can we rate public support for democracy in a comparable way? Cross-national equivalence of democratic attitudes in the World Value Survey. *Social Indicators Research* 104:271–286.

Asparouhov, Tihomir, and Bengt O. Muthén. 2014. Multi-group factor analysis Alignment. *Structural Equation Modeling* 21:1–14. https://doi.org/10.1080/10705511.2014.919210

Beierlein, Constanze, Eldad Davidov, Peter Schmidt, Shalom H. Schwartz and Beatrice Rammstedt. 2012. Testing the discriminant validity of Schwartz' Portrait Value Questionnaire items—A replication and extension of Knoppen and Saris (2009). *Survey Research Methods* 6:25–36.

Bilsky, Wolfgang, Michael Janik and Shalom H. Schwartz. 2011. The structural organization of human values—Evidence from three rounds of the European Social Survey (ESS). *Journal of Cross-Cultural Psychology* 42:759–776. https://doi.org/10.1177/0022022110362757

Brown, Timothy A. 2015. *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Byrne, Barbara M., Richard J. Shavelson and Bengt O. Muthén. 1989. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin* 105:456–466. https://doi.org/10.1207/s15328007sem1102_8

Chen, Fang F. 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling* 14:464–504. https://doi.org/10.1080/10705510701301834

Chen, Fang F. 2008. What happens if we compare chopsticks with forks? The impact of making inappropriate comparison in cross-cultural research. *Journal of Personality and Social Psychology* 95:1005–1018. https://doi.org/10.1037/a0013193

Cieciuch, Jan, and Eldad Davidov. 2012. A comparison of the invariance properties of the PVQ-40 and the PVQ-21 to measure human values across German and Polish samples. *Survey Research Methods* 6:37–48.

Cieciuch, Jan, and Eldad Davidov. 2015. Establishing measurement invariance across online and offline samples. A tutorial with the software packages Amos and Mplus. *Studia Psychologica* 15: 83–99. https://doi.org/10.21697/sp.2015.14.2.06

Cieciuch, Jan, and Shalom H. Schwartz. 2012. The number of distinct basic values and their structure assessed by PVQ-40. *Journal of Personality Assessment* 94:321–328. https://doi.org/10.1080/00223891.2012.655817

Cieciuch, Jan, Eldad Davidov and René Algesheimer. 2016. The stability and change of value structure and priorities in childhood: A longitudinal study. *Social Development* 25:503–527. https://doi.org/10.1111/sode.12147

Cieciuch, Jan, Eldad Davidov, Peter Schmidt, René Algesheimer and Shalom H. Schwartz. 2014. Comparing results of an exact versus an approximate (Bayesian) measurement invariance test: A cross-country illustration with a new scale to measure 19 human values. *Frontiers in Psychology* 982:1–10. https://doi.org/10.3389/fpsyg.2014.00982

Cieciuch, Jan, Eldad Davidov and Peter Schmidt. 2018. Using alignment optimization in establishing measurement invariance. In *Cross-cultural analysis: Methods and applications, 2nd edition*, eds. Eldad Davidov, Peter Schmidt, Jaak Billiet and Bart Meuleman. New York: Routledge Taylor & Francis Group.

Cieciuch, Jan, Eldad Davidov, Michele Vecchione, Constanze Beierlein and Shalom H. Schwartz. 2014a. The cross-national invariance properties of a new scale to measure 19 basic human values. A test across eight countries. *Journal of Cross-Cultural Psychology* 45:764–779. https://doi.org/10.1177/0022022114527348

Cieciuch, Jan, Eldad Davidov, Michele Vecchione and Shalom H. Schwartz. 2014b. A hierarchical structure of basic human values in a third-order confirmatory factor analysis. *Swiss Journal of Psychology* 73:177–182. https://doi.org/10.1024/1421-0185/a000134

Cieciuch, Jan, Eldad Davidov, René Algesheimer and Peter Schmidt. 2017. Testing for approximate measurement invariance of human values in the European Social Survey. *Sociological Methods & Research*. https://doi.org/10.1177/0049124117701478

Cieciuch, Jan, Shalom H. Schwartz and Eldad Davidov. 2015. Values, social psychology of. In *International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, v. 25*, ed. James D. Wright, 41–46. Oxford: Elsevier.

Cieciuch, Jan, Shalom H. Schwartz and Michele Vecchione. 2013. Applying the refined values theory to past data: What can researchers gain? *Journal of Cross-Cultural Psychology* 44:1215–1234. https://doi.org/10.1177/0022022113487076

Coromina, Lluis, and Eldad Davidov. 2013. Evaluating measurement invariance for social and political trust in Western Europe over four measurement time points (2002–2008). *ASK Research & Methods* 22:35–52.

Davidov, Eldad. 2008. A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods* 2:33–46.

Davidov, Eldad. 2009. Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective. *Political Analysis* 17:64–82.

Davidov, Eldad. 2010. Testing for comparability of human values across countries and time with the third round of the European Social Survey. *International Journal of Comparative Sociology* 51:171–191. https://doi.org/10.1177/0020715210363534

Davidov, Eldad, and Pascal Siegers. 2010. Comparing basic human values in East and West Germany. In *Komparative empirische Sozialforschung* (Comparative empirical social research)*,* eds. Tilo Beckers, Klaus Birkelbach, Jörg Hagenah, and Ulrich Rosar, 43–63. Wiesbaden: VS.

Davidov, Eldad, Jan Cieciuch, Peter Schmidt, Bart Meuleman and René Algesheimer. 2015. The comparability of measurements of attitudes toward immigration in the European Social Survey: Exact versus approximate measurement equivalence. *Public Opinion Quarterly* 79: 244–266. https://doi.org/10.1093/poq/nfv008

Davidov, Eldad, Hermann Dülmer, Jan Cieciuch, Anabel Kuntz, Daniel Seddig and Peter Schmidt 2016. Explaining measurement nonequivalence using multilevel structural equation modeling: The case of attitudes toward citizenship rights. *Sociological Methods & Research.* https://doi.org/10.1177/0049124116672678

Davidov, Eldad, Hermann Dülmer, Elmar Schlueter, Peter Schmidt and Bart Meuleman. 2012. Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology* 43:558–575. https://doi.org/10.1177/0022022112438397

Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt and Jaak Billiet. 2014. Measurement equivalence in cross-national research. *Annual Review of Sociology* 40:55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Davidov, Eldad, Peter Schmidt and Shalom H. Schwartz. 2008. Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly* 72:420–445. https://doi.org/10.1093/poq/nfn035

De Beuckelaer, Alain, and Gilbert Swinnen. 2018. Biased latent variable mean comparisons due to measurement noninvariance: A simulation study. In *Cross-cultural research: Methods and applications*, *2nd edition,* eds. Eldad Davidov, Peter Schmidt, Jaak Billiet and Bart Meuleman, 127–156. New York: Routledge Taylor & Francis Group.

Döring, Anna, Shalom H. Schwartz, Jan Cieciuch, Patrick J. F. Groenen, Valentina Glatzel, Justyna Harasimczuk, Nicole Janowicz, Maya Nyagolova, Rebecca E. Scheefer, Matthias Allritz, Taciano L. Milfont and Wolfgang Bilsky. 2015. Cross-cultural evidence of value structures and priorities in childhood. *British Journal of Psychology* 106:675–699. https://doi.org/10.1111/bjop.12116

Durkheim, Émile. 1897/1964. *Suicide*. Glencoe, IL: Free Press.

Goerres, Achim, Markus B. Siewert and Claudius Wagemann. 2019. Internationally comparative research designs in the social sciences: Fundamental issues, case selection logics, and research limitations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00600-2.

Guenole, Nigel. 2016. The importance of isomorphism for conclusions about homology: A Bayesian multilevel structural equation modeling approach with ordinal indicators. *Frontiers in Psychology* 7:289. https://doi.org/10.3389/fpsyg.2016.00289

Hitlin, Steven, and Allyn Piliavin. 2004. Values: Reviving a dormant concept. *Annual Review of Sociology* 30:359–393. https://doi.org/10.1146/annurev.soc.30.012703.110640

Hofstede, Geert. 2000. *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations, 2nd edition.* Beverly Hills, CA: Sage.

Horn, John L., and John J. McArdle. 1992. A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* 18:117–144. https://doi.org/10.1080/03610739208253916

Inglehart, Ronald, and Wayne E. Baker. 2000. Modernization, cultural change, and the persistence of traditional values. *American Sociological Review* 65:19–51.

Jak, Suzanne, Frans J. Oort and Conor V. Dolan. 2013. A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling* 20:265–282. https://doi.org/10.1080/10705511.2013.769392

Jöreskog, Karl G. 1971. Simultaneous factor analysis in several populations. *Psychometrika* 36:409–426. https://doi.org/10.1007/BF02291366

Kluckhohn, Clyde. 1951. Values and value-orientations in the theory of action: An exploration in definition and classification. In *Toward a general theory of action*, eds. Talcott Parsons and Edward A. Shils, 388–433. Cambridge, MA: Harvard University Press.

Knoppen, Desirée, and Willem Saris. 2009. Do we have combine values in the Schwartz' human values scale? A comment on the Davidov studies. *Survey Research Methods* 3:91–103.

Lomazzi, Vera. 2018. Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, data, analyses: A journal for quantitative methods and survey methodology (mda)* 12:77–103. https://doi.org/10.12758/mda.2017.09

Magun, Vladimir, Maxim Rudnev and Peter Schmidt. 2016. Within- and between-country value diversity in Europe: A typological approach. *European Sociological Review* 32:189–202. https://doi.org/10.1093/esr/jcv080

Marsh, Herbert W., Jiesi Guo, Philip D. Parker, Benjamin Nagengast, Tihomir Asparouhov, Bengt O. Muthén and Theresa Dicke. 2017. What to do when scalar invariance fails: The extended alignment method to multi-group factor analysis comparison of latent means across many groups. *Structural Equation Modeling*. https://doi.org/10.1037/met0000113

Meitinger, Katharina. 2017. Necessary but insufficient: Why measurement invariance tests need online probing as a complementary tool. *Public Opinion Quarterly* 81:447–472. https://doi.org/10.1093/poq/nfx009.

Merkle, Edgar C., and Yves Rosseel. 2016. blavaan: Bayesian structural equation modelling via parameter expansion. *arXiv: 1511.05604v2 [stat.CO]*. Retrieved from https://arxiv.org/abs/1511.05604 on June 4, 2018.

Meuleman, Bart. 2019. Multilevel structural equation modeling for cross-national comparative research. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00605-x.

Millsap, Roger E. 2011. *Statistical approaches to measurement invariance*. New York: Routledge.

Munck, Ingrid, Carolyn Barner and Judith Torney-Purta. 2017. Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009. The alignment method applied to IEA CIVED and ICCS. *Sociological Methods & Research*. https://doi.org/10.1177/0049124117729691

Muthén Bengt O. 1994. Multilevel covariance structure analysis. *Sociological Methods & Research* 22:376–398. https://doi.org/10.1177/0049124194022003006

Muthén, Bengt O., and Tihomir Asparouhov. 2013. BSEM measurement invariance analysis. *Mplus Web Notes* 17:1–48.

Muthén, Bengt O., and Tihomir Asparouhov. 2014. IRT studies of many groups: The alignment method. *Frontiers in Psychology* 978:1–7. https://doi.org/10.3389/fpsyg.2014.00978

Muthén, Bengt O., and Tihomir Asparouhov. 2017. Recent methods for the study of measurement invariance with many groups. Alignment and random effects. *Sociological Methods & Research*. https://doi.org/10.1177/0049124117701488

Muthén, Linda K., and Bengt O. Muthén. 1998–2014. *Mplus user's guide.* Los Angeles: Muthén & Muthén.

Oberski, Daniel L. 2014. Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis* 22:45–60. https://doi.org/10.1093/pan/mpt014

Rokeach, Milton. 1973. *The nature of human values*. New York, NY: Free Press.

Rudnev, Maksim, Ekaterina Lytkina, Eldad Davidov, Peter Schmidt and Andreas Zick. 2018a. Testing measurement invariance for a second-order factor: A cross-national test of the alienation scale. *Methods, data, analyses: A journal for quantitative methods and survey methodology (mda)* 12:47–76. https://doi.org/10.12758/mda.2017.11

Rudnev, Maxim, Vladimir Magun and Shalom Schwartz. 2018b. Relations among higher order values around the world. *Journal of Cross-Cultural Psychology* 49(8):1165–1182. https://doi.org/10.1177/0022022118782644

Ruelens, Anna, Bart Meuleman and Ides Nicaise. 2016. Examining measurement isomorphism of multilevel constructs: The case of political trust. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1799-6

Schafer, Joseph L., and John W. Graham. 2002. Missing values: Our view of the state of the art. *Psychological Methods* 7:147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schwartz, Shalom H. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, *vol. 25*, ed. Mark Zanna, 1–65. London, UK: Academic Press.

Schwartz, Shalom H. 2003. A proposal for measuring value orientations across nations. In *Questionnaire development package of the European Social Survey*, 259–319. Retrieved from www.europeansocialsurvey.org, June 30, 2016.

Schwartz, Shalom H., and Jan Cieciuch. 2016. Values. In *The ITC international handbook of testing and assessment*, eds. Frederick T. L. Leong, Dave Bartram, Fanny M. Cheung, Kurt F. Geisinger and Dragos Iliescu, 106–119. Oxford: Oxford University Press.

Schwartz, Shalom H., Jan Cieciuch, Michelle Vecchione, Eldad Davidov, Ronald Fischer, Constanze Beierlein, Alice Ramos, Markku Verkasalo, Jan-Erik Lönnqvist, Kursad Demirutku, Ozlem Dirilen-Gumus and Mark Konty. 2012. Refining the theory of basic individual values. *Journal of Personality and Social Psychology* 103:663–688. https://doi.org/10.1037/a0029393

Schwartz, Shalom H., Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris and Vicki Owens. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology* 32:519–542.

Sokolov, Boris. 2018. The index of emancipative values: Measurement model Misspecifications. *American Political Science Review* 112:395–408.

Steenkamp, Jan-Benedict E. M., and Hans Baumgartner. 1998. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research* 25:78–90. https://doi.org/10.1086/209528

Steinmetz, Holger. 2018. Estimation and comparison of latent means across cultures. In *Cross-cultural analysis: Methods and applications*, 2nd edition, eds. Eldad Davidov, Peter Schmidt, Jaak Billiet and Bart Meuleman, 95–126. New York: Routledge Taylor & Francis Group.

Steinmetz, Holger Rodrigo Isidor, Naissa Baeuerle. 2012. Testing the circular structure of human values: A meta-analytical structural equation modelling approach. *Survey Research Methods* 6:61–75

van de Schoot, Rens, Anouck Kluytmans, Lars Tummers, Peter Lugtig, Joop Hox and Bengt O. Muthén. 2013. Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology* 770:1–15. https://doi.org/10.3389/fpsyg.2013.00770

Vandenberg, Robert J., and Charles E. Lance. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods* 3:4–69. https://doi.org/10.1177/109442810031002

Weber, Max. 1905/1958. *The protestant ethic and the spirit of capitalism*. New York: Scribner's.

Welzel, Christian, and Ronald F. Inglehart. 2016. Misconceptions of measurement equivalence: Time for a paradigm shift. *Comparative Political Studies* 49:1068–1094.

Zercher, Florian, Peter Schmidt, Jan Cieciuch and Eldad Davidov. 2015. The comparability of the universalism value over time and across countries in the European Social Survey: Exact versus approximate measurement invariance. *Frontiers in Psychology* 733:1–11. https://doi.org/10.3389/fpsyg.2015.00733

Jan Cieciuch 1974, Dr., University of Zurich and Cardinal Stefan Wyszynski University in Warsaw. Areas of research: structure and development of personality with focus on values, traits, personal identity, and well-being, also from a cross-cultural perspective.

Eldad Davidov 1971, Prof. Dr., University of Cologne and University of Zurich. President of the European Survey Research Association (ESRA) between 2015 and 2017. Areas of research: structural equation modeling especially applied to cross-cultural and longitudinal survey data. In his research he analyzes human values and attitudes toward immigrants or other minorities.

Peter Schmidt 1942, Prof. em. Dr., University of Giessen. Areas of research: foundations and applications of generalized latent variable models, and especially structural equation models. Applications include cross-country, repeated cross-sections and panel data. The substantive topics deal with values, the reasoned action approach, attitudes toward minorities, national identity and innovation. Together with A. Heath, E. Green, E. Davidov and A. Ramos, he was a member of the Question Design Team for the immigration module of the ESS 2014.

René Algesheimer 1973, Prof. Dr., University of Zurich. Director of the University Research Priority Program on Social Networks. Areas of research: how social structures shape individuals' behavior, and correspondingly how individuals' behavior reproduces social structures. More recently he has worked on the effects of social networks on influence processes, opinion diffusion, risk preferences, social learning, human values, and cooperation.

How Much Do Sources of Happiness Vary Across Countries? A Review of the Empirical Literature

Olga Stavrova

Abstract This article presents a review of empirical research exploring cross-national differences in the correlates of subjective well-being (SWB). I start by giving an overview of the concept of SWB across psychological, sociological, and economic literature. Measures of SWB have good cross-cultural validity, yet there is currently little consensus regarding the cultural universality of the definition of happiness. An overview of existing empirical literature points toward robust cross-national differences in mean levels of SWB that are associated with national differences in wealth and other socioeconomic, political, and cultural factors. The degree to which individual-level variables are associated with SWB is also subject to cross-national variations. Many individuals' characteristics contribute to happiness to the extent that they are beneficial, socially desirable, and aspired to in a particular sociocultural context. These results are discussed in light of two theoretical approaches (institutional and fit hypotheses). Directions for future research are proposed.

Keywords Life satisfaction · Cross-national comparitive research · Subjective well-being · Culture · Person-culture fit

Internationale Unterschiede in den Einflussfaktoren auf das Glück: Übersicht über die empirische Literatur

Zusammenfassung Dieser Artikel stellt eine Übersicht der empirischen Forschung über internationale Unterschiede in den Korrelaten des subjektiven Wohlbefindens (SWB) dar. Zuerst wird eine Übersicht der Konzeptualisierung und Operationalisie-

O. Stavrova (✉)
Department of Social Psychology, Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands
E-Mail: O.Stavrova@uvt.nl

 Springer

rung des Konzepts SWB in der psychologischen, soziologischen und ökonomischen Literatur gegeben. Obwohl es in der Glücksforschung keine Übereinstimmung über die interkulturelle Universalität der Definition von Glück gibt, ergaben die Messinstrumente von SWB eine gute interkulturelle Validität. Ein Überblick empirischer Studien zeigt robuste internationale Unterschiede in den Durchschnittswerten von SWB, die mit nationalen Unterschieden im Wohlstand und anderen sozioökonomischen, politischen und kulturellen Faktoren zusammenhängen. Der Zusammenhang zwischen Glück und individuellen Merkmalen unterliegt ebenfalls internationalen Unterschieden. Individuelle Faktoren tragen zum höheren SWB insofern bei, als sie in einem bestimmten soziokulturellen Kontext sozial erwünscht und angestrebt sind. Die Ergebnisse werden im Licht von 2 theoretischen Ansätzen (institutionelle Hypothese und Person-Umwelt-Passungstheorie) diskutiert. Es werden Vorschläge für zukünftige Forschung gemacht.

1 Introduction

The question of the nature and causes of human happiness has been a source of major interest in philosophy since antiquity, with different schools of thought offering different answers (McMahon 2006). In contrast to philosophy, it is only since the middle of the twentieth century that social scientists have discovered this research field. Since then, research on happiness, or subjective well-being (SWB), as this concept is often referred to in the social sciences, has been very fruitful, and attracted scholars from diverse disciplines, including psychology, sociology, and economics, who have published more than 80,000 articles on that topic in total over the last 30 years (according to the Web of Science).

Most of this work has been dedicated to revealing factors that contribute to a higher vs. a lower SWB. Some studies have focused on individual characteristics, such as personality traits, values, and beliefs or life circumstances. Others explored contextual or country-level predictors, such as cross-national differences in social policies, socioeconomic conditions and culture. Finally, a third group of studies combined the investigation of individual- and country-level factors exploring how they come together to affect individuals' happiness. Studies exploring individual-level predictors probably represent the largest portion of the literature and have been the target of several review papers in psychology (Diener et al. 1999; Lyubomirsky et al., 2005a, b). Studies exploring national-level predictors have been the major focus of sociological and economic literature on SWB and have also been reviewed before (Di Tella et al. 2003; Frey and Stutzer 2005). Yet, there have been no reviews of the third stream of empirical research, the one that combines the examination of individual- and country-level predictors of SWB. Therefore, the present review focuses on this third group of studies. Nevertheless, to put these studies into context, I also include an overview of empirical studies that examined individual- and country-level predictors independently of each other. I will start by discussing existing

conceptualizations, definitions, and measurements of the concept of SWB, including the question of its cross-cultural validity (see also Cieciuch et al. 2019). I will then proceed to present an overview of empirical findings pertaining to individual- and country-level predictors, as well as their joint effects on SWB. Finally, I will discuss the results in the light of existing theories and outline directions for future research.

2 The Concept of Subjective Well-being

Probably due to the interdisciplinary nature of SWB research (it is an active area of research across diverse disciplines, mainly including psychology, sociology, and economics), there are currently dozens of concepts that are studied under the umbrella term of subjective well-being: life satisfaction, avowed happiness (Wilson 1967), subjective happiness (Lyubomirsky and Lepper 1999), authentic happiness (Seligman 2002), affect balance (Bradburn 1969), experienced utility, and objective happiness (Kahneman 2000), to name but a few.

Even though research on SWB is largely interdisciplinary in nature, it has long been dominated by psychologists. This is not surprising, as SWB is an individual-level concept that deals with differences in individuals' subjective perception and evaluation of reality; and the way people see reality is the core business of psychology, a science of the human mind. In contrast, despite the rising level of interest in the concept of happiness in sociology and economics, it is still rarely discussed in the respective textbooks and journals (Veenhoven 2008). This relative lack of interest has been explained by these disciplines' main interest in understanding societal problems rooted in objective reality or conditions, rather than understanding how people feel about reality and about their conditions, and why (Veenhoven 2008). Accordingly, the most elaborate conceptualization of happiness stems from psychological literature and will be the focus in the present review. At the same time, as I explore an interplay of individual and contextual (country-level) predictors of happiness, and the latter have been almost exclusively the focus of sociology and economics (with the exception of cross-cultural research), I will provide an overview of how happiness is conceptualized in these disciplines as well.

2.1 The Concept of SWB in Psychology

The most widely accepted conceptualization of SWB in psychology was proposed in the 1980s by Ed Diener, who also coined this term. SWB was defined as "a general area of scientific interest, rather than a single specific construct" (Diener et al. 1999, p. 277), with life satisfaction and affect representing the most widely studied concepts within the field of subjective well-being. Life satisfaction represents a cognitive aspect of SWB; it refers to individuals' evaluation of how close their life is to what constitutes an ideal life in their view. Affect represents the emotional component of SWB. It is typically operationalized as frequency and intensity of positive and negative emotions experienced during the last several weeks (Diener 1984; Kahneman and Deaton 2010).

Both the cognitive and affective components of SWB are subjective and depend on individuals' personal evaluation and judgment. Also, they are both hedonic, rather than eudaimonic, in nature. Current psychological literature on well-being distinguishes between its hedonic and eudaimonic components (Ryan and Deci 2001). Eudaimonic well-being is grounded in the philosophy of Aristotle, who suggested that a life of virtue and realization of one's potential constitutes true happiness (eudaimonia). The Aristotelian conception of happiness has been further developed by humanistic and positive psychologists (Ryff 1989; Seligman 2002). Instruments measuring eudaimonic well-being typically include such concepts as meaning and purpose, personal growth and development, rather than mere satisfaction. For example, Ryff (1989) considers happiness (psychological well-being) as consisting of six dimensions: autonomy, personal growth, mastery, positive relatedness, life purpose, and self-acceptance.

In contrast, hedonic—both cognitive and affective—well-being emphasizes individuals' global evaluation and feeling regarding their lives. SWB is retrospective in nature, meaning that measures of life satisfaction or affect represent memory-based reports. Although such retrospective reports of happiness do not completely overlap with moment-to-moment happiness (e.g., measures of happiness collected during different activities and times of the day) (Fredrickson 2000; Redelmeier and Kahneman 1996), they are important as they often represent the basis of individuals' decision making. For example, it is the retrospective, stored-in-memory judgment of happiness one experienced in a particular restaurant that is likely to affect one's decision to visit this restaurant again (Stavrova 2014).

2.2 The Concept of SWB in Sociology and Economics

Sociological research has started showing an interest in the studies of happiness since the 1980s. Similar to psychology, a variety of terms referring to happiness have been used by sociologists as well. While some of these terms overlap with the ones used by psychologists, others are new and denote the particularities of the use of the concept in sociology in general. For example, while the term "happiness" and "subjective well-being" are used by both psychologists and sociologists as an umbrella term, sociologists see it as a synonym of "quality of life" and "individual and social welfare" (Veenhoven 2012). Regardless of these differences, the definition of happiness adopted in sociology is quite similar to the psychological one: happiness is defined in sociology as "the degree to which an individual judges the overall quality of his/her own life-as a whole favorably" (Veenhoven 1984). Yet, in contrast to psychologists, who differentiate between the cognitive and the affective components, sociologists see life satisfaction and happiness as a combination of a "cognitive comparison with standards of the good life (contentment) and affective information from how one feels most of the time (hedonic level of affect)" (Veenhoven 2008). In a tradition of social constructionist theories (Berger and Luckman 1966), sociologists share a constructionist view of happiness, assuming that happiness represents individuals' construction of reality, which is largely determined by comparative thinking: people are happy as long as their life meets their expectations and is not worse than a neighbor's life (Veenhoven 2008). As the main measurement

instrument, sociological research typically uses a single happiness item that directly asks individuals how happy they are with their lives as a whole.

Interestingly, this very measure of happiness is also quite often used in economics (Frey and Stutzer 2002a; 2005). The concept itself is however considered in a quite different light. In contrast to the social constructionist view shared by sociologists (and psychologists, although they would not call it that), economists use happiness as a tool to measure the desirability of objective reality, or in other words, for economists, happiness represents a measure of utility (Frey and Stutzer 2014). Economists' interest in happiness is just a result of their attempts to find the best operationalization of the utility that individuals derive from using different products, institutional or social policy changes (Di Tella et al. 2003; Di Tella and MacCulloch 2006; Frey and Stutzer 2002a, b).

2.3 SWB Measurement

A variety of reliable and valid instruments have been designed to assess SWB and its different components. One of the most widely known instruments used to measure life satisfaction in psychology and sociology is the Satisfaction With Life Scale (Diener et al. 1985). It consists of five items and has a high degree of reliability and external validity (Pavot and Diener 1993). It shows a high level of self-informant consistency (Schneider and Schimmack 2010) and converges with objective measures of satisfaction (e.g. smiling behavior, use of negative emotion words in texts) (Liu et al. 2015; Settanni and Marengo 2015). It has also been shown to have high temporal stability (0.54 within 4 years; Diener et al. 2013), showing that—in contrast to what some studies have suggested (Schwarz 1990, but see Yap et al. 2016)—satisfaction with life is not a momentary assessment that is potentially subject to framing, item order, or weather effects (Diener et al. 2013; Pavot and Diener 2008).

Most large-scale cross-national studies typically rely on a single-item scale of life satisfaction, asking participants to indicate how satisfied they are with their lives overall (a 10-point Likert scale is usually used). This single-item measure of life satisfaction has been shown to strongly correlate with the Satisfaction With Life Scale and to have good external validity as well (Cheung and Lucas 2014; Jovanović 2016). Due to its brief form and ease of use, it has been included in many large-scale cross-national surveys. As a result, most cross-national findings on life satisfaction so far are based on the data using this measure.

Regarding the affective component of SWB, a number of measures have been developed as well, with Bradburn's Affect Balance scale (Bradburn 1969) and Positive and Negative Affect Schedule (PANAS; Watson et al. 1988) being the scales used most often. These measures assess the frequency of experiencing a series of positive and negative emotions over the past 4 weeks.

Cognitive (life satisfaction) and affective measures of SWB show moderate to strong correlations with each other (Diener et al. 1999), but do demonstrate distinct patterns of correlations with other constructs. For example, while higher income is positively related to life satisfaction, it does not predict more positive emotions (Kahneman and Deaton 2010). On the other hand, fulfillment of psychological needs,

such as social affiliation, respect, and autonomy, is a stronger predictor of affect than it is of life satisfaction (Diener et al. 2010).

Importantly, existing SWB measures show a good level of cross-cultural validity. For example, a couple of studies showed that instruments developed to measure SWB in the West show acceptable levels of reliability, convergent and discriminant validity in other cultural contexts (Whisman and Judd 2016; see also Cieciuch et al. 2019). For example, the Satisfaction with Life Scale, which was developed for use in American samples, showed the same factor structure in samples of Serbian adolescents (Jovanović 2016), elderly Mexican people (López-Ortega et al. 2016), Korean elementary school children (Lim 2015), as well as adolescents and young adults in Portugal (Silva et al. 2015) and Italy (Di Fabio and Gori 2016). Also, reports of life satisfaction converge with more objective validity indicators, such as reports of positive vs. negative events, not only in American but in other cultures as well (Balatsky and Diener 1993).

Somewhat less consensus exists with respect to the extent of cross-cultural differences in folks theories of happiness, that is, lay beliefs about the nature and sources of happiness. On the one hand, a stream of literature in cross-cultural psychology tends to emphasize cross-cultural differences in lay (and even in dictionary, Oishi et al. 2013) definitions of happiness (Uchida and Kitayama 2009). These studies showed that North American vs. East Asian and East European participants think about happiness differently, with the former seeing it exclusively positively, and the later having a more ambivalent attitude towards happiness, including its negative aspects, such as the fact that it does not last long and may even cause envy on the part of others (Joshanloo et al. 2014). At the same time, cross-cultural commonalities in lay beliefs about happiness have been shown as well. For example, a recent exploration of lay definitions of happiness showed a certain degree of similarity, with an emphasis on inner harmony being considered central to happiness across the twelve countries studied, including the U.S., India, as well as some Eastern European and Latin American countries (Delle Fave et al. 2016).

While cognitive and affective components of SWB might be equally important, disciplines differ in what component they mostly focus on, with psychological research being interested in both, while studies conducted in sociology and economics are mostly restricted to life satisfaction or overall happiness (forgoing the distinction between cognitive and affective components altogether). As a result, most large-scale cross-national datasets include a measure of life satisfaction ("Taking all things together, how satisfied you are with your life these days?") and overall happiness ("Taking all things together, would you say you are very happy, quite happy, not very happy, not at all happy?"), but no measures of affective well-being, and most findings on cross-cultural variability in mean levels and correlates of SWB, are restricted to life satisfaction and overall happiness.

3 Individual Predictors of SWB

At the individual level, personality has been shown to be one of the strongest predictors of happiness (Lyubomirsky, Sheldon and Schkade 2005). In fact, differences in

stable individual dispositions are assumed to account for about half of the variance in SWB (Lyubomirsky et al., 2005b). Among the Big Five personality traits, neuroticism has been shown to be the strongest negative predictor of life satisfaction (Steel et al. 2008). Neurotic individuals are moody, experience frequent mood swings, get upset and stressed out easily, and respond to negative stimuli with greater anxiety. It comes as no surprise that neurotic individuals tend to report lower levels of life satisfaction and happiness.

Besides neuroticism, extraversion and conscientiousness typically show an independent positive association with SWB (Hayes and Joseph 2003; Lucas et al. 2000; Steel et al. 2008; Suldo et al. 2015). Extraverted individuals feel comfortable being around other people, start conversations easily, and like social attention; they are talkative and easy going. As a result, extraverted individuals are more likely to develop a sense of belonging, have more friends and acquaintances, are more likely to report high-quality relationships with others, less likely to feel lonely and socially excluded—all these attributes being important components of a happy life (Diener and Ryan 2009). In addition, social perception studies show that extraverted individuals are also perceived as being more likeable and preferred as communication partners and friends (Feiler and Kleinbaum 2015). What we see here is that higher levels of extraversion make one's social life easier and satisfy one's need to belong (Baumeister and Leary 1995), thus, representing an important factor contributing to happiness.

Conscientious individuals like order, pay attention to details, follow schedules and plans, are dutiful and responsible. Given that all these characteristics facilitate goal achievement and success, conscientious individuals report higher SWB than their less conscientious counterparts do (Hayes and Joseph 2003; Suldo et al. 2015). Other dispositional traits reflecting one's strength of will, such as the trait of self-control, as well as one's perception of control over one's life, show consistent positive associations with SWB as well (Hofmann et al. 2014; Lachman and Agrigoroaei 2010). Importantly, even though early research has almost unanimously assumed that differences in these basic personality traits result in different levels of SWB (Diener et al. 2003), more recent studies have shown that SWB can also affect personality development (Specht et al. 2013). Using the longitudinal data from Germany (German Socio-Economic Panel Study), Specht and colleagues (2013) showed that individuals with a higher (vs. lower) baseline level of life satisfaction were more likely to become more conscientious and less neurotic over time.

Finally, individuals' feelings and beliefs regarding the "self" represent another factor influencing well-being. Specifically, higher levels of self-esteem are positively associated with life satisfaction (Cheng and Furnham 2003). Individuals with high self-esteem are less likely to experience stressful life events (Orth and Luciano 2015) and are more likely to be liked and respected by others (Reitz et al. 2016).

Individual differences in goals, values, and beliefs are associated with SWB as well. The importance of goal attainment has been acknowledged in psychology since Maslow's pyramid of needs (Maslow 1943). Goals have also been seen as a crucial component of happiness in sociological theories. For example, the Social Production Function Theory (Ormel et al. 1999) postulates that people's SWB is contingent on

achieving five universal goals (stimulation, comfort, status, behavioral confirmation, and affection).

Conversely, another stream of literature suggested that it does not matter much whether or not goals are attained, but rather it is the type of goals people pursue that is crucial for SWB. For example, pursuing avoidance goals (e.g. not failing an exam) is associated with lower SWB, while endorsing approach goals (e.g. getting a great grade in an exam) is related to higher SWB (Elliot et al. 1997). Pursuing materialistic goals was shown to lead to a lower SWB (Kasser and Ryan 1993), and pursuing goals emanating from intrinsic rather than extrinsic motivations is related to an increased SWB (Sheldon and Kasser 1998). Research into values and SWB has also backed up the importance of the self-determination view (Ryan and Deci 2001). Specifically, it has been shown that self-direction, stimulation, and achievement values are positively related to positive emotions, whereas security, conformity, and tradition values are negatively associated with positive emotions (Sagiv and Schwartz 2000).

Finally, beliefs about the world have also been shown to affect happiness. For example, individuals who believe in a just world are more likely to report higher life satisfaction and more frequent positive affect than their counterparts with a lower level of belief in a just world (Correia et al. 2009; Lucas et al. 2013). Furthermore, conservative political ideology (Napier and Jost 2008), sexist ideology (Hammond and Sibley 2011), religious beliefs (Diener et al. 2011), and interpersonal trust (Tokuda et al. 2010) are also often cited as sources of life satisfaction. However, as most of these studies are based on correlational data, it remains unclear whether holding certain types of beliefs makes one happier and more satisfied, or whether happy and satisfied people are more likely to endorse certain types of beliefs.

Sociodemographic factors, life circumstances and events have been shown to explain individual differences in happiness as well. The findings indicate that having a job (Carroll 2007), earning good money (Luhmann et al. 2011), as well as being married or having a long-term partner (Kim and MacKenry 2002) are all positive predictors of SWB. In contrast, unemployment, divorce, widowhood, and poverty contribute to a lower SWB (Lucas et al. 2004; Williams and Dunne-Bryant 2006). However, these effects are probably bidirectional such that life satisfaction can "predispose" individuals to certain life events. For example, prospective studies have shown higher levels of baseline life satisfaction to be associated with higher chances of marriage and lower risks of divorce and unemployment (Luhmann et al. 2013). It should be noted that the explanatory power of sociodemographic factors is very small compared to that of the Big Five and other personality traits. According to Lyubomirsky et al. (2005b), while personality accounts for about 50% of variance in SWB, individual differences in sociodemographics only explain about 10%.

At a theoretical level, the relative importance of personality predictors provides support to the so-called top-down theories of happiness, according to which individuals' stable predispositions with a large genetic component shape individuals' happiness level. This view has also been expressed in the set-point theories of happiness (Headey and Wearing 1992). According to these theories, individuals have a genetically determined default level of happiness. Different life circumstances and behaviors may change it in the short run, reducing or increasing individuals'

happiness that will however ultimately return to its set point. These theories have been supported by a strong temporal stability that happiness measures are known to possess, as well as by surveys of twins. For example, one highly cited twin study supported the set-point assertion but showed that happiness levels in monozygotic twins correlate at 0.40, and in dizygotic twins at just 0.08 (Lykken and Tellegen 1996). Similarly, long-term panel studies have shown the test–retest correlation of life satisfaction to reach high values (around 0.50) across decades (Schimmack and Oishi 2005).

In contrast, the bottom-up theories assume that happiness is a result of being in a beneficial vs. threatening environment or experiencing positive as opposed to negative events. In other words, "a happy individual is happy precisely because he or she experiences many happy moments" (Brief et al. 1993). Consistent with this idea, large-scale panel studies have shown happiness to be subject to major life events (Lucas 2007), thereby questioning the postulates of the set-point theory. Not all events are capable of affecting individuals' default level of happiness, though: while periods of unemployment were shown to leave long-term "scarring" effects, the happiness brought about by getting married was short-lived (Lucas 2007; Lucas et al. 2004).

While many life events lie outside individuals' control and responsibility, there are ways in which, researchers believe, people can still shape their own happiness. Specifically, individuals can change their SWB through "effortful activities". Indeed, individual differences in everyday behavior are assumed to account for about 40% of variance in SWB (Lyubomirsky et al. 2005b). What are the activities that promote happiness?

One of these activities is prosocial behavior. The idea that prosociality is an important source of happiness harkens back to ancient philosophy, with Aristotle and Plato believing that a life of virtue is the only possible route to happiness. Recent psychological and sociological research has provided support for this idea from Antiquity, showing that altruistic behavior, volunteering and charity donations are associated with increased life satisfaction (Aknin et al. 2013; Dunn et al. 2008; Schwartz et al. 2003; Stavrova et al. 2013b). Importantly, the benefits of altruism were supported in experimental research as well. For example, Aknin et al. (2013) provided their participants with a small sum of money and instructed half of them to spend it on themselves and the other half to spend it on others. By the end of the day, the members of the latter group (prosocial spending) reported higher happiness levels than those in the former group (selfish spending).

Not only prosocial, but also merely social activities have been recognized as a source of SWB. Experience sampling studies (in these studies, participants are signaled with a smartphone application several times per day within a certain fixed time period, such as a week, and asked to fill in a small survey about their current experiences, thoughts, and feelings; this method makes it possible to study individuals' daily experiences without relying on their memory) have shown that individuals report the lowest positive affect when they are alone. In fact, people are happier when they are with their clients and bosses than just alone (Kahneman 2000). Different measures of social engagement, the presence of friends and a romantic partner were all associated with higher levels of SWB (Lucas and Dyrenforth 2006).

Finally, recent studies suggest that genetic and environmental factors (life circumstances, life events, the activities one chooses to engage in) are not independent of each other (Plomin 1994). For example, in a theory of person–situation transactions, individuals self-select into certain environments based on their personality predispositions (Caspi et al. 1989). To give an example, neurotic individuals are more likely to withdraw from social events, thus, further exacerbating their unhappiness. As a result, individuals' life outcomes (including happiness) represent a joint product of both genetic predispositions (e. g. personality) and environmental factors. This enables characteristics of one's national context to be recognized as being particularly important "environmental" factors. The following section explores existing findings on country-level predictors of SWB.

4 Country-level Predictors of SWB

Countries show substantial differences in average SWB levels. For example, analyses combining individual- and country-level predictors suggest that regional and national differences explain large amounts of variation in individuals' life satisfaction (Bonini 2008). Also, national SWB scores show a surprisingly high level of temporal stability. For example, in a 30-year period, the Japanese level of life satisfaction did not substantially deviate from a 6 on a 10-point scale, whereas that of the Danes fluctuated around an 8 (Veenhoven 1993). Which country-level characteristics explain differences in national levels of SWB?

Multiple studies converge on differences in national wealth as one of the most important country-level variables related to life satisfaction (Diener et al. 2003). Correlations between gross domestic product (GDP) and the national level of life satisfaction are typically in the range of 0.50–0.60, with individuals living in wealthier countries reporting higher life satisfaction compared to individuals in poorer ones (Di Tella et al. 2003). Interestingly, the associations between wealth and life satisfaction at the individual level are substantially smaller: individuals' wealth (e. g. personal income) correlates with their life satisfaction at approximately 0.10. One explanation of why wealth seems so important at the country level is that wealthy nations typically score higher on a wide range of other characteristics that have been shown to have beneficial consequences for their citizens' SWB as well: political freedom, civil rights, good governance, low crime rates and low social inequality (Dorn et al. 2007; Oishi et al. 2011; Ott 2011). Although there have been attempts to disentangle the effects of these different factors, they have been largely unsuccessful due to strong intercorrelations between the predictors and a small number of countries available for the analyses (Diener et al. 2003).

Besides socioeconomic and political differences, cultural dimensions have been shown to explain country-level differences in SWB as well. Among them, individualism vs. collectivism and uncertainty avoidance showed the most consistent associations with SWB. Individualism (vs. collectivism) represents one of the most widely studied and recognized dimensions of culture (Hofstede 2001; Markus and Kitayama 1991; Triandis 1995). Individualistic cultures are characterized by an emphasis on individuals' (vs. groups') needs and values, values of autonomy, and

independence, whereas collectivistic cultures are defined by prioritizing groups' values and well-being over those of individuals, security, and interdependence values and a stronger differentiation between in-groups and out-groups. People in collectivistic cultures tend toward an interdependent self-concept, defining themselves primarily as members of social groups. Individualism vs. collectivism, as well as many other dimensions of culture, have been quantified. That is, nearly a hundred national cultures have been assigned an individualism-collectivism score (as well as scores on other cultural dimensions such as uncertainty avoidance), which represents nationally aggregated individuals' responses to items yielding their preferred values, norms, attitudes, and behaviors. This has allowed researchers to explore the potential associations between cultural dimensions and SWB. As a result of this exploration, residents of individualistic countries were shown to score higher on life satisfaction and positive affect than residents of collectivistic countries did (Diener and Suh 2000; Hofstede 2001). Although this effect can be partly explained by the fact that individualistic cultures tend to be wealthier, it is consistent with different psychological perspectives highlighting the importance of freedom of choice and autonomy for well-being (e. g. the Self-Determination Theory; Deci and Ryan 2000; Ryan and Deci 2000).

The cultural dimension of uncertainty avoidance describes cultural differences in the tolerance of uncertainty and ambiguity (Hofstede 2001). While people in general show an aversion towards uncertainty, this is particularly so in cultures characterized by higher levels of uncertainty avoidance. Individuals in cultures with higher scores in uncertainty avoidance are particularly uncomfortable with uncertain and ambiguous situations, and rely on strict rules and regulations as a way of coping with uncertainty. In contrast, cultures with low scores on uncertainty avoidance show greater tolerance toward uncertainty; they perceive changeable environments and an uncertain future not as a threat but as an opportunity. Given that uncertainty is inherent to virtually any aspect of our lives, residents of cultures high in uncertainty avoidance tend toward lower SWB scores than cultures that are more tolerant of uncertainty (Hofstede 2001).

It is noteworthy that even though these studies examined the effect of culture as a macro-level characteristic, it is less clear whether the mechanism of these effects operates at a macro- or a micro- (i.e. individual) level. In fact, being part of an individualistic culture can strengthen an individual's SWB either because he/she is then more likely to hold an individualistic orientation him/herself or because living in an individualistic culture gives one more autonomy from others. Similarly, a country's level of uncertainty avoidance might be negatively associated with its citizens' happiness, either because the citizens are then more likely to be uncomfortable with uncertainty themselves or because being surrounded by individuals who are intolerant of uncertainty makes one's life stressful regardless of one's own uncertainty orientation. Examining the effects of these cultural orientations at both individual and country level would foster our understanding of these processes.

Even though no studies have undertaken this endeavor with respect to the dimensions of culture discussed above, several studies have attempted to differentiate between the effects of the same constructs measured at individual and country level in other areas. For example, it has been shown that more trusting individuals report

higher levels of life satisfaction than their less trusting counterparts, and that residents of more trusting cultures score higher on life satisfaction than residents of less trusting cultures do (Tokuda et al. 2010). Importantly, the effects at the individual and the country level were independent of each other: individuals living in countries with a high level of social trust are happier than individuals living in countries with a lower level, regardless of their personal trust score (Tokuda et al. 2010). Similarly, individual, national, and regional levels of neuroticism were shown to be robust negative predictors of SWB (Rentfrow et al. 2008; Steel and Ones 2002). For example, Steel and Ones (2002) demonstrated a strong negative association between neuroticism and life satisfaction aggregated at the country level. This negative relationship was replicated at the state level in the U.S., such that states with a high aggregated neuroticism score tend toward a lower SWB than states with a lower level (Rentfrow et al. 2009). Most recently, multilevel analyses of individual neuroticism and regional neuroticism in Germany showed that living in a state that was neurotic was negatively related to life satisfaction, regardless of individuals' own neuroticism score (Stavrova 2015a). Regardless of how neurotic a person is, being surrounded by neurotic individuals (that is, living in a region with a high level of neuroticism) has a detrimental effect on one's happiness.

At the same time, some individuals' characteristics seem to affect SWB only when aggregated at the macro-level. For example, a series of studies have shown that an average level of education in cities was positively related to a city-average SWB, whereas the association between education and happiness at the individual level was negligible (Florida et al. 2013). Similarly, national IQ was shown to positively predict the happiness of nations, whereas between-individual differences in intelligence are not a robust predictor of individuals' SWB (Veenhoven and Choi 2012). Taken together, these results suggest that some individual characteristics do not affect SWB directly, but only when aggregated at a higher level (city, region, country), probably via shaping societal living conditions and cultural climates.

Finally, according to a number of psychological perspectives, country-level differences in SWB might be explained by cultural differences in self-enhancement and the cultural importance of happiness. One of the most robust findings in cross-cultural psychology is the variation in self-serving and self-enhancement motives and tendencies, with individualistic cultures showing stronger self-enhancement than collectivistic ones (Heine and Hamamura 2007). Multiple psychological studies have shown that people in individualistic cultures tend to consider themselves as scoring higher on a range of positive traits and as having better future prospects than most other people (Dunning et al. 2004; Weinstein 1980). Such self-serving biases have been shown to be less common in collectivistic cultures, suggesting that cross-cultural differences in the mean level of SWB might be at least partly a result of a stronger self-enhancement tendency in individualistic (vs. collectivistic) nations.

Supporting this conclusion, several studies demonstrated that self-enhancement tendencies affect autobiographical memory (Ross and Qi 2010; Ross and Wilson 2002). Study participants in individualistic cultures tended to remember events that made them proud of themselves as being subjectively more recent than events that made them feel ashamed, while no difference was observed in East Asian participants (Ross and Wilson 2002). Following up on this finding, another study compared

measures of happiness in U.S. and East Asian individuals that were either retrospective ("How satisfied were you with your life this week?") vs. immediate ("How satisfied were you with your life today?"), but averaged over seven days. The usual differences in life satisfaction between East and West emerged when retrospective but not immediate measures were used. In other words, when asked to report their satisfaction within the last week from memory, U.S. Americans were more satisfied than East Asians, although daily reports revealed no differences, suggesting that cross-cultural differences in mean levels of life satisfaction might be partly explained by cultural differences in self-enhancement and autobiographical memory biases (Oishi 2002). Hence, although cross-national differences in mean levels of SWB are large, temporally stable, and show consistent correlations with countries' socioeconomic, political and cultural characteristics, it remains to be explored to what extent these differences are "true" or just reflect cultural differences in in-depth psychological processes, such as memory biases.

5 A Joint Examination of Individual- and Country-level Predictors of SWB

Countries might not only differ in mean levels of SWB, but also in its correlates. Are sources of happiness culturally universal, or do they vary across countries? And are there any systematic patterns in these variations?

In contrast to sociological research tradition, psychological research often tends to assume universality in its theory and findings (Oishi et al. 2009). This tendency is evident in the recently criticized propensity of psychological researchers to draw conclusions about human nature from data obtained from WEIRD (Western, Educated, Industrialized, Rich, Democratic) samples (Henrich et al. 2010), as well as in theories of universal human needs and motivation. For example, Maslow advanced the idea of universal needs as far back as in the 1940s—needs inherent to human nature (Maslow 1943). Striving for universality can be observed in later theoretical work as well. For example, Ryff's psychological model of well-being postulates the existence of six universal needs (autonomy, growth, relationships, purpose in life, environmental mastery, and self-acceptance), the fulfillment of which leads to well-being (Ryff 1989), whereas Deci and Ryans' self-determination theory advances the existence of three such universal needs (autonomy, competence, relatedness) (Deci and Ryan 2000).

While intuitively appealing, such a universalistic approach has been only partially supported by empirical research. Even though positive relationships, self-esteem and other presumably universal sources of SWB typically show positive associations with life satisfaction across cultures, the magnitude of such associations often differs (Diener and Diener 1995). Although substantial efforts have been directed at uncovering systematic patterns in these variations, theory building has lagged behind. I give below an overview of empirical findings and theoretical approaches aimed at explaining cross-national variations in correlates of SWB in sociology, economics, and psychology.

Using the search terms "happiness or life satisfaction or subjective well-being" and "cross-national or cross-cultural" resulted in 1519 articles in academic journals in PsychINFO, PsychARTICLES, EconLIT and an additional 470 in Sociological Abstracts. A brief look at the listed articles showed that most of them do not explore cross-national variability in individual-level associations, but rather explore either predictors of SWB at the individual level or national differences in the mean level of SWB. Therefore, the search was refined by adding the term "multilevel" (as exploring cross-level interactions typically requires a multilevel analysis). This modification resulted in 54 articles in academic journals in PsychINFO, PsychAR-TICLES, EconLIT and an additional 12 in Sociological Abstracts. For this review, I selected empirical papers that explicitly indicated having explored both individual- and country-level predictors of SWB in the abstract. Further relevant empirical papers were detected via studying the literature cited in these articles as well as the literature that cited them. Finally, a number of relevant empirical papers—that were not detected using the search terms described above but that I was familiar with through my work—were considered as well. The list of reviewed papers is presented in the Appendix (Table 1). While this search strategy does not render the present review comprehensive, it nonetheless makes it possible to detect major theoretical and empirical trends in the literature.

An examination of the empirical work showed that most of the existing empirical results can be categorized as adopting one of two approaches: I describe the first one as the "institutional hypothesis", while the second one is typically referred to in the literature as the "fit hypothesis".

5.1 The Institutional Hypothesis

The theoretical reasoning behind most sociological and economic research exploring cross-national variability in correlates of SWB can be summarized as follows: individuals' characteristics contribute to happiness to the extent that macro-level conditions are favorable to individuals with these characteristics—what I refer to as the "institutional hypothesis". Below, I summarize research findings supporting (and refuting) the institutional hypotheses across different domains, including employment, family life, and health.

5.1.1 Employment

The adverse consequences of job loss for happiness and life satisfaction have been largely acknowledged in the literature (Carroll 2007). Cross-national studies in sociology and economics additionally explored whether labor market policies, including unemployment benefits and employment protection regulations, could mitigate the negative effect of unemployment. For example, Voßemer et al. (2017) studied the role of cross-national differences in labor market policies in shaping the well-being and health of unemployed individuals across 26 European countries. Consistent with the institutional hypothesis, they found that unemployed individuals were better off in countries with more (vs. less) generous unemployment benefits. This finding was further refined by Ochsen and Welsch (2012), who distinguished between the effects

of employment protection policies and unemployment benefits, using the data from nearly 400,000 individuals in ten European countries, from 1975 to 2002. Their analyses showed that employment protection and a higher benefit replacement rate were positively associated with the life satisfaction of everyone, with employment protection being particularly appreciated by the employed but higher replacement rates—by the unemployed. Having said that, not all the studies supported the institutional hypothesis. For example, Eichhorn's (2014) analyses of the data from the European Values Study including 28 countries showed that the effect of unemployment on life-satisfaction was not moderated by unemployment benefits. In another study, a generous social policy regarding unemployment benefits has been shown to mitigate the negative effect of financial hardship on the SWB of self-employed individuals across 31 European countries (Annink et al. 2016), suggesting that the beneficial effect of supportive institutions might be more specific than previously assumed, at least with respect to unemployment benefits.

5.1.2 Family Life

Another type of social benefits that has been explored as a factor potentially contributing to the well-being of social groups in need of support are family and parental benefits. Most studies examining the effect of parenthood on SWB have shown that parents typically report lower SWB levels than childless individuals (e.g. Luhmann et al. 2012). Can the burden of parenthood be alleviated by welfare state support? Glass et al. (2016) showed that the negative effect of parenthood on happiness was weaker in countries with more (vs. less) generous welfare policies supporting families. It has been suggested that welfare state support directed at parents can compensate for the stress of parenthood, reducing the disparities in happiness between parents and nonparents (see also Hank and Steinbach 2019). Generous welfare state support has also been shown to mitigate the negative effect of being childless in old age. Specifically, Neuberger and Preisner (2017) used the data of the Survey of Health, Ageing and Retirement in Europe (including 19 countries) and the English Longitudinal Study of Ageing—both large-scale surveys including respondents aged 50 and older—to show that generous social benefits are associated with a higher SWB for childless elderly individuals.

5.1.3 Health Care and Aging

Another line of research explored the role of health care policies in mitigating the effect of ill-health on SWB (see also Pförtner et al. 2019). For example, Kööts-Ausmees and Realo (2015) used the data from 32 countries (using the data of the European Social Survey) and showed that even though the association between self-reported health status and life satisfaction is positive across countries, it was weaker in countries with higher (vs. lower) government spending on health care programs. That is, investing in health care appears to reduce the negative effect of ill-health on life satisfaction: becoming ill in a country with meager health care spending might result in accumulating personal debt or refusing medical treatment altogether, ultimately resulting in further health deterioration.

The generosity of the welfare state was shown not only to compensate for individuals' health problems, but also to make life easier for the elderly. For example, Moor, de Graaf and Komter (2013) showed that older individuals are more satisfied with life in countries with better welfare services targeted at the elderly. Specifically, they detected a positive association between the share of the elderly in institutional care homes (used as an indicator of the welfare generosity directed at the elderly) and the life satisfaction of elderly individuals in general. Similarly, the negative consequences of financial hardship among the elderly are buffered by the generosity of the welfare state. Specifically, Niedzwiedz et al. (2015), using the data of elderly individuals (Survey of Health, Ageing, and Retirement in Europe) living in Southern, Scandinavian, Postcommunist, and Bismarckian welfare regimes, demonstrated that the negative effect on SWB exerted by low socioeconomic status was attenuated in Scandinavian regimes (which are known for their generous social spending).

5.1.4 Migration

In light of the ongoing migration and refugee crisis in the Europe, several studies examined the role of social policies in migrants' satisfaction (see also Careja 2019). Kogan et al. (2017) showed that the life satisfaction of immigrants in Europe (18 EU countries were analyzed) depends much more on natives' attitudes toward migration than it does on legal immigration regulations and policies. A welcoming social climate including positive migration attitudes was shown to make a happy migrant. Heizmann and Böhnke (2018) distinguished between migrants from the EU and other countries, and found that Kogan and colleagues' findings primarily apply to migrants from the EU, whereas the life satisfaction of migrants from outside the EU does benefit from inclusive integration policies.

5.1.5 Gender

Besides specific policies directed at improving the living conditions of one or another group, cross-national sociological research considered inequalities between different social groups. For example, Başlevent and Kirmanoğlu (2017) explored the happiness gap between working women and housewives across 29 European countries. They showed that working women were happier than housewives and that this happiness gap increased with increasing country-level gender equality in economic participation, educational attainment, and political empowerment. Cross-national differences in gender equality were also shown to explain between-country variation in gender differences in happiness, with women being more satisfied in more gender-equal countries (Tesch-Roemer et al. 2008). However, this finding was not replicated among adolescents. Looze et al. (2017) used individual-level data of over 150,000 adolescents from 34 European and North American countries. Their analyses indicated that the national-level of gender equality (e. g. women's economic participation and decision-making power in politics and business) showed a positive association with life satisfaction in both boys and girls to the same extent. Similarly, despite some support for a positive effect of gender equality on women's happiness, researchers have also noticed that the happiness of American women has been in

decline since the 1970s, despite growing gender equality (Stevenson and Wolfers 2009). Potentially, increases in gender equality resulted in a higher role ambiguity and role conflict in women who are now supposed to succeed in both family life and at work (see also Grunow 2019). These findings point toward a limitation of the institutional approach, showing that an objective improvement in living conditions might do little to enhance happiness, unless it is followed by the requisite changes in the sociocultural climate.

5.1.6 Miscellaneous

Although most research testing the "institutional hypothesis" explored the role of social policies, institutional and structural differences can more generally foster or undermine the well-being of one or another group as well. For example, Tay et al. (2014), who analyzed individual-level data from over 100 countries, showed that the positive association between individuals' income and SWB is enhanced in countries with higher national corruption levels (probably as it allows one to bribe more officials—an ability crucial for survival in corrupt countries). Sortheix and Schwartz (2017) examined the effect of personal values from Schwartz's values theory on life satisfaction across 25 European countries and showed that while the effects of some values did not show meaningful cross-cultural variations (e.g. high benevolence and hedonism values were related to higher life satisfaction, whereas high power and security values were related to lower life satisfaction), the effects of other values were less universal[1]. For example, individuals scoring high on achievement value were more satisfied with life if they lived in countries with lower scores in the Human Development Index, whereas individuals scoring lower on achievement values were happier in countries with higher scores in the Human Development Index. A strong achievement orientation might potentially foster prosperity, which is more important for well-being in developing than in developed countries.

5.2 The Fit Hypothesis

The role of sociocultural climate has been the focus of psychological research exploring cross-cultural differences in correlates of SWB. This line of research advanced what can be called a "fit hypothesis", according to which individuals' characteristics contribute to SWB to the degree to which these characteristics are widespread and, consequently, socially desirable in a particular culture (Stavrova 2014; Stavrova et al. 2013a, b). This theoretical perspective is grounded in decades of psychological research on normative conformity, social sanctions, social identity and person–environment fit (Abrams et al. 2002; Cialdini and Goldstein 2004; Kristof-Brown et al. 2005).

Studies on the psychology of social processes showed that people like similar others more than dissimilar others (Byrne 1961) and judge members of their in-group more positively than members of their out-group (Tajfel et al. 1971). Similarly, peo-

---

[1] Personal values represent individual characteristics (other than personality traits) that reflect individuals' guiding principles in life (Schwartz 1992).

ple judge group members holding opinions or views that deviate from the group's average as socially unattractive and unlikeable; they are more likely to ostracize and exclude them from the group (Abrams et al. 2002; Ouwerkerk et al. 2005). This idea is supported by the studies on conformity and norm violations, showing that people whose behavior or opinions deviate from social norms are often subject to informal social sanctions, such as disapproval and exclusion (Christensen et al. 2004; Pool et al. 1998). In fact, as shown by the research on the backlash effect, behaviors typically related to social approval and respect might backfire and result in disrespect when executed by those for whom they represent a violation of social norms (Rudman and Fairchild 2004). For example, volubility, expressions of anger and assertive negotiation techniques have overwhelmingly positive consequences for men, but negative ones for women (Brescoll 2011; Brescoll and Uhlmann 2008). People are often aware of the backlash that being different than the majority can elicit. This is illustrated by the so called Spiral of Silence phenomenon that describes the hesitancy with which minority (vs. majority) opinion holders express their opinions (Noelle-Neumann 1974). For example, individuals who believe that their opinion is discrepant from (vs. congruent with) the majority's opinion are slower (that is, more hesitant) when it comes to publicly sharing their opinion (Bassili 2003).

Finally, studies in organizational psychology (Kristof-Brown et al. 2005) have demonstrated the importance of fit between individuals and their immediate work contexts. These studies have shown that individuals whose values, goals, and personality are similar to those in their organization or team are more likely to have higher job satisfaction, commitment and performance than individuals who are dissimilar to their immediate work environment (Kristof-Brown et al. 2005).

The idea of fit has transcended cross-cultural literature on correlates of SWB as well. Cross-cultural studies in psychology traditionally focus on comparing the predictive validity of the same characteristic across several countries with different scores in important cultural dimensions, such as individualism vs. collectivism. Individualistic cultures value positive emotions, autonomy and self-concern, whereas collectivistic cultures place emphasis on relationships concern, interdependence and normative conformity (Oyserman et al. 2002). Consistent with the fit perspective, relationship concern was shown to be a stronger predictor of SWB in collectivistic cultures, whereas self-concern was a stronger predictor of SWB in individualistic ones (Mesquita and Karasawa 2002). Similarly, satisfaction with one's freedom, experiences of positive emotions and valuing pleasure (hedonism) were more strongly associated with life satisfaction in individualistic countries than in collectivistic ones (Joshanloo and Jarden 2016; Oishi et al. 1999; Suh et al. 1998). Taken together, these findings suggest that people are happy and satisfied with life to the extent that their characteristics are in harmony with cultural values.

Although cultural differences along the individualism vs. collectivism dimension are well established, they might not represent the optimal way to explore the fit hypothesis. Individualism–collectivism (as well as any other cultural dimension) describes cultural differences not just in one but in multiple values, needs, traditions, customs and the like, and thus lack the specificity required for a straightforward test of the fit idea.

A most recent stream of research therefore explored the fit hypothesis more directly by investigating the interactions between individuals' characteristics and the degree to which these characteristics are widespread and socially desirable in a given cultural context. These two aspects of normative behaviors—being common and socially desirable—represent core attributes of social norms in social psychological literature (Cialdini et al. 1990). They are referred to as descriptive (as they describe behaviors that people commonly exercise under certain circumstances) and injunctive (as they describe behaviors that people believe one should exercise under certain circumstances). It should be noted that this approach differs from the dominant view in the sociological literature that only recognizes the injunctive (but not the descriptive) component as part of the definition of social norms (Coleman 1990). Using the social norms approach to explain cross-national differences produced promising results by demonstrating the importance of fit for SWB in different life domains. For example, the happiness gap between employed and unemployed individuals was shown to vary as a function of a country-level social norm to work, with the unemployed being especially dissatisfied with their lives in countries with a strong social work ethic (Stavrova et al. 2011).

The person–culture fit pattern also emerged with respect to other areas of human life, such as marital status and parenthood. For example, while cohabiting women are often shown to report lower SWB than their married counterparts, this effect was shown to be restricted to countries with a strong norm for women to get married (Stavrova et al. 2012). Similarly, the misery of single parents and parents raising a child out of wedlock (cohabiting) could be substantially alleviated in cultural contexts with tolerant norms regarding childbearing practices. In countries where childbearing in cohabiting unions and by single parents is common and normatively accepted, there was a smaller gap in life satisfaction between married and cohabiting or single parents than in countries where these alternative family models were not accepted (Stavrova and Fetchenhauer 2015a, 2015b).

Interestingly, support for the fit hypothesis was also demonstrated with respect to less visible individual characteristics such as beliefs, worldviews, and ideologies. A series of studies showed that the positive effect of religious beliefs on SWB was restricted to religious countries (Gebauer et al. 2017; Stavrova et al. 2013). Similarly, holding specific secular beliefs, such as a belief in scientific–technological progress, was associated with a higher SWB to the extent that this belief was common in a particular cultural context (Stavrova et al. 2016). Conservative political ideology was a positive predictor of SWB only in above-average conservative contexts (Stavrova and Luhmann 2016a). Finally, even the effect of characteristics that are sometimes described as representing universal sources of happiness, such as virtue and prosociality (Aknin et al. 2013) has been shown to vary across countries, following a fit pattern. For example, helping others was related to higher life satisfaction in countries where helping was common than in countries where helping others was rare (Oarga et al. 2015). Similarly, the positive effect of civic virtue on life satisfaction turned negative in countries where above-average virtuous behaviors in economic games was subject to sanctions (Stavrova et al. 2013b).

What mechanisms account for the fit effect? On the one hand, individuals who deviate from the majority behavior (or in other words who violate social norms) are

likely to be sanctioned by their fellow group members. Indeed, multiple studies in social psychology and sociology have shown that individuals expressing a deviant opinion are more likely to be disliked by others and even ostracized (Christensen et al. 2004; Wood et al. 1997). In turn, social exclusion represents one of the most important factors that negatively affect psychological functioning (Williams 2007; Williams et al. 2000). Similarly, loneliness has been shown to hinder psychological well-being, resulting in poor health and even increased mortality risks (Cacioppo and Cacioppo 2014; Cacioppo et al. 2010). It follows that pursuing a lifestyle that significantly deviates from the social norms in one's surroundings is likely to give rise to negative behaviors from others, resulting in social isolation and therefore damaging one's sense of well-being. In brief, this explanation assumes that the reason behind the unhappiness of norm-deviant individuals are social sanctions—the toll of disrespect and the pain of social exclusion.

At the same time, social norms work not only because people fear punishment (social sanctions) in case of noncompliance, but also because they are often internalized and become part of individuals' selves. In this case, a lower SWB of norm-deviant individuals might be explained by their self-disappointment and a feeling of guilt associated with failing to live up to one's own standards. Psychological studies have shown that people tend to make an effort to minimize the discrepancy between their actual and ideal self-concept (Crocker and Knight 2005). This mechanism is most likely to be at work in the case of involuntary nonconformity, such as involuntary unemployment (in a country with a strong work ethic), cohabitation (in a country with strong traditional family norms) or single parenthood (in a country with a strong two-parent family norm). Yet, nonconformity is sometimes more voluntary: People usually have some degree of freedom to decide whether they want to support a minority political ideology or become an atheist in a religious country. The negative consequences of nonconformity are therefore hard to explain in such cases by individuals' self-disappointment or guilt associated with falling short of reaching their ideal selves. After all, they could become their ideal selves if they just changed their views.

Several papers have attempted to explore which of the two mechanisms discussed above—deviance from social vs. personal norms—can best explain the negative effect of a lack of fit. In a cross-cultural study of unemployed individuals' well-being, Stavrova et al. (2011) included a measure of the norm to work at both individual and national level (the latter was obtained by aggregating individuals' responses to the norm to work scale). If individuals' deviation from their personal norm of work and the resulting feeling of self-discrepancy and guilt represent the underlying mechanisms of the fit effect, then the interaction between individuals' employment status and the national norm to work should vanish once researchers control for individual differences in the personal norm to work. Yet this did not take place: the stronger the country-level norm to work was, the less happy the unemployed were, regardless of how strong or weak their personal norm to work was. Similar results were obtained in a couple of other studies that explored the role of social vs. personal norms. For example, in a study of 43 European countries, Stavrova and Fetchenhauer (2015b) showed that single parents were particularly dissatisfied with life in countries with a strong two-parent family norm. Importantly, this effect was

present even among single parents who themselves did not hold a two-parent family norm, pointing to the working of the social sanction mechanism of the fit effect. The same conclusions were reached in a study of cohabiting parents' happiness: cohabiting (vs. married) parents were less happy in countries with strong traditional family norms, regardless of whether they themselves supported these norms or not (Stavrova and Fetchenhauer 2015a). Taken together, these findings provide support for the social sanctions (rather than norm internalization) explanation of the fit effect.

To summarize, the literature reviewed points to substantial cross-cultural variability in the degree to which individuals' characteristics are associated with SWB and thus challenges the assumption of the universality of human happiness. Existing literature seems to converge on the idea that multiple individuals' characteristics contribute to happiness to the extent that they are common and socially desirable in a particular sociocultural context. The cross-national variability in effect sizes can be considered as not only statistically but also practically significant. In fact, in some domains, the particularities of a national context do not only make an individual-level association smaller, but actually make it vanish or even reverse. Knowing what country-level characteristics switch the predictors of SWB on and off and why represents an important step in assessing the cross-national generalizability of SWB findings.

6 Directions for Future Research

Although cross-national comparative research on SWB has been greatly facilitated by the inclusion of measures of life satisfaction in most cross-national large-scale survey programs, it is still in its infancy. The vast majority of cross-cultural studies in psychology still involve comparing findings across two or three countries or cultural contexts. While this approach usually uses highly valid instruments and sophisticated (experimental or longitudinal) study designs, it does not permit any definitive conclusions to be drawn about the role of country-level characteristics in explaining between-country differences in the results. To give an example, observing a positive association between self-esteem and happiness in individualistic Canada and a zero correlation in collectivistic Japan is informative, yet it tells us little about the role of individualism in driving these differences.

In this sense, studies that rely on large-scale cross-national survey datasets and statistically test the effect of contextual characteristics on within-country relationships (e. g. using a multilevel analysis) represent an important methodological advancement. The availability of basic SWB measures in multiple large-scale cross-national studies currently allows researchers to explore the variations in correlates of life satisfaction across more than 100 countries. Yet, this research is restricted by the measures available in such datasets. While personality characteristics (e. g. the Big Five personality traits) represent the most important predictors of SWB, there is currently not a single publicly available cross-national dataset that includes reliable measures of personality. On the contrary, such datasets usually include measures of sociodemographic characteristics, values, attitudes, and behaviors—aspects that were shown to explain relatively little variance in SWB (Lyubomirsky et al.

2005b). As a result, we know quite a lot about the cross-cultural variability of factors whose overall impact on SWB is small, but we know very little about the cross-cultural variability of the most important predictors of SWB, such as personality. As SWB is not only a topic of interest to psychologists, but also to sociologists and economists, it is highly recommended for cross-national large-scale survey projects to include measures of basic dimensions of personality to fill in this knowledge and data infrastructure gap.

Another limitation that is associated with the nature of available cross-national data is that such data are usually cross-sectional and, thus, do not allow causal inferences. Nevertheless, most studies using cross-national data tend to interpret their results in causal terms, typically implying a causal effect of individuals' characteristics on their happiness (see also Schmidt-Catran et al. 2019). Indeed, most theoretical models in sociology and economics consider happiness as an outcome, not as a predictor. However, recent longitudinal studies have shown that individual differences that usually serve to predict SWB might well be its outcomes. These studies have shown that life satisfaction and happiness contribute to better health and longevity (Danner et al. 2001; Diener and Chan 2011; Stavrova 2019), as well as career success (Cropanzano and Wright 1999; Rose and Stavrova 2019), and can even trigger important life events such as marriage or separation (Luhmann et al. 2013; Stavrova and Luhmann 2016b). These findings make the interpretation of established patterns of cross-cultural variations in correlates of SWB problematic. For example, the fit perspective assumes that endorsing culturally shared beliefs contributes to SWB. Alternatively, it is plausible that (1) SWB contributes to the endorsement of culturally shared beliefs or (2) that both SWB and the endorsement of culturally shared beliefs are driven by further unassessed factors (e.g. high self-control might result in both higher SWB and conformity). It is possible that not only the strength but also the causal direction of the associations between individual characteristics and SWB varies across countries. To conclude, extending cross-national studies on correlates of SWB to include longitudinal or experimental (when appropriate) data in select countries will become crucial in advancing this research field.

While most empirical work exploring cross-national variability in correlates of SWB falls under one of the two theoretical approaches described here—institutional and fit hypotheses—there are multiple further macro-level indicators reflecting different aspects of individuals' living conditions (wealth, social inequality, political freedom, etc.) that can be used to explain cross-national differences in the correlates of SWB. The most prominent example is the role of national wealth (GDP per capita) in the association between individuals' income and their SWB. A couple of studies have shown that the positive association between individuals' income and their SWB is stronger in poorer countries than in wealthier ones (Diener et al. 2010; Schyns 2002), suggesting that money buys happiness when it ensures that basic needs are met (access to clean water, medical care, etc.). An exploration of how country-level differences in further objectively measurable characteristics such as income inequality, human development or particularities of the political regime might be a worthwhile endeavor for future cross-national comparative research.

It should be acknowledged that cross-national comparative research on hedonic components of well-being has been flourishing despite these limitations. In contrast, however, cross-cultural studies on eudaimonic well-being are virtually nonexistent. Yet, eudaimonic well-being—in particular its currently most often studied component, meaning in life—is of increasing importance for both individuals and society to function (Hill and Turiano 2014; Stavrova and Luhmann 2016b). In fact, monitoring and seeking to improve citizens' sense of meaning in life has been discussed as a potential matter of public policy (Steger 2014). Hence, exploring cross-cultural differences in mean levels and correlates of meaning in life might represent an important direction for future studies.

On a related note, the theoretical developments in the area of cross-national SWB research might be used to explain cross-national differences in predictors of physical health. Indeed, a series of recent studies have detected that cross-national variations in the associations between religiosity and physical health show a fit pattern as well. Across nearly 60 countries, religious individuals only enjoyed better physical health than nonreligious individuals in countries where religiosity was common and socially desirable. In contrast, in secular countries, being religious was not associated with any physical health advantage. This pattern was detected across the U.S. census regions and was even extended to mortality risks. Religiosity was only related to lower mortality risks in highly religious regions (e.g. the Bible Belt; Stavrova 2015b).

It is noteworthy that certain statistical techniques, such as multilevel regression or multilevel structural equation modeling, made testing both the institutional and the fit hypotheses possible. Nevertheless, a substantial number of countries is typically required for this technique to lead to reliable conclusions, and most of the studies reviewed here barely satisfied these requirements (Hox 2002). Besides including more countries, one way to address this limitation is to move to a lower level of analysis (e.g. studying regions within one or several countries). Adopting this method might be particularly promising when trying to disentangle the institutional and the fit hypotheses: while institutional effects most often exist at national level, the effect of social norms and culture can be tested at lower levels as well.

Finally, understanding what drives cross-cultural differences in correlates of SWB might represent a tool for studying the mechanisms behind the associations between individual characteristics and happiness. For example, research on the positive effect of religiosity on life satisfaction has considered multiple potential mechanisms, including increased social networks (Lim and Putnam 2009), healthy behaviors (Wallace and Forman 1998), as well as social approval and respect (Gebauer et al. 2017), etc. The fact that the positive effect of religiosity is restricted to religious countries can be interpreted as providing support to the social approval mechanism. In other words, religiosity is only related to increased life satisfaction in contexts where it represents a source of higher social esteem, approval and recognition, suggesting that it is these social benefits (rather than healthy behaviors or other factors) that represent the underlying mechanism of the effect.

7 Conclusions

Why study happiness? Besides feeling good, happiness is associated with many positive life outcomes. Happy people benefit from better health and longevity (Danner et al. 2001), career success (Cropanzano and Wright 1999), and positive relationships (Stavrova and Luhmann 2016b) than their less happy counterparts (for a review, see Lyubomirsky et al. 2005a). Happiness promotes helping and prosociality (Dulin and Hill 2003), broadens individuals' cognitive repertoire (Fredrickson and Branigan 2005) and involvement with approach goals (Elliot and Thrash 2002). Hence, the study of happiness and factors promoting individuals' happiness appears an important endeavor with the potential to improve the human condition.

Cross-national comparative research on SWB has enjoyed fast-growing interest from a number of social science disciplines. These studies made substantial contributions to our understanding of cultural differences in mean levels and correlates of SWB. Further extending empirical studies to include longitudinal and experimental designs and promoting theory building in the area might represent the most immediate steps that future research should undertake.

Appendix

Table 1 Overview of the empirical papers. Author's own work

Paper	Individual-level variables	Country-level variables	Number of countries	Main finding
Voßemer et al. (2017)	Unemployment; insecure employment	Passive and active labor market policies; employment protection legislation	26	Higher unemployment benefits are related to a weaker negative effect of unemployment on SWB. Higher active labor market policies expenditures are associated with stronger negative effects of unemployment on SWB
Ochsen and Welsch (2012)	Employment+ sociodemographics	Employment protection and unemployment benefits policies	10	The positive effect of employment protection on SWB is stronger in employed persons of intermediate age (compared to women/housewives and older people). The effect of generous unemployment insurance on SWB is stronger in women/housewives, older people and the unemployed

Table 1 (Continued)

Paper	Individual-level variables	Country-level variables	Number of countries	Main finding
Eichhorn (2014)	Unemployment	Unemployment benefits; inflation rate; country-level share of the elderly	28	The effect of unemployment on life satisfaction is not moderated by unemployment benefits. Higher inflation rates and higher proportions of the elderly in the population are associated with a stronger negative effect of unemployment on SWB
Annink et al. (2016)	Financial hardship in self-employed individuals	Unemployment benefits policies	31	Financial hardship has a weaker effect on the SWB of self-employed individuals in countries with the presence of unemployment benefits policy
Glass et al. (2016)	Parenthood	Policies allowing paid time off and childcare subsidies	22	The negative effect of parenthood on SWB is weaker in countries with more general policies implementing paid time off and childcare subsidies
Neuberger and Preisner (2017)	Parenthood in old age	Gross domestic product; social service expenditures	19	The positive effect of parenthood is stronger in countries with lower Gross Domestic Product and higher social service spending
Kööts-Ausmees and Realo (2015)	Subjective health	Health care spending	32	The positive association between subjective health and SWB is stronger in countries with lower government spending on health care programs
Niedzwiedz et al. (2015)	Life course socio-economic index	Welfare state regimes	13	The negative effect of a low socioeconomic index on SWB is attenuated in Scandinavian welfare state regimes
Moor et al. (2013)	Old age	Welfare state services targeted at the elderly	47	Older people living in countries with a higher (vs. lower) share of older adults in institutional care facilities report higher SWB. The replacement rates for pensions were not related to the elderly's SWB
Kogan et al. (2017)	Migration status	Natives' attitudes toward immigrants; immigration policies (Migration Integration Policy Index)	18	Positive natives' attitudes towards immigrants have a positive relationship with immigrants' SWB, while immigration regulations and policies are unrelated to immigrants' SWB

Table 1 (Continued)

Paper	Individual-level variables	Country-level variables	Number of countries	Main finding
Heizmann and Böhnke (2018)	Migration status: national citizens, EU citizens and third-country nationals	Natives' attitudes toward immigrants; immigration policies (Migration Integration Policy Index)	25	The SWB of third-country nationals is most strongly (positively) affected by migrant-friendly policy-making, whereas the SWB of EU migrants is mostly affected by positive natives' attitudes toward immigrants
Başlevent and Kirmanoğlu (2017)	Working women vs. housewives	Gender inequality index	29	Working women report higher SWB than housewives, especially in countries with greater gender equality
Tesch-Roemer et al. (2008)	Gender	Gender inequality index	57	Women are more satisfied in more gender egalitarian countries
Looze et al. (2017)	Gender (among adolescents)	Gender inequality index	34	The gender inequality index is equally associated with life satisfaction among both boys and girls (their SWB is higher in more gender-equal countries)
Tay et al. (2014)	Household income	National corruption level	150	The positive association between individuals' income and SWB is enhanced in countries with higher national corruption levels
Sortheix and Schwartz (2017)	Personal values	Cultural egalitarianism index (based on Schwartz cultural values); Human Development Index	32/25	In countries with a lower (vs. higher) Cultural Egalitarianism index, SWB is associated more positively with openness, more negatively with conservation, less negatively with self-enhancement, and less positively with self-transcendence values. A similar pattern emerged for Human Development Index
Oishi et al. (1999)	Financial satisfaction; esteem needs?	Individualism; GDP	39	The positive association between financial satisfaction and SWB is stronger in poorer (vs. wealthier) nations. The positive association between the satisfaction of esteem needs and SWB is stronger in individualistic (vs. collectivistic) nations
Suh et al. (1998)	Emotions	Individualism	61	Emotions are a stronger predictor of SWB in individualistic (vs. collectivistic) countries

Table 1 (Continued)

Paper	Individual-level variables	Country-level variables	Number of countries	Main finding
Joshanloo and Jarden (2016)	Hedonism values	Individualism	19	Hedonism values are more strongly (positively) associated with SWB in individualistic (vs. collectivistic) countries
Stavrova et al. (2011)	Unemployment status	Social work ethic; national unemployment rates	28	The negative effect of unemployment on SWB was alleviated in counties with a weak social work ethic; national unemployment rates did not moderate the effect of unemployment on SWB
Stavrova et al. (2012)	Gender and partnership arrangement (marriage vs. cohabitation)	Gender role norms	30	Cohabiting women report lower SWB than married women do, especially in countries with conservative gender role norms; gender role norms had no effect on cohabiting vs. married men's SWB
Stavrova et al. (2013a)	Personal religiosity	Social norm of religiosity	64	Religiosity is positively associated with SWB, and this relationship is especially strong in countries with a stronger (vs. weaker) norm of religiosity
Stavrova et al. (2013b)	Civic virtue	Country-level civic virtue; country-level antisocial punishment (punishment of high contributors in public goods games) rates	73/13	Civic virtue is positively associated with SWB, but less so in countries with a weak country-level civic virtue and high country-level antisocial punishment rates
Oarga et al. (2015)	Prosocial behavior	Country-level norm of prosociality	23	Prosocial behavior is positively related to SWB, especially in countries with a strong norm of prosociality
Stavrova and Fetchenhauer (2015b)	Marital status and parenthood	Country-level two-parent family norms; individualism	43	Parenthood more negatively affects the SWB of single than of partnered individuals, especially in collectivistic countries and counties with a strong two-parent family norm
Stavrova and Fetchenhauer (2015a)	Marital status and parenthood	Country-level childbearing norms	24	Cohabiting parents report lower SWB than married parents do, but only in countries with a strong norm proscribing childbearing in cohabiting unions

Table 1 (Continued)

Paper	Individual-level variables	Country-level variables	Number of countries	Main finding
Stavrova et al. (2016)	Belief in scientific-technological progress	Country-level belief in scientific-technological progress	72	The positive association between belief in scientific-technological progress and SWB is stronger in countries with a stronger (vs. weaker) average belief in scientific–technological progress
Stavrova and Luhmann (2016a)	Political ideology	Country-level political ideology	92	Political conservatism is positively associated with SWB, especially in countries with stronger average political conservatism
Roex and Rözer (2018)	Unemployment status	Country-level social work ethic	31	The negative effect of unemployment on SWB is stronger in countries with a strong work ethic (especially for men and the long-term unemployed)
Van de Velde et al. (2017)	Religious service attendance; frequency of prayer	Country-level religiosity	29	The negative association between service attendance and depression is weaker in less (vs. more) religious countries. The positive association between frequency of prayer and depression is stronger in less (vs. more) religious countries

References

Abrams, Dominic, José Marques, Nicola Bown, and Melanie Dougill. 2002. Anti-norm and pro-norm deviance in the bank and on the campus: two experiments on subjective group dynamics. *Group Processes & Intergroup Relations* 5:163–182.

Aknin, Lara B., Christopher P. Barrington-Leigh, Elizabeth W. Dunn, John F. Helliwell, Justine Burns, Robert Biswas-Diener, Imelda Kemeza, Paul Claire E.Ashton-James Nyende, and Michael I. Norton. 2013. Prosocial spending and well-being: cross-cultural evidence for a psychological universal. *Journal of Personality and Social Psychology* 104:635–652.

Annink, Anne, Marjan Gorgievski, and Laura Den Dulk. 2016. Financial hardship and well-being: a cross-national comparison among the European self-employed. *European Journal of Work and Organizational Psychology* 25:645–657.

Balatsky, Galina, and Ed Diener. 1993. Subjective well-being among Russian students. *Social Indicators Research* 28:225–243.

Bassili, John N. 2003. The minority slowness effect: Subtle inhibitions in the expression of views not shared by others. *Journal of Personality and Social Psychology* 84:261–276.

Baumeister, Roy F., and Mark R. Leary. 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin* 117:497–529.

Başlevent, Cem, and Hasan Kirmanoğlu. 2017. Gender inequality in Europe and the life satisfaction of working and non-working women. *Journal of Happiness Studies* 18:107–124.

Berger, Peter L., and Thomas Luckman. 1966. *The social construction of reality: a treatise in the sociology of knowledge*. Garden City: Doubleday.

Bonini, Astra. 2008. Cross-national variation in individual life satisfaction: effects of national wealth, human development, and environmental conditions. *Social Indicators Research* 87:223–236.

Bradburn, Norman M. 1969. *The structure of psychological well-being*. Chicago: Aldine.

Brescoll, Victoria L. 2011. Who takes the floor and why. *Administrative Science Quarterly* 56:622–641.

Brescoll, Victoria L., and Eric Luis Uhlmann. 2008. Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science* 19:268–275.

Brief, Arthur P., Ann H. Butcher, Jennifer M. George and Karen E. Link. 1993. Integrating bottom-up and top-down theories of subjective well-being: the case of health. *Journal of Personality and Social Psychology* 64:646–653.

Byrne, Donn. 1961. Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology* 62:713–715.

Cacioppo, John T., and Stephanie Cacioppo. 2014. Social relationships and health: the toxic effects of perceived social isolation. *Social and Personality Psychology Compass* 8:58–72.

Cacioppo, John T., Louise C. Hawkley, and Ronald A. Thisted. 2010. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago health, aging, and social relations study. *Psychology and Aging* 25:453–463.

Careja, Romana. 2019. Immigrants' labor market outcomes: Contributions from multilevel studies. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00597-8.

Carroll, Nick. 2007. Unemployment and psychological well-being. *Economic Record* 83:287–302.

Caspi, Avshalom, Daryl J. Bem and Glen H. Elder. 1989. Continuities and consequences of interactional styles across the life course. *Journal of Personality* 57:375–406.

Cheng, Helen, and Adrian Furnham. 2003. Personality, self-esteem, and demographic predictions of happiness and depression. *Personality and Individual Differences* 34:921–942.

Cheung, Felix, and Richard E. Lucas. 2014. Assessing the validity of single-item life satisfaction measures: results from three large samples. *Quality of Life Research* https://doi.org/10.1007/s11136-014-0726-4.

Christensen, P. Niels, Hank Rothgerber, Wendy Wood and David C. Matz. 2004. Social norms and identity relevance: a motivational approach to normative behavior. *Personality and Social Psychology Bulletin* 30:1295–1309.

Cialdini, Robert B., and Noah J. Goldstein. 2004. Social influence: compliance and conformity. *Annual Review of Psychology* 55:591–622.

Cialdini, Robert B., Raymond R. Reno and Carl A. Kallgren. 1990. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58:1015–1026.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt and René Algesheimer. 2019. How to obtain comparable measures for cross-national comparisons. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00598-7.

Coleman, James S. 1990. *Foundations of social theory*. Cambridge: Harward University Press.

Correia, Isabel, Maria Toscano Batista and Maria Luisa Lima. 2009. Does the belief in a just world bring happiness? Causal relationships among belief in a just world, life satisfaction and mood. *Australian Journal of Psychology* 61:220–227.

Crocker, Jennifer, and Katherine M. Knight. 2005. Contingencies of self-worth. *Current Directions in Psychological Science* 14:200–203.

Cropanzano, Russell, and Thomas A. Wright. 1999. A 5-year study of change in the relationship between well-being and job performance. *Consulting Psychology Journal: Practice and Research* 51:252–265.

Danner, D. Deborah, David A. Snowdon and Wallce. V. Friesen. 2001. Positive emotions in early life and longevity: findings from the nun study. *Journal of Personality and Social Psychology* 80:804–813.

Deci, Edward L., and Richard M. Ryan. 2000. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 11:227–268.

Delle Fave, Antonella, Ingrid Brdar, Marié P. Wissing, Ulisses Araujo, Alejandro Castro Solano, Teresa Freire, María Del Rocío Hernández-Pozo, Paul Jose, Tamás Martos, Hilde E. Nafstad, Jeanne Nakamura, Kamlesh Singh and Lawrence Soosai-Nathan. 2016. Lay definitions of happiness across nations: the primacy of inner harmony and relational connectedness. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.00030

Diener, Ed. 1984. Subjective well-being. *Psychological Bulletin* 95:542–575.

Diener, Ed, and Micaela Y. Chan. 2011. Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being* 3:1–43.

Diener, Ed, and Marissa Diener. 1995. Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology* 68:653–663.

Diener, Ed, and Katherine Ryan. 2009. Subjective well-being: a general overview. *South African Journal of Psychology* 39:391–406.

Diener, Ed, and Eunkook Suh. 2000. *Culture and subjective well-being*. Cambridge: A Bradford Book, The MIT Press.

Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. The satisfaction with life scale. *Journal of Personality Assessment* 49:71–75.

Diener, Ed, Ronald Inglehart, and Louis Tay. 2013. Theory and validity of Life Satisfaction Scales. *Social Indicators Research* 112:497–527.

Diener, Ed, Weiting Ng, James Harter, and Raksha Arora. 2010. Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology* 99:52–61.

Diener, Ed, Shigehiro Oishi, and Richard E. Lucas. 2003. Personality, culture, and subjective well-being: emotional and cognitive evaluations of life. *Annual Review of Psychology* 54:403–425.

Diener, Ed, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith. 1999. Subjective well-being: Three decades of progress. *Psychological Bulletin* 125:276–302.

Diener, Ed, Louis Tay, and David G. Myers. 2011. The religion paradox: if religion makes people happy, why are so many dropping out? *Journal of Personality and Social Psychology* 101:1278–1290.

Di Fabio, Annamaria, and Alessio Gori. 2016. Measuring adolescent life satisfaction: psychometric properties of the satisfaction with life scale in a sample of Italian adolescents and young adults. *Journal of Psychoeducational Assessment* 34:501–506.

Di Tella, Rafael, and Robert MacCulloch. 2006. Some uses of happiness data in economics. *Journal of Economic Perspectives* 20:25–25.

Di Tella, Rafael, Robert MacCulloch, and Andrew J. Oswald. 2003. The macroeconomics of happiness. *The Review of Economics and Statistics* 85:809–827.

Dorn, David, Justina A.V. Fischer, Gebhard Kirchgässner, and Alfonso Sousa-Poza. 2007. Is it culture or democracy? The impact of democracy and culture on happiness. *Social Indicators Research* 82:505–526.

Dulin, Patrick L., and R.D. Hill. 2003. Relationships between altruistic activity and positive and negative affect among low-income older adult service providers. *Aging & Mental Health* 7:294–299.

Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton. 2008. Spending money on others promotes happiness. *Science* 319:1687–1688.

Dunning, David, Heath Chip, and Jerry M. Suls. 2004. Flawed self-assessment: implications for health, education, and the workplace. *Psychological Science in the Public Interest* 5:69–106.

Eichhorn, Jan. 2014. The (non-)effect of unemployment benefits: variations in the effect of unemployment on life-satisfaction between EU countries. *Social Indicators Research* 119:389–404.

Elliot, Andrew, and Todd Thrash. 2002. Approach-avoidance motivation in personality approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology* 82:804–818.

Elliot, Andrew, M. Sheldon Kenon, and Marcy A. Church. 1997. Avoidance personal goals and subjective well-being. *Personality and Social Psychology Bulletin* 23:915–927.

Feiler, Daniel C., and Adam M. Kleinbaum. 2015. Popularity, similarity, and the network extraversion bias. *Psychological Science* 26:593–603.

Florida, Richard, Charlotta Mellander, and Peter J. Rentfrow. 2013. The happiness of cities. *Regional Studies* 47:613–627.

Fredrickson, Barbara L. 2000. Extracting meaning from past affective experiences: the importance of peaks, ends, and specific emotions. *Cognition and Emotion* 14:577–606.

Fredrickson, Barbara L., and Christine Branigan. 2005. Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion* 19:313–332.

Frey, Bruno, and Alois Stutzer. 2002a. *Happiness and economics. How the economy and institutions affect well-being*. Pinceton: Princeton University Press.

Frey, Bruno, and Alois Stutzer. 2002b. What can economists learn from happiness research? *Journal of Economic Literature* 40:402–435.

Frey, Bruno, and Alois Stutzer. 2005. Happiness research: state and prospects. *Review of Social Economy* 62:207–228.

Frey, Bruno, and Alois Stutzer. 2014. Economic consequences of mispredicting utility. *Journal of Happiness Studies* 15:937–956.

Gebauer, Jochen E., Constantine Sedikides, Felix D. Schonbrodt, Wiebke Bleidorn, P. Jason Rentfrow, J. Potter, and Sam D. Gosling. 2017. The religiosity as social value hypothesis: a multi-method replica-

tion and extension across 65 countries and three levels of spatial aggregation. *Journal of Personality and Social Psychology* 113:e18–e39.

Glass, Jennifer, Robin W. Simon, and Matthew A. Andersson. 2016. Parenthood and happiness: effects of work-family reconciliation policies in 22 OECD countries. *American Journal of Sociology* 122:886–929.

Grunow, Daniela. 2019. Comparative analyses of housework and its relation to paid work: Institutional contexts and individual agency. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00601-1.

Hammond, Matthew, and Chris Sibley. 2011. Why are benevolent sexists happier? *Sex Roles* 65:332–343.

Hank, Karsten, and Anja Steinbach. 2019. Families and their institutional contexts: The role of family policies and legal regulations. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00603-z.

Hayes, Natalie, and Stephen Joseph. 2003. Big 5 correlates of three measures of subjective well-being. *Personality and Individual Differences* 34:723–727.

Headey, Bruce, and A. Wearing. 1992. *Understanding happiness: a theory of subjective well-being*. Melbourne: Longman Cheshire.

Heine, Steven J., and Takeshi Hamamura. 2007. In search of East Asian self-enhancement. *Personality and Social Psychology Review* 11:1–24.

Heizmann, Boris, and Petra Böhnke. 2018. Immigrant life satisfaction in Europe: the role of social and symbolic boundaries. *Journal of Ethnic and Migration Studies*. https://doi.org/10.1080/1369183X.2018.1438252

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral & Brain Sciences* 33:61–135.

Hill, Patrick L., and Nicholas A. Turiano. 2014. Purpose in life as a predictor of mortality across adulthood. *Psychological Science* 25:1482–1486.

Hofmann, Wilhelm, Maike Luhmann, Rachel R. Fisher, Kathleen D. Vohs, and Roy F. Baumeister. 2014. Yes, but are they happy? Effects of trait self-control on affective well-being and life satisfaction. *Journal of Personality* 82:265–277.

Hofstede, Geert. 2001. *Culture's consequences: comparing values, behaviors, institutions and organizations across nations*. London: SAGE.

Hox, Joop. 2002. *Multilevel analysis: techniques and applications*. New Jersey: Lawrence Erlbaum.

Joshanloo, Mohsen, and Aaron Jarden. 2016. Individualism as the moderator of the relationship between hedonism and happiness: a study in 19 nations. *Personality and Individual Differences* 94:149–152.

Joshanloo, Mohsen, Zarina K. Lepshokova, Tatiana Panyusheva, Natalia Amerkhanova, Wai-Ching Poon, Victoria Wai-lan Yeung, Suresh Sundaram, Mustapha Achoui, Ryosuke Asano, Tasuku Igarashi, Saori Tsukamoto, Muhammad Rizwan, Imran Ahmed Khilji, Maria Cristina Ferreira, Joyce S. Pang, Lok Sang Ho, Gyuseog Han, Jaechang Bae, and Ding-Yu Jiang. 2014. Cross-cultural validation of fear of happiness scale across 14 national groups. *Journal of Cross-Cultural Psychology* 45:246–264.

Jovanović, Veljko. 2016. The validity of the satisfaction with life scale in adolescents and a comparison with single-item life satisfaction measures: a preliminary study. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation* 25:3173–3180.

Kahneman, Daniel. 2000. Experienced utility and objective happiness: a moment-based approach. In *Choices, values and frames*, ed. Daniel Kahneman, Amos Tversky, 673–692. New York: Cambridge University Press.

Kahneman, Daniel, and Angus Deaton. 2010. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences* 107:16489–16493.

Kasser, Tim, and Richard M. Ryan. 1993. A dark side of the American dream: correlates of financial success as a central life aspiration. *Journal of Personality and Social Psychology* 65:410–422.

Kim, Hyoun, and Patrick C. MacKenry. 2002. The relationship between marriage and psychological well-being: a longitudinal analysis. *Journal of Family Issues* 23:885–911.

Kogan, Irena, Jing Shen, and Manuel Siegert. 2017. What makes a satisfied immigrant? Host-country characteristics and immigrants' life satisfaction in eighteen European countries. *Journal of Happiness Studies* 19:1783–1809.

Kööts-Ausmees, Liisi, and Anu Realo. 2015. The association between life satisfaction and self-reported health status in Europe. *European Journal of Personality* 29:647–657.

Kristof-Brown, Amyl, Ryan D. Zimmerman, and Erin C. Johnson. 2005. Consequences of individuals' fit at work: a meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology* 58:281–342.

Lachman, Margie E., and Stefan Agrigoroaei. 2010. Promoting functional health in midlife and old age: Long-term protective effects of control beliefs, social support, and physical exercise. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0013297

Lim, Young-Jin. 2015. Psychometric characteristics of the Korean version of the Satisfaction With Life Scale adapted for children. *Canadian Journal of School Psychology* 30:246–251.

Lim, Chaeyoon, and Robert D. Putnam. 2009. Praying alone is no fun: religion, social networks, and subjective well-being. In *Paper presented at the meeting of the Canadian Institute for Advanced Research*. Toronto, Ontario: Institute for Advanced Research.

Liu, Pan, William Tov, Michal Kosinski, David J. Stillwell, and Lin Qiu. 2015. Do Facebook status updates reflect subjective well-being? *Cyberpsychology, Behavior, and Social Networking* 18:373–379.

de Looze, Margaretha E., Tim Huijts, Gonneke Stevens, Torbjorn Torsheim, and Wilma Vollebergh. 2017. The happiest kids on earth. Gender equality and adolescent life satisfaction in Europe and North America. *Journal of Youth and Adolescence* 47:1073–1085.

Lucas, Richard. 2007. Adaptation and the set-point model of subjective well-being: does happiness change after major life events? *Current Directions in Psychological Science* 16:75–79.

Lucas, Richard E., and Portia S. Dyrenforth. 2006. Does the existence of social relationships matter for subjective well-being? In *Self and relationships: Connecting intrapersonal and interpersonal processes*, ed. Kathleen D. Vohs, Eli J. Finkel, 254–273. New York: Guilford.

Lucas, Richard, A. Clark, Yannis Georgellis, and E. Diener. 2004. Unemployment alters the set point for life satisfaction. *Psychological Science* 15:8–13.

Lucas, Todd, Ludmila Zhdanova, Craig Wendorf, and Sheldon Alexander. 2013. Procedural and distributive justice beliefs for self and others: multilevel associations with life satisfaction and self-rated health. *Journal of Happiness Studies* 14:1325–1341.

Lucas, Richard E., Ed Diener, Alexander Grob, Eunkook M. Suh, and Liang Shao. 2000. Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology* 79:452–468.

Luhmann, Maike, Wilhelm Hofmann, Michael Eid, and Richard Lucas. 2012. Subjective well-being and adaptation to life events: a meta-analysis. *Journal of Personality and Social Psychology* 102:592–615.

Luhmann, Maike, Richard E. Lucas, Michael Eid, and E. Diener. 2013. The prospective effect of life satisfaction on life events. *Social Psychological and Personality Science* 4:39–45.

Luhmann, Maike, Ulrich Schimmack, and Michael Eid. 2011. Stability and variability in the relationship between subjective well-being and income. *Journal of Research in Personality* 45:186–197.

Lykken, David, and Auke Tellegen. 1996. Happiness is a stochastic phenomenon. *Psychological Science* 7:186–189.

Lyubomirsky, Sonja, and Heidi S. Lepper. 1999. A measure of subjective happiness: preliminary reliability and construct validation. *Social Indicators Research* 46:137–155.

Lyubomirsky, Sonja, M. Kennon Sheldon, and David Schkade. 2005b. Pursuing happiness: the architecture of sustainable change. *Review of General Psychology* 9:111–131.

Lyubomirsky, Sonja, Laura King, and Ed Diener. 2005a. The benefits of frequent positive affect: does happiness lead to success? *Psychological Bulletin* 131:803–855.

López-Ortega, Mariana, Sara Torres-Castro, and Oscar Rosas-Carrasco. 2016. Psychometric properties of the satisfaction with life scale (SWLS): secondary analysis of the Mexican Health and Aging Study. *Health and Quality of Life Outcomes*. https://doi.org/10.1186/s12955-016-0573-9

Markus, Hazel R., and Shinobu Kitayama. 1991. Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review* 98:224–253.

Maslow, Abraham H. 1943. A theory of human motivation. *Psychological Review* 50:380–396.

McMahon, Darrin M. 2006. *Happiness: A history*. New York: Atlantic Monthly Press.

Mesquita, Batja, and Mayumi Karasawa. 2002. Different emotional lives. *Cognition and Emotion* 16:127–141.

Moor, Nienke, Paul M. de Graaf, and Aafke Komter. 2013. Family, welfare state generosity and the vulnerability of older adults: a cross-national study. *Journal of Aging Studies* 27:347–357.

Napier, Jaime L., and John T. Jost. 2008. Why are conservatives happier than liberals? *Psychological Science* 19:565–572.

Neuberger, Franz Stephan, and Klaus Preisner. 2017. Parenthood and quality of life in old age: the role of individual resources, the welfare state and the economy. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1665-6

Niedzwiedz, Claire L., Jill P. Pell, and Richard Mitchell. 2015. The relationship between financial distress and life-course socioeconomic inequalities in well-being: cross-national analysis of European welfare states. *American Journal of Public Health* 105:2090–2098.

Noelle-Neumann, Elizabeth. 1974. The spiral of silence: a theory of public opinion. *Journal of Communication* 24:43–51.

Oarga, Cristina, Olga Stavrova, and Detlef Fetchenhauer. 2015. When and why is helping others good for well-being? The role of belief in reciprocity and conformity to society's expectations. *European Journal of Social Psychology* 45:242–254.

Ochsen, Carsten, and Heinz Welsch. 2012. Who benefits from labor market institutions? Evidence from surveys of life satisfaction. *Journal of Economic Psychology* 33:112–124.

Oishi, Shigehiro. 2002. The experiencing and remembering of well-being: a cross-cultural analysis. *Personality and Social Psychology Bulletin* 28:1398–1406.

Oishi, Shigehiro, Jesse Graham, Selin Kesebir, and Iolanda Costa Galinha. 2013. Concepts of happiness across time and cultures. *Personality and Social Psychology Bulletin* 39:559–577.

Oishi, Shigehiro, Selin Kesebir, and Ed Diener. 2011. Income inequality and happiness. *Psychological Science* 22:1095–1100.

Oishi, Shigehiro, Selin Kesebir, and Benjamin H. Snyder. 2009. Sociology: a lost connection in social psychology. *Personality and Social Psychology Review* 13:334–353.

Oishi, Shigehiro, Ed Diener, Richard Lucas, and Eunkook Suh. 1999. Cross-cultural variations in predictors of life satisfaction: Perspectives from needs and values. *Personality and Social Psychology Bulletin* 25:980–990.

Ormel, Johan, Siegwart Lindenberg, Nardi Steverink, and Lois M. Verbrugge. 1999. Subjective well-being and social production functions. *Social Indicators Research* 46:61–90.

Orth, Ulrich, and Eva C. Luciano. 2015. Self-esteem, narcissism, and stressful life events: testing for selection and socialization. *Journal of Personality and Social Psychology* 109:707–721.

Ott, Jan. 2011. Government and happiness in 130 nations: good governance fosters higher level and more equality of happiness. *Social Indicators Research* 102:3–22.

Ouwerkerk, Jaap W., Norbert L. Kerr, Marcello Gallucci, and Paul A.M. van Lange. 2005. Avoiding the social death penalty: ostracism and cooperation in social dilemmas. In *The social outcast*, ed. D. Williams Kipling, Joseph P. Forgas, and William von Hippel, 321–332. New York: Psychology Press.

Oyserman, Daphna, Heather M. Coon, and Markus Kemmelmeier. 2002. Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin* 128:3–72.

Pavot, William, and Ed Diener. 1993. Review of the satisfaction with life scale. *Psychological Assessment* 5:164–172.

Pavot, William, and Ed Diener. 2008. The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology* 3:137–152.

Pförtner, Timo-Kolja, Holger Pfaff and Frank J. Elgar. 2019. The role of welfare state characteristics for health and inequalities in health from a cross-national perspective: A critical research synthesis. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00606-w.

Plomin, Robert. 1994. *Genetics and experience: the interplay between nature and nurture*. London: SAGE.

Pool, Gregory J., Wendy Wood, and Kira Leck. 1998. The self-esteem motive in social influence: agreement with valued majorities and disagreement with derogated minorities. *Journal of Personality and Social Psychology* 75:967–975.

Redelmeier, Donald A., and Daniel Kahneman. 1996. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66:3–8.

Reitz, Anne K., Frosso Motti-Stefanidi, and Jens B. Asendorpf. 2016. Me, us, and them: testing sociometer theory in a socially diverse real-life context. *Journal of Personality and Social Psychology* 110:908–920.

Rentfrow, Peter Jason, Samuel D. Gosling, and Jeff Potter. 2008. A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science* 3:339–369.

Rentfrow, Peter Jason, Charlotta Mellander, and Richard Florida. 2009. Happy states of america: a state-level analysis of psychological, economic, and social well-being. *Journal of Research in Personality* 43:1073–1082.

Roex, Karlijn L.A., and Jesper J. Rözer. 2018. The social norm to work and the well-being of the short- and long-term unemployed. *Social Indicators Research* 139:1037–1064.

Rose, Damaris, and Olga Stavrova. 2019. Does life satisfaction predict reemployment? Evidence form German panel data. *Journal of Economic Psychology* 72:1–11.

Ross, Michael, and Wang Qi. 2010. Why we remember and what we remember: culture and autobiographical memory. *Perspectives on Psychological Science (Sage Publications Inc.)* 5:401–409.

Ross, Michael, and Anne E. Wilson. 2002. It feels like yesterday: self-esteem, valence of personal past experiences, and judgments of subjective distance. *Journal of Personality and Social Psychology* 82:792–803.

Rudman, Laurie A., and Kimberly Fairchild. 2004. Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology* 87:157–176.

Ryan, Richard M., and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55:68–78.

Ryan, Richard M., and Edward L. Deci. 2001. On happiness and human potential: a review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology* 52:141–166.

Ryff, Carol D. 1989. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology* 57:1069–1081.

Sagiv, Lilach, and Shalom H. Schwartz. 2000. Value priorities and subjective well-being: direct relations and congruity effects. *European Journal of Social Psychology* 30:177–198.

Schimmack, Ulrich, and Shigehiro Oishi. 2005. The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology* 89:395–406.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Schneider, Leann, and Ulrich Schimmack. 2010. Examining sources of self-informant agreement in life-satisfaction judgments. *Journal of Research in Personality* 44:207–212.

Schwartz, Shalom H. 1992. Universals in the content and structure of values: theory and empirical tests in 20 countries. In *Advances in experimental social psychology*, ed. M. Zanna, 1–65. New York: Academic Press.

Schwartz, Carolyn, Janice Bell Meisenhelder, Ma Yunsheng, and George Reed. 2003. Altruistic social interest behaviors are associated with better mental health. *Psychosomatic Medicine* 65:778–785.

Schwarz, Norbert. 1990. Feelings as information: Informational and motivational functions of affective states. In *Handbook of motivation and cognition*, ed. E. Tory Higgings, Richard M. Sorrentino, 527–561. New York: Guilford.

Schyns, Peggy. 2002. Wealth of nations, individual income and life satisfaction in 42 countries: a multilevel approach. *Social Indicators Research* 60:5–40.

Seligman, Martin. 2002. *Authentic happiness: using the new positive psychology to realize your potential for lasting fulfillment*. New York: Free Press.

Settanni, Michele, and Davide Marengo. 2015. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.01045

Sheldon, Kennon M., and T. Kasser. 1998. Pursuing personal goals: skills enable progress, but not all goal progress is beneficial. *Personality and Social Psychology Bulletin* 24:1319–1331.

Silva, Ana Daniela, Maria do Céu Taveira, Cátia Marques, and Valdiney V. Gouveia. 2015. Satisfaction with life scale among adolescents and young adults in Portugal: extending evidence of construct validity. *Social Indicators Research* 120:309–318.

Sortheix, Florencia M., and Shalom H. Schwartz. 2017. Values that underlie and undermine well-being: variability across countries. *European Journal of Personality* 31:187–201.

Specht, Jule, Boris Egloff, and Stefan C. Schmukle. 2013. Examining mechanisms of personality maturation: the impact of life satisfaction on the development of the big five personality traits. *Social Psychological and Personality Science* 4:181–189.

Stavrova, Olga. 2014. *Fitting in and getting happy: How conformity to societal norms affects subjective well-being*. Frankfurt a. M.: Campus.

Stavrova, Olga. 2015a. How regional personality affects individuals' life satisfaction: a case of emotional contagion? *Journal of Research in Personality* 58:1–5.

Stavrova, Olga. 2015b. Religion, self-rated health, and mortality. *Social Psychological and Personality Science* 6:911–922.

Stavrova, Olga. 2019. Having a happy spouse is associated with lowered risk of mortality. *Psychological Science* 25:2189–2198. https://doi.org/10.1177/0956797619835147.

Stavrova, Olga, and Detlef Fetchenhauer. 2015a. Married and cohabiting parents' well-being: the effects of a cultural normative context across countries. *Journal of Social and Personal Relationships* 32:601–632.

Stavrova, Olga, and Detlef Fetchenhauer. 2015b. Single parents, unhappy parents? Parenthood, partnership and the cultural normative context. *Journal of Cross-Cultural Psychology* 46:134–149.

Stavrova, Olga, and Maike Luhmann. 2016a. Are conservatives happier than liberals? Not always and not everywhere. *Journal of Research in Personality* 63:29–35.

Stavrova, Olga, and Maike Luhmann. 2016b. Social connectedness as a source and consequence of meaning in life. *The Journal of Positive Psychology* 11:470–479.

Stavrova, Olga, Daniel Ehlebracht, and Detlef Fetchenhauer. 2016. Belief in scientific-technological progress and life satisfaction: the role of personal control. *Personality and Individual Differences* 96:227–236.

Stavrova, Olga, Detlef Fetchenhauer and Thomas Schlösser. 2012. Cohabitation, gender, and happiness: a cross-cultural study in thirty countries. *Journal of Cross-Cultural Psychology* 43:1063–1081.

Stavrova, Olga, Detlef Fetchenhauer and Thomas Schlösser. 2013a. Why are religious people happy? The effect of the social norm of religiosity across countries. *Social Science Research* 42:90–105.

Stavrova, Olga, Thomas Schlösser, and Detlef Fetchenhauer. 2011. Are the unemployed equally unhappy all around the world? The role of the social norms to work and welfare state provision in 28 OECD countries. *Journal of Economic Psychology* 32:159–171.

Stavrova, Olga, Thomas Schlösser, and Detlef Fetchenhauer. 2013b. Are virtuous people happy all around the world? Civic virtue, antisocial punishment, and subjective well-being across cultures. *Personality and Social Psychology Bulletin* 39:927–942.

Steel, Piers, and Deniz S. Ones. 2002. Personality and happiness: a national-level analysis. *Journal of Personality and Social Psychology* 83:767–781.

Steel, Piers, Joseph Schmidt, and Jonas Shultz. 2008. Refining the relationship between personality and subjective well-being. *Psychological Bulletin* 134:138–161.

Steger, Michael. 2014. Is it time to consider meaning in life as a public policy priority? *Ewha Journal of Social Sciences*. https://ssrn.com/abstract=2580707

Stevenson, Betsey, and Justin Wolfers. 2009. The paradox of declining female happiness. *American Economic Journal: Economic Policy* 1:190–225.

Suh, Eunkook, Ed Diener, Shigehiro Oishi, and Harry C. Triandis. 1998. The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology* 74:482–493.

Suldo, Shannon, Devon Minch, and Brittany Hearon. 2015. Adolescent life satisfaction and personality characteristics: investigating relationships using a five factor model. *Journal of Happiness Studies* 16:965–983.

Tajfel, Henri, M.G. Billig, R.P. Bundy, and Claude Flament. 1971. Social categorization and intergroup behaviour. *European Journal of Social Psychology* 1:149–178.

Tay, Louis, Mitchel N. Herian, and Ed Diener. 2014. Detrimental effects of corruption and subjective well-being: Whether, how, and when. *Social Psychological and Personality Science* 5:751–759.

Tesch-Roemer, Clemens, Andreas Motel-Klingebiel, and Martin J. Tomasik. 2008. Gender differences in subjective well-being: comparing societies with respect to gender equality. *Social Indicators Research* 85:329–349.

Tokuda, Yasuharu, Seiji Fujii, and Takashi Inoguchi. 2010. Individual and country-level effects of social trust on happiness: the Asia Barometer Survey. *Journal of Applied Social Psychology* 40:2574–2593.

Triandis, H.C. 1995. *Individualism and collectivism*. Boulder: Westview.

Uchida, Yukiko, and Shinobu Kitayama. 2009. Happiness and unhappiness in east and west: themes and variations. *Emotion* 9:441–456.

Veenhoven, Ruut. 1984. *Conditions of happiness*. Dordrecht: Reidel Publ. Co.

Veenhoven, Ruut. 1993. *Happiness in nations: Subjective appreciation of life in 56 nations 1946–1992*. Rotterdam: Erasmus University.

Veenhoven, Ruut. 2008. Sociological theories of subjective well-being. In *The science of subjective well-being: a tribute to Ed Diener*, ed. Michael Eid, Randy Larsen, 44–61. New York: Guilford.

Veenhoven, Ruut. 2012. Cross-national differences in happiness: cultural measurement bias or effect of culture? *International Journal of Wellbeing* 2:333–353. https://doi.org/10.5502/ijw.v2.i4.4

Veenhoven, Ruut, and Yowon Choi. 2012. Does intelligence boost happiness? Smartness of all pays more than being smarter than others. *International Journal of Happiness and Development* 1:5–27.

Van de Velde, Sarah, Koen Van der Bracht, and Veerle Buffel. 2017. The relation between religion and depression in Europe: the moderating role of the religious context. *International Journal of Comparative Sociology* 58:515–532.

Voßemer, Jonas, Michael Gebel, Kadri Täht, Marge Unt, Björn Högberg, and Mattias Strandh. 2017. The effects of unemployment and insecure jobs on well-being and health: the moderating role of labor market policies. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1697-y

Wallace, John M., and Tyrone A. Forman. 1998. Religion's role in promoting health and reducing risk among American youth. *Health Education & Behavior* 25:721–741.

Watson, Deryn, A. Clark Lee, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS Scale. *Journal of Personality and Social Psychology* 54:1063–1070.

Weinstein, Neil D. 1980. Unrealistic optimism about future life events. *Journal of Personality & Social Psychology* 39:806–820.

Whisman, Mark A., and Charles M. Judd. 2016. A cross-national analysis of measurement invariance of the Satisfaction With Life Scale. *Psychological Assessment* 28:239–244.

Williams, Kipling D. 2007. Ostracism. *Annual Review of Psychology* 58:425–452.

Williams, Kristi, and Alexandra Dunne-Bryant. 2006. Divorce and adult psychological well-being: clarifying the role of gender and child age. *Journal of Marriage and Family* 68:1178–1196.

Williams, Kipling D., Christopher K.T. Cheung, and Wilma Choi. 2000. Cyberostracism: effects of being ignored over the Internet. *Journal of Personality and Social Psychology* 79:748–762.

Wilson, Warner R. 1967. Correlates of avowed happiness. *Psychological Bulletin* 67:294–306.

Wood, Wendy, P. Niels Christensen, Michelle R. Hebl, and Hank Rothgerber. 1997. Conformity to sex-typed norms, affect, and the self-concept. *Journal of Personality and Social Psychology* 73:523–535.

Yap, Stevie C.Y., Jessica Wortman, Ivana Anusic, S. Glenn Baker, Laura D. Scherer, M. Brent Donnellan, and Richard E. Lucas. 2016. The effect of mood on judgments of subjective well-being: nine tests of the judgment model. *Journal of Personality and Social Psychology* 113:939–961.

Olga Stavrova 1985, Dr. rer.pol., Assistant Professor at Tilburg University (The Netherlands). Areas of research: social psychology, personality psychology. Publications: Cynical beliefs about human nature and income: longitudinal and cross-cultural analyses. Journal of Personality and Social Psychology 110, 2015 (with D. Ehlebracht); Having a happy spouse is associated with lowered risk of mortality. Psychological Science 2019; Fitting in and getting happy. How conformity to societal norms affects subjective well-being. Frankfurt a. M. 2014.

Is the Influence of Religiosity on Attitudes and Behaviors Stronger in less Religious or more Religious Societies? A Review of Theories and Contradictory Evidence

Pascal Siegers

Abstract Cross-cultural studies of religion have shown that there is substantial variation in the effects of religiosity on moral and social outcomes across countries. Many authors have addressed the question of how the effects of individual religiosity depend on the religious context. The findings, however, are contradictory, casting doubts on the validity of results. The current paper reviews the existing research on how the religious context moderates the effects of individual religiosity. It reveals the limitations of existing studies, and these limitations might explain the contradictory findings. Most notably, authors tend to assume similar effects from all religions without discussing the possibility of the effects that they find being confounded with differences across denominations. Given the limited set of available comparative data, there is a high risk of selection bias because almost all the studies use data from the same limited number of international survey projects. Moreover, the operationalizations of religiosity and religious context do not sufficiently reflect the theoretical approaches. The paper gives recommendations to improve future research on how religious contexts shape the effects of individual religiosity.

Keywords Multilevel analysis · Religion · Moral communities · Intrinsic religiosity · Religious networks · Moderation

P. Siegers (✉)
GESIS Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6–8, 50667 Köln, Germany
E-Mail: pascal.siegers@gesis.org

🖄 Springer

Ist der Einfluss der Religiosität auf Einstellungen und Verhalten in säkularen oder religiösen Gesellschaften stärker? Ein Überblick über Theorien und widersprüchliche Ergebnisse

Zusammenfassung Die international vergleichende Religionsforschung hat eine erhebliche Variation in den Effekten von Religiosität auf moralische und soziale Einstellungen und Verhaltensweisen nachgewiesen. Viele Studien haben die Bedeutung der religiösen Kontexte für die Richtung und Stärke der Effekte von Religiosität auf verschiedene abhängige Variablen untersucht. Die Ergebnisse dieser Studien sind jedoch widersprüchlich und lassen deshalb Zweifel an ihrer Validität aufkommen. Dieser Beitrag gibt einen Überblick über theoretische Ansätze und empirische Befunde aus der Forschung dazu, wie sich religiöse Kontexte auf die Effekte individueller Religiosität auswirken. Dabei zeigt sich, dass bestehende Studien Schwachstellen aufweisen, welche für die widersprüchlichen empirischen Befunde verantwortlich sein könnten. Insbesondere gibt es in der komparativen Sozialforschung die Tendenz, allen Religionen ähnliche Effekte zu unterstellen, ohne zu prüfen, ob möglicherweise konfessionelle Unterschiede vorliegen. Zudem besteht angesichts der kleinen Zahl von verfügbaren international vergleichenden Umfrageprogrammen das Risiko eines „selection bias", weil fast alle Studien mit den wenigen gleichen Datensätzen arbeiten. Schließlich spiegeln die Operationalisierungen von Religiosität und religiösem Kontext die verwendeten theoretischen Ansätze häufig nicht ausreichend wider. Die Arbeit schließt mit Empfehlungen zur Verbesserung der zukünftigen Forschung darüber, wie religiöse Kontexte die Effekte individueller Religiosität beeinflussen.

Schlüsselwörter Mehrebenenanalyse · Religion · Moralische Gemeinschaften · Intrinsische Religiosität · Religiöse Netzwerke · Moderation

1 Introduction

Individual religiosity remains one of the most important predictors of individual attitudes and behaviors across various domains of the social sciences. The effects of religiosity were often considered to be universal and independent of their social and cultural context (Stark 2001). The growing amount of comparative data from large-scale cross-cultural survey programs—especially the World Values Survey and the European Values Study—challenged this assumption of universality. On the one hand, religious contexts in terms of denominational culture or aggregate religiosity have been studied extensively (Dülmer 2014; Norris and Inglehart 2004). On the other hand, comparative studies have revealed significant variation in the effects of religiosity across countries (see Stavrova 2019).

The availability of cross-national comparative data triggered the spread of multilevel analysis among social scientists for analyzing hierarchical data (for example, individuals nested in countries or students nested in schools). Multilevel regression allows us to study effects of contextual attributes on (1) individual-level outcomes (e.g., effects of religious contexts on moral attitudes or subjective health) and (2) the

effects of individual-level predictors (e.g., predicting variation in effects of religiosity). Thus, multilevel analysis has become an essential tool to go beyond macro correlations of aggregate measures of religion and religiosity. Multilevel analysis allows us to statistically test hypotheses about the relationship between religious contexts and the effects of individual religiosity.

These new regression techniques for cross-cultural comparative research have complemented comparative research based on macro-analytical approaches. The latter type of research focuses on how the religious context moderates the effects of individual religiosity for many different outcomes, such as attitudes about abortion (Adamczyk 2008), divorce (Finke and Adamczyk 2008), homosexuality (Adamczyk and Pitt 2009), euthanasia (Verbakel and Jaspers 2010), pro-social orientations and behaviors (Stavrova and Siegers 2014), outgroup prejudice (Doebler 2015b), social capital and trust (Ruiter and De Graaf 2006), life satisfaction (Stavrova et al. 2013), wellbeing (Hayward and Elliott 2014), and health (Huijts and Kraaykamp 2011).

All these studies examined the conditions for religiosity to be effective in influencing individuals' attitudes and behaviors. This approach is particularly important because studying the moderation of religiosity by contextual attributes is a method to identify the mechanisms underlying the individual-level associations between religiosity and social outcomes. Nevertheless, there is no systematic feedback from cross-cultural comparative studies that lends itself to use in theorizing about religion because the comparative papers focus mostly on particular outcomes. A systematic evaluation of the findings across different outcomes has not yet been published.

The current paper aims to fill this gap, explaining the specificity of multilevel models from the comparative study of religious contexts, and reviews the relevant results, focusing on the moderation of the effects of religiosity by religious contexts for different dependent variables. Thus, I try to evaluate whether the results allow for conclusions about the mechanism linking religiosity to social and moral outcomes. The limitation to religious contexts follows pragmatic considerations. There are insufficient studies considering other contextual factors that may moderate the effects of individual religiosity for a meaningful review (with the exception of Adamczyk and Pitt 2009; Luria et al. 2017).

The results of the review reveal that the existing evidence is contradictory. Improving our understanding of how religiosity and religious contexts shape individuals' attitudes and behaviors requires more appropriate use to be made of the available methods of quantitative comparative analysis. By pointing out the limitations of the existing research, the review helps to improve future research.

Some terminological clarifications are useful to facilitate reading of the paper. Throughout this paper, I use the term "religiosity" to address the attributes of individuals (i.e., the intensity of belief and practice), the term "contextual religiosity" to address aggregate religiosity at the contextual level (e.g., average church attendance), and the term "religious culture" to point to denominational and congregational differences in contexts.

The first section gives an overview of different comparative designs for studies of religious contexts and discusses the specificity of multilevel analysis. The second section summarizes the most important theoretical approaches to the moderation of religiosity by religious context. The third section summarizes the findings from

existing research for different groups of outcomes. The fourth section discusses the most important limitations of existing studies, and the paper concludes with recommendations on how to improve future research.
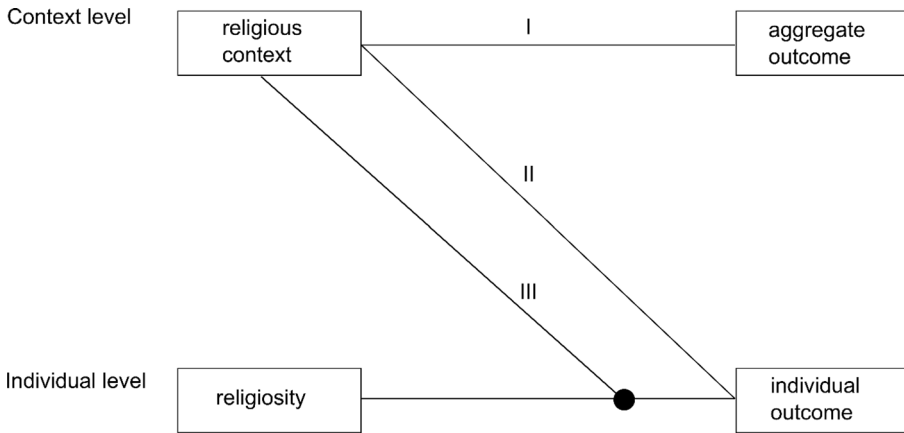
2 Comparative Designs for Studying Religious Contexts

Studies about how religious contexts influence social outcomes have a very long tradition in sociology. Figure 1 summarizes the most important approaches. The first line (I) refers to studies based on correlations between attributes of religious contexts and aggregate-level outcomes. The most prominent example of a macro-level study is Durkheim's seminal work about suicide, in which he demonstrated systematic variation in suicide rates across different religious communities, arguing that religious traditions that provide stronger social integration will reduce individual dispositions toward suicide (Durkheim 1897). Another example is the original article suggesting the moral communities hypothesis. Stark et al. (1980) correlated aggregate levels of religiosity with aggregate measures of deviance to show that religiosity was more effective when it came to prohibiting deviant behavior and crime in more religious contexts (see also Stark 1996). Macro-analytical approaches have been widely used in research about secularization because the quality of the religious context plays a key role in theories of religious change (Pickel 2010).[1] Secularization theory argues that heterogeneity undermines the validity of religious teachings by relativizing religious claims as to truth, thus paving the way for religious disaffiliation (Bruce 2006). The market model, in contrast, states that heterogeneity of religious supply creates competition between religious organizations, which increases the quality of the religious services provided to individuals, and ultimately increases religious participation (Stark and Finke 2000).

Other macro-analytical studies examined correlations between religious contexts and social outcomes at the context level. Using data from the World Values Surveys, Norris and Inglehart (2004) demonstrate that average religiosity decreased in most economically advanced societies, but that at the same time, lower contextual religiosity in a country is associated with lower average birth rates. The result is that although secularization is evident in many economically advanced countries, the overall number of religious people worldwide is on the increase because more babies are born in religious countries. Most of the work about value change by Inglehart and colleagues (for example Inglehart and Welzel 2005) assumes that a decrease in contextual religiosity is part and parcel of a shift from traditional (including religious) to secular-rational values. Inglehart argues that this change is the consequence of economic growth, and that the trajectories of change are path dependent and highly influenced by a society's prevalent religious culture (Inglehart and Baker 2000).

These macro-analytical studies have their limitations. The most important is the risk of an ecological fallacy when interpreting correlations between aggregate at-

---

[1] In fact, secularization theory is a framework for explaining religious change at the context level. This paper, in contrast, reviews studies about the social consequences of variation in religious context for individual-level outcomes.

Fig. 1 Overview of comparative designs for studying effects of religious contexts. Author's own work. I. Macro level correlation between context level attributes. II. Direct effect of religious context on individual level dependent variable. III Interaction effect between religious context and individual level religiosity (i. e., the moderation of effects of religiosity by religious context)

tributes (Gnaldi et al. 2018; Schmidt-Catran et al. 2019). Regarding the particular case of *religious* context, a major methodological problem was found that affects all studies that use the Herfindahl index for measuring religious heterogeneity. All correlations found between the context measures and religious outcomes (especially religious participation when measured as aggregate church attendance and all measures correlated with participation) might be a mathematical artifact and should be interpreted with caution (Voas et al. 2002). Also, macro analyses are not suited to studying individual-level outcomes such as moral attitudes or subjective health.

Until recently, it was not possible to model the link between context attributes and individual-level attributes. Some authors included attributes of context into standard ordinary least squares regression (Welch et al. 1991). The test statistics (e.g., t-values) for the contextual attributes are likely to be underestimated in this case because the number of cases for inferential testing is based on the individual-level units (i.e., the number of respondents in the survey), although the true number of cases is the number of context units included in the analysis (e.g., the number of countries available in a dataset). This leads to hypotheses being confirmed incorrectly (Hox 2002).

The advent of multilevel regression has enabled researchers to test hypotheses about the relationship between religious contexts and dependent variables at the individual level (such as health or social and moral attitudes; see line II in Fig. 1). Furthermore, multilevel regression can easily be extended to the explanation of the variation of effects across contexts. The third line (III) in Fig. 1 represents this relationship. The line at the bottom of Fig. 1 represents the effect of religiosity on attitudes or behaviors. This relationship potentially varies across contexts and might systematically depend on the religious context. The filled circle in Fig. 1 represents the moderation of the individual-level effect by the religious contexts. In more technical terms, it is a cross-level interaction effect between the variables

measuring individual religiosity and religious context.[2] Studying this moderation is relevant to research about religion because it makes it possible to unveil the mechanisms underlying the correlations found at the micro level (i.e., the line at the bottom of Fig. 1) by identifying the contextual conditions for religiosity to influence attitudes and behavior.[3] Most recent comparative studies about religious context use a fully fledged multilevel design, studying the direct effects from contexts on the individual-level outcomes (line II), and the moderation of the individual-level effects by contextual variables (line III). The remainder of this paper will therefore focus on a review of recent studies using multilevel regressions.

3 Five Theoretical Approaches: Explaining the Moderation of Individual Religiosity by Religious Contexts

Reviewing the existing research reveals that there are five theoretical frameworks used for explaining cross-cultural differences in the effects of religiosity. For each of the five theoretical approaches, we summarize the basic theoretical assumptions and the expected moderation of the religious context.

3.1 Moral Communities and Secular Temptation

Most studies of how the religious context influences the effects of individual religiosity are based on the "moral community" thesis that was introduced in the early 1980s in research about crime and deviant behavior (Stark et al. 1980; Tittle and Welch 1983). The theory is based on Durkheim's definition of religion as a "moral community" (Durkheim 1968 [1912]).[4] For Durkheim, religion serves to maintain social cohesion by creating and sustaining a frame of shared normative orientations. The most important factor for integration into the "moral community" is ritual participation. During the ritual, individuals experience transcendental reality, which is the sanctified symbol of the society itself (the "totem"). To be part of the community, individuals must adhere to the religious beliefs and to the social and moral rules derived from them. This means that individual religiosity should foster conformity with religious rules (Welch et al. 1991). From a Durkheimian perspective, religion is the most important condition for social integration and social cohesion.

It therefore comes as no surprise that scholars of deviant behavior (re-)introduced the moral community thesis into social research (Stark 1996). The starting point for this research was ambivalent evidence about the negative effects of individual religiosity on deviant behavior in the United States. Whereas some studies

---

[2] An alternative method for estimating the interaction between context factors and individual-level attributes are fixed effects models using dummy variables for the context units. The effect of the context attitude on the individual-level outcome cannot be estimated in this case. But interaction effects (macro*micro variable) can be tested. This shortcoming might explain why fixed effects models have not been used in studies of religious context.

[3] The approach is naturally not limited to religious context variables. Other contextual moderators for individual religiosity can also be studied.

[4] Durkheim uses the word "moral community" as a synonym for "church."

reported effects, others did not. This contradictory evidence raised questions about the conditions for individual religiosity to inhibit deviant and criminal behavior. The assumption was that individual religiosity needs a supportive religious context to be effective at fostering conformity because strong communities can more easily sustain shared behavioral norms. In less religious contexts, churches do not have the same capacity to enforce the acceptance of behavioral norms, and individual religiosity consequently has a weaker influence on behavior (Stark 1996). The mechanism relating individual religiosity to attitudes and behaviors is social control and sanctioning of non-conformity.

From the perspective of secularization theory, the moral community thesis was (re-)interpreted as "secular temptation" (Scheepers et al. 2002). During the process of secularization, churches' social control of moral and social attitudes weakens, and the moral community dissolves. Conformity with religious norms becomes more difficult to enforce, and even religious individuals exhibit lower levels of norm compliance. They succumb to the "secular temptation" to live without strict behavioral or moral rules. The conclusion is the same as for the moral community thesis. The differences between religious and non-religious individuals are less pronounced in less religious societies.

If Durkheim's theory about social integration through religious participation is taken seriously, it requires an operationalization in terms of collective religious practice. For example, exposure to the preaching of the clergy has been shown to be an important condition for religious norms to be effective in individuals' lives (Spenkuch and Tillmann 2018).

3.2 The Detrimental Effect of "Religious Heterogeneity"

The moral communities thesis focuses on the contextual level of religiosity (i.e., the average level of religious participation within a social group), and does not address the question of how homogeneous (i.e., shared among the members of the society) religious identities and beliefs are. In many countries, however, different religious traditions coexist, and individuals have diverging religious identities. Since the Reformation, Europe has been divided between Protestant and Catholic Churches, and North America has been characterized by a patchwork of different Protestant congregations, Catholics, and many other religious traditions.

Differing somewhat from the moral community thesis, the religious heterogeneity thesis argues that homogeneity of the religious context is a prerequisite to foster conformity with religious norms (Olson and Li 2015). Religious heterogeneity, in contrast, is expected to undermine the effects of individual religiosity when religious norms and teachings of different religious traditions contradict each other or when religious identities are strongly shaped by conflicts between religious groups (Taylor 2007). In such situations, religious identities are instrumentalized for social bonding (i.e., closing the group to external people and influences) and foster social differentiation instead of social integration because conflicts between congregations reduce trust among members of the society and reduce capacity for social cooperation. This mechanism of conflict is detrimental to effects of religiosity otherwise strengthened

by supportive religious context, such as the positive effects on pro-social behavior and social capital.

The moral community thesis and the religious heterogeneity thesis can be combined into a single theoretical framework. Religious homogeneity is then a precondition for moral communities to be effective, because some degree of homogeneity or consistency in religious teachings might be a necessary condition for the validity of the moral community hypothesis (Olson and Li 2015).

3.3 Religious Networks and Social Support

A third theoretical approach does not stress the importance of religious teachings and rules for influencing individual attitudes and behaviors, but rather emphasizes the role that contacts among religious people play for social outcomes. The key assumption is that in societies with higher average levels of worship attendance, religious people are more likely to have contacts with other religious people. Within these religious networks, individuals can recruit each other into activities and mutually reinforce specific attitudes even if these are not religious in nature (e.g., participation in sports clubs or voting behavior; Ruiter and De Graaf 2006) and obtain social support when facing difficult situations (e.g., grief, sickness, or unemployment; Pargament et al. 1998).

This contact mechanism is more effective when religious individuals have more opportunities for contact with co-religionists because with higher average levels of religious participation (but not belief), religious networks are larger and the effects of individual service attendance are reinforced. This rationale for justifying the influence of context on the individual-level effect of religiosity is therefore particularly prominent in studies about volunteering (Ruiter and De Graaf 2006).

The difference from the moral community hypothesis might be subtle, but it is important. The moral community thesis stresses the importance of exposure to religious rules that influence individuals' attitudes and behaviors. The religious networks thesis underscores the importance of contacts with fellow believers in congregations and communities. These produce specific outcomes that are not necessarily linked to religious teachings.

The interrelatedness between individual religious participation and the contextual level of religious participation is clearly stated for the religious networks thesis. No one can create and maintain social ties when attending services in an empty church or praying alone at home.

3.4 Religious Defense in Secular Societies

In contrast to the three approaches presented so far, two other theories expect effects of individual religiosity to be stronger in more secular religious contexts. The first is the "religious defense" thesis and the second is the "intrinsic religiosity" thesis.

Religious defense refers to situations in secular societies in which not only religiosity has become a matter of individual choice (Taylor 2007), but religious individuals also form a minority. This is the prevalent situation in many Western European societies, especially countries such as the Czech Republic, the Nether-

lands, the United Kingdom, and Eastern Germany, and especially in urban milieus (Pickel 2010; Pollack and Rosta 2015). Under such conditions, the choice to be religious is not a default option for individuals, but rather requires explicit choices and justification in everyday life (e.g., in the workplace or at school). Being religious, then, involves greater individual effort in the absence of a supportive religious context.

As a consequence, religious individuals in secular contexts will invest more effort in sustaining their religiosity (i.e., by engaging in a particularly intense religious practice) and in the religious upbringing of their offspring (Kelly and De Graaf 1997; Scheepers et al. 2002). This results in stronger bonding within religious communities. Conformity with religious rules becomes an important marker of religious identities, and thus for drawing the boundaries between the religious community and the secular (or religious) other. From this perspective, religious communities succeed even better in maintaining their cohesion if there is no supportive context. Religious defense assumes the opposite of the moral communities hypothesis: Individual religiosity has a stronger effect on attitudes and behavior in less religious contexts.

3.5 Intrinsic vs. Extrinsic Religiosity

Another approach arguing that a secular context strengthens the effects of individual religiosity on attitudes and behaviors assumes that the religious context shapes motivations for religious beliefs. This starts from the assumption that in highly religious societies, being religious is, de facto, a social norm (Taylor 2007). Non-religious individuals are perceived as deviant and therefore less trustworthy, likable, etc. In some countries, basic social and political rights are even conditional on belonging to a religious denomination. In such situations, non-belief is equivalent to incomplete social integration of individuals, with negative effects on social support and life satisfaction (Stavrova et al. 2013). In the United States, for example, non-religiosity is frequently associated with amorality (Zuckerman 2012). As a result, some individuals will also be religious for utilitarian considerations, such as to avoid negative sanctions for non-belief. Although they participate in religious rituals and report belief in God, religious beliefs and teachings are not the primary source of guidance when forming attitudes or choosing behaviors. Such utilitarian motivations for religiosity have been called "extrinsic religiosity" and have been distinguished from "intrinsic religiosity," in which the religious belief is central to an individual's life (Allport and Ross 1967). Extrinsic religiosity is present if an individual's religiosity is a means for the satisfaction of personal and social needs that are not necessarily related to religion, such as social support and recognition or individual comfort. Beliefs are not deeply anchored in the truth of the religious teachings, but are rather based on instrumental evaluation. Therefore, Allport and Ross (1967) did not expect strong obedience to religious norms based on extrinsically motivated religiosity. Intrinsic religiosity, in contrast, describes religiosity in which transcendental beliefs are central to individuals' lives. In the Christian context, such a belief would be the relationship that individuals have with the personal God. Intrinsically motivated individuals obtain meaning and guidance for their lives from religion and

will therefore conform to religious teachings because they are convinced of their truth.

Intrinsic and extrinsic motives are not exclusive and can be present simultaneously. The relative importance of intrinsic and extrinsic religious motivation changes depending on the level of contextual religiosity. When religiosity loses its status as a normative requirement for social integration, extrinsic motivations for religiosity become less important because non-religiosity is not sanctioned and more individuals decide to be non-religious (Saroglou 2011; Stavrova and Siegers 2014).

Following this approach, secularization has twofold effects on the association between individual religiosity and social and moral outcomes. On the one hand, the decline in the churches' social significance results in a more permissive and liberal moral climate (Halman and Van Ingen 2015). On the other hand, individuals remaining religious during the process of secularization are more likely to be intrinsically religious, and therefore also more likely to conform to religious norms. Stavrova and Siegers (2014) demonstrated that the correlation between measures of general religiosity and indicators of intrinsic religiosity is stronger in countries where being religious is not a social norm. This means that the difference between religious and non-religious individuals in terms of social outcomes (i.e., the effect of individual religiosity) will be larger in more secularized societies.

The theoretical rationale of the intrinsic religiosity hypothesis focuses on the centrality of religious beliefs for individuals (Huber 2007), rather than on exposure to religious teachings or contacts with fellow citizens. Measurements of religiosity for testing this hypothesis should use an operationalization in terms of core religious beliefs (e.g., belief in a personal God) and the relevance of religion for individuals' lives. Similarly, the religious context is best measured as a norm for being religious, not the average religiosity (Stavrova and Siegers 2014).

Figure 2 summarizes the expected pattern for the main effects of religious contexts (line II in Fig. 1) and the moderation of individual religiosity by religious context (line III in Fig. 1) for four different theoretical approaches presented in this section (the moral community hypothesis subsumes the religious heterogeneity hypothesis).

The visualization of the expected effects (Fig. 2) is based on a multilevel equation shown in Eq. 1 using the example of moral attitudes as the dependent variable (Snijders and Bosker 1999):[5]

$$\widehat{MORAL}_{ij} = \gamma_{00} + \gamma_{10}\text{IREL}_{ij} + \gamma_{01}\text{CREL}_j + \gamma_{11}\text{IREL}_{ij}\text{CREL}_j \qquad (1)$$

In this equation, the predicted moral attitudes ($\widehat{MORAL}_{ij}$) depicted on the y-axis are a function of individual religiosity (IREL_{ij}) depicted on the x-axis and contextual religiosity (CREL_j). The subscript ij refers to individual i nested in country j. γ_{00} denotes the intercept; this means the intersection of the regression line with the y-axis at the mean of $CREL$, which is the solid line. γ_{10} denotes the effect of individual religiosity on the moral attitudes at the mean of $CREL$ (i.e., the solid lines in Fig. 2) and γ_{01} is the main effect of contextual religiosity on the

---

[5] The equation is only valid for predicted values. The variance components have to be added to the equation for observed values of the outcome.
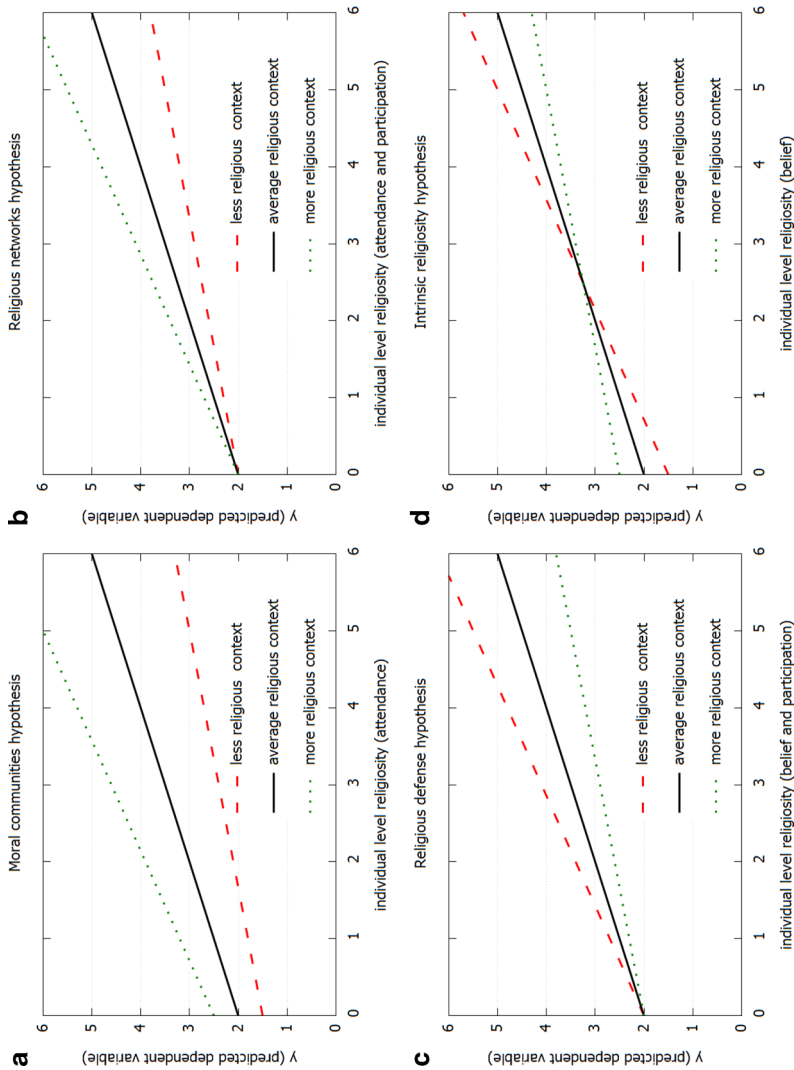
Fig. 2 Overview of expected moderation effects from theoretical approaches. Author's own work. The panels show the expected pattern in effects of individual religiosity depending on religious context for (**a**) the moral communities hypothesis, (**b**) the religious networks hypothesis, (**c**) the religious defense hypothesis, and (**d**) the intrinsic religiosity hypothesis

outcome variable. The main effect is visible in Fig. 2 in the distances between the intersections of the dotted, bold, and dashed lines with the y-axis. Finally, γ_{11} is the cross-level interaction (i.e., the difference in steepness between the dotted, the solid, and the dashed lines). The functions shown in Fig. 2 assume a positive effect of individual religiosity on the outcome variable (with x-axis values between 0 = non-religious and 6 = very religious). The lines represent three different types of religious contexts: The solid line is a context with average religiosity, the dotted line represents above-average religious contexts, and the dashed line corresponds to below-average religious contexts.

The differences in the intersections of the lines with the y-axis show the main effect of contextual religiosity. If the dotted line is above the solid line, the effect is positive, whereas if the dashed line is above the solid line, the main effect is negative. The steepness of the solid line shows the effect of individual religiosity at the average level of country religiosity (i.e., the average difference between religious and non-religious individuals). The steepness of the dotted and dashed lines represents the effect for more and less religious contexts, respectively.

Figure 2a shows that the moral community hypothesis assumes a positive main effect and a positive interaction. This means that countries that are more religious are characterized by higher levels of the outcome (i.e., moral conservativism), and the effect of individual religiosity on the outcome is stronger in more religious countries. This means that religious individuals are particularly conservative (compared to the non-religious) in more religious countries, because the mechanism of social sanctions of immoral behaviors (in the perception of the churches) is more effective.

The religious networks hypothesis (Fig. 2b) does not assume a specific main effect of religious context (but a main effect might exist). The expected interaction is positive because there are more contact opportunities for religious people in more religious contexts. The dotted line for the more religious contexts is steeper than the other lines because religious networks are larger where more people attend worship.

The religious defense hypothesis (Fig. 2c) assumes a negative interaction between religious contexts and individual religiosity. The dashed line is steeper than the dotted line. The difference between non-religious and religious people is greater in less religious countries because religious individuals follow religious values more closely in less religious contexts. The figure does not assume a main effect, but there could be one without contradicting the theoretical assumptions.

Finally, the intrinsic religiosity hypothesis expects a positive main effect combined with a negative interaction (Fig. 2d). This means that the outcome (e.g., moral conservativism) is more widespread in more religious countries, but the difference between religious and non-religious people is greater in less religious contexts. The reason for this is that extrinsic motivations for being religious lose importance in more secular societies, and those individuals who remain religious will conform to religious teachings. The intrinsic religiosity hypothesis and the moral communities hypothesis expect identical main effects but opposite directions for the moderation.

4 Contradictory Evidence from Existing Studies

The review of results is restricted to dependent variables for which at least two independent studies including cross-level interactions between individual religiosity and religious contexts have been published. Three research topics generated sufficient output for a meaningful review: (1) moral attitudes, (2) volunteering and trust, and (3) health and wellbeing.

As mentioned in the introduction, the objective is not a systematic meta-analysis, but rather a review with a focus on how the religious contexts interact with individual religiosity when influencing attitudes, behavior, and wellbeing.

4.1 Moral Attitudes

Religiosity is the most prominent predictor of moral attitudes in social research. The basic assumption is that religions prohibit various behaviors, including abortion, divorce, homosexuality, suicide, euthanasia, violence, crime, and substance use, and promote pro-social orientations and behaviors (Dülmer 2014; see Halman and Gelissen 2019). Religious individuals are expected to conform to the rules of their religious communities.

Therefore, it is not surprising that many studies have addressed the question of whether the religious context moderates the relationship between religiosity and moral attitudes. Although not all studies use the Durkheimian terminology, the *moral community hypothesis* has frequently been used as a theoretical rationale to justify hypotheses about the cross-country variation in effect sizes of religiosity. Even if these studies vary greatly in terms of the operationalization of religiosity, the operationalization of religious context, the sample of countries included in the comparative designs, and the specification of the multilevel models, there is evidence in support of the assertion that a supportive religious context reinforces religions' ability to enforce conformity with religious norms.

Some authors use indices for moral attitudes as outcomes (Adamczyk and Pitt 2009; Finke and Adamczyk 2008; Storm 2016), whereas other studies focus on specific attitudes such as those regarding homosexuality (Adamczyk et al. 2016; Adamczyk and Pitt 2009; Doebler 2015a), euthanasia (Verbakel and Jaspers 2010),[6] or suicide (Boyd and Chung 2012; Stack and Kposowa 2011), but all the studies report that the association of religiosity with more restrictive moral attitudes is stronger in more religious countries. Regarding the main effects, the studies show that moral permissiveness is lower in more religious contexts.

The operationalization of religiosity and religious context in these studies does not, however, correspond to the theoretical framework. The Durkheimian approach would require focusing on religious participation, but in most cases, composite scores for religiosity or an indicator of the importance of religion in respondents' lives that does not measure ritual practice are used. Finke and Adamczyk (2008), for example, do not find a cross-level interaction between religious context and church attendance, but rather between religious context and the importance of religion in life. They argue that this is due to a lack of measurement invariance for the attendance question in surveys covering countries with varying religious cultures.

Results contradicting the moral communities thesis have also been published. Based on the intrinsic religiosity thesis, Stavrova and Siegers (2014) show that the effect of religiosity on moral attitudes is stronger in countries in which social enforcement of religion is stronger (i.e., in contexts that are more religious on average). The authors argue that religiosity is related to moral attitudes only if belief is a matter of free choice, and not—as the moral communities thesis suggests—if a supportive religious context fosters conformity with religious rules. It is noteworthy

---

[6] Verbakel and Jaspers (2010) do not expect the effect of religiosity to be stronger in religious countries, but rather in secular countries. The results are nevertheless consistent with the moral community thesis.

that they partially use the same database and find main effects that are very similar to those found in most of the moral community studies reported above.

Finally, two studies found that the effect of religiosity on moral attitudes is stronger in countries with a self-expressionist culture compared to survivalist cultures (Adamczyk and Pitt 2009; Boyd and Chung 2012). They argue that in self-expressionist cultures, the restrictive teachings of the churches compete with more permissive secular authorities, thus reducing the churches' moral authority. Only religious individuals continue to follow religious rules in more self-expressionist contexts. The finding is relevant for this review because self-expressionist culture is negatively correlated with contextual religiosity (Inglehart 2006). This result supports the religious defense hypothesis. However, the measurement of self-expression values includes attitudes about moral issues (especially attitudes toward homosexuality), and thus there is a risk of confounding the measurement and the dependent variable when using self-expression values to predict moral attitudes.

The puzzle with regard to these contradictory findings is that the data used for all of the studies stem from the same two large survey programs (i.e., the European Values Study and the World Values Survey). Given that the measurement of individual religiosity is very similar in all studies, the different directions of the cross-level interactions result either from different operationalizations of the context or from different selections of country samples (see below).

4.2 Social Capital: Volunteering and Trust

The relationship between religiosity and social capital is a classical topic of social science research. Cross-cultural research stresses the importance of the religious networks hypothesis to justify the positive association between church attendance and volunteering. The basic assumption is that religious people recruit other religious people into volunteering when they meet them at religious services or other church-related activities. The most influential study was published by Ruiter and De Graaf (2006) a decade ago using data from early waves of the World Values Survey. They find a positive main effect of contextual religiosity for general volunteering, but not for membership in civil society associations. Using regional data from the US, Oarga et al. (2015) do not, however, find evidence of a contextual effect of religion on volunteering. Regarding the moderation, Ruiter and De Graaf (2006) report a negative cross-level interaction, indicating that the effects of religiosity are weaker in more religious countries, which is at odds with the religious network hypothesis. The authors argue that in more religious countries, both religious and non-religious people volunteer more, such that there is no additional benefit from recruiting in churches, confirming the results reported in different studies about religion and volunteering (Oarga et al. 2015; Stavrova and Siegers 2014). All in all, there is conclusive evidence that the positive effect of religiosity on volunteering is stronger in less religious contexts. This partly confirms the intrinsic religiosity hypothesis, although there is no main effect of religious context on volunteering.

For interpersonal trust—another important component of social capital—the results of existing studies are mostly inconclusive. Some studies have found that Protestant countries have higher levels of trust, although the differences between

Protestants, Catholics, and non-religious people are mostly small at the individual level (Dingemans and Van Ingen 2015; Traunmüller 2011). More important is the discussion about the question of whether religious heterogeneity undermines trust in a society. In a recent study, Olson and Li (2015) showed that trust is lower in countries with high contextual religiosity and high religious heterogeneity, whereas neither of the indicators influences trust separately. This finding confirms assumptions of the religious heterogeneity hypothesis. Using regional data from Germany, Traunmüller (2011) did not find any contextual effects. In contrast, Dingemans and Van Ingen (2015) found that higher religious heterogeneity increases trust. The analysis also revealed that religious heterogeneity reduces the detrimental effect of belief in God on trust, thus refuting the religious heterogeneity hypothesis.

Taken as a whole, studies about how religious contexts and individual religiosity influence interpersonal trust find that Protestant cultures exhibit higher trust. The evidence regarding heterogeneity is, however, inconclusive. A potential reason for this is that the studies include very different selections of countries in the analysis, which might lead to specific forms of selection bias (see the section about limitations).

4.3 (Subjective) Health and Wellbeing

Many studies from the sociology of health report evidence that religious participation is associated with better health outcomes (Nicholson et al. 2009; Weaver et al. 2006). Two broader mechanisms underlying this relationship are discussed in the literature. On the one hand, religiosity is assumed to suppress health-deteriorating behaviors such as drinking and smoking. Behaviors then mediate the effect of individual religiosity, and any direct effect of religiosity on health should disappear given an appropriate model specification (Headey et al. 2014). On the other hand, religiosity is considered a resource to cope with stressors and enhance individuals' coping abilities. Such resources are social support from religious communities and psychological support from transcendental beliefs. For example, most monotheistic religions include a dual structure of pastoral care combined with a concept of other-worldly salvation. This structure gives individuals a sense of meaning when facing health problems or other stressful situations.

The theoretical justification for the moderation of individual-level effects by religious contexts is based on the moral community thesis, which argues that the benefits of religiosity for health will be stronger in more religious contexts (Huijts and Kraaykamp 2011; Stavrova 2015).[7] Two studies confirm this expectation. The first is based on data from the World Values Survey covering all parts of the world and religious traditions. Whereas the correlation between religiosity and health is negative in countries with low average religiosity, it is positive in countries that are more religious (Stavrova 2015). The second study uses data from the US General Social Survey that are linked to the US National Death Index and analyzes regional differences in religious context. This study confirms the finding from the cross-cul-

---

[7] Stavrova (2015) uses a somewhat different terminology, referring to cultural fit to denote that religious individuals conform to the cultural context, which is equivalent to the moral community thesis.

tural comparison that individual religiosity reduces mortality more in more religious regions (Stavrova 2015).

Similar results are also reported in a study concerned with subjective health status and happiness using data from the World Values Study. Religiosity has a positive effect on health and life satisfaction when the religious context is supportive (Hayward and Elliott 2014). Conceptualizing religious context in terms of weak or strong social norms, and thus stressing conformity more than social support in highly religious societies, Stavrova et al. (2013) find that the positive effect of religiosity on happiness is stronger in highly religious countries. These results support the moral community thesis for health and wellbeing as outcomes. The study by Hayward and Elliott (2014) also shows that the positive effects of religiosity become negative in countries that have low average religiosity and where there are state restrictions on religious freedom. The religious composition of society interacts with social religious norms and legal regulations concerning religions in conditioning the impact of religiosity.

Nevertheless, two studies using data from the European Social Survey challenge the generalization of these results. Although they also find significant variation in the effects of religiosity on subjective health across European countries, they do not confirm that the effects of religiosity depend on the average religious participation of the country context (Huijts and Kraaykamp 2011; Nicholson et al. 2009). Regarding health outcomes, the moral community thesis is not confirmed for a sample of European countries.

Religious influences on other outcomes were also analyzed from a comparative perspective; however, very few studies are available, such that a review would not be meaningful. One of these outcomes is *deviant behavior*—the topic that was at the origin of the moral communities thesis, but only a small number of studies have been published recently and these are barely comparable (Stavrova and Siegers 2014). Other studies have addressed *outgroup prejudice* (Doebler 2015b) and *voting behavior* (Goldberg 2014).

Table 1 summarizes the findings from the review. The top lines of the table summarize the expected empirical pattern and the underlying social mechanism. The only unambiguous result from this review is that no study reports evidence supporting the religious networks hypothesis. Nevertheless, it would be premature to discard the theory. It has only been tested for social capital. From a theoretical point of view, the religious network hypothesis could also be tested with deviant behavior and health as outcomes, because social support from religious communities could strengthen individual resilience.

Regarding all other theories, the results are contradictory when comparing the different studies. The situation is particularly puzzling when it comes to moral attitudes. Evidence has been published supporting the moral communities hypothesis and the intrinsic religiosity hypothesis. They contradict each other regarding the expected pattern for the cross-level interactions (Fig. 2). The studies are based on the same datasets (World Values Survey/European Values Study) and use similar operationalization of religiosity and contextual religiosity. Deciding which of the theories produces the better result when it comes to explaining social outcomes requires a more rigorous empirical comparison of the two theories using operationalizations

Table 1 Overview of the results of the review by theoretical approaches and individual-level outcomes

	Moral communities	Religious heterogeneity	Religious networks	Religious defense	Intrinsic religiosity
Expected pattern of moderation	Stronger effect of religiosity in more religious contexts	Stronger effect in more religiously homogeneous contexts	Stronger effect of religiosity in more religious contexts	Stronger effect of religiosity in less religious contexts	Stronger effect of religiosity in less religious contexts
Social mechanism	Supportive religious contexts increase the capacity of religious communities to enforce religious norms	Competing religious communities undermine the effectiveness of the respective teachings	Contacts among fellow believers in parishes and congregations	Increasing investment in maintaining religious identities in non-religious societies	Intrinsic religiosity is more important in less religious societies for guiding individuals' choices

Table 1 (Continued)

	Moral communities	Religious heterogeneity	Religious networks	Religious defense	Intrinsic religiosity
Outcomes					
Moral attitudes	Main effect: Higher contextual religiosity is associated with less permissive moral attitudes (Finke and Adamczyk 2008; Verbakel and Jaspers 2010; Stack and Kposowa 2011; Doebler 2015a; Adamczyk et al. 2016)	–	–	Main effect: More self-expressive cultures show higher permissiveness (Adamczyk and Pitt 2009; Boyd and Chung 2012)	Main effect: Stronger norm for being religious is associated with less permissive moral attitudes (Stavrova and Siegers 2014)
	Interaction: Stronger associations of religiosity with moral attitudes in supportive religious contexts (Finke and Adamczyk 2008; Verbakel and Jaspers 2010; Stack and Kposowa 2011; Doebler 2015a; Adamczyk et al. 2016)	–	–	Interaction: Effect of religiosity is stronger in more self-expressionist contexts (Adamczyk and Pitt 2009; Boyd and Chung 2012)	Interaction: Weaker associations with moral attitudes in countries with a social norm for being religious (Stavrova and Siegers 2014)

Table 1 (Continued)

	Moral communities	Religious heterogeneity	Religious networks	Religious defense	Intrinsic religiosity
Social capital	–	Main effect: Contradictory evidence on the role of heterogeneity for civic engagement and trust (Traunmüller 2011; Dingemans and Van Ingen 2015; Olson and Li 2015)	Main effect: Contradictory evidence on the association of contextual religiosity with volunteering (Ruiter and De Graaf 2006; Oarga et al. 2015)	–	Main effect: Country-level norm of being religious is not associated with volunteering (Stavrova and Siegers 2014)
	–	–	Interaction: Effect of religiosity weaker in more religious contexts contradicting the hypothesis (Ruiter and De Graaf 2006; Oarga et al. 2015)	–	Interaction: Weaker associations with volunteering in countries with a social norm for being religious (Stavrova and Siegers 2014)
Health and well-being	Inconclusive evidence. Religiosity more beneficial for health outcomes in supportive religious contexts (Hayward and Elliott 2014; Stavrova 2015) but not in European sample of countries (Nicholson et al. 2009; Huijts and Kraaykamp 2011)	–	–	–	–

– No study from this perspective available. Author's own work.

that are based on the theoretically relevant dimensions of religiosity (e.g., participation vs. belief). Supporting evidence was published for all three outcomes reviewed for the intrinsic religiosity hypothesis, but only one paper is available, so that more tests are needed to evaluate the validity of the theory.

For social capital and health/wellbeing, the evidence reported from other theoretical perspectives is contradictory, and does not support any of the theories.

Evaluating and comparing the overall validity of the theories and the mechanisms they assume to cause the correlation between religiosity and social outcomes requires more rigorous tests of the theoretical claims across different outcomes. The review shows that scholars apply the theories to outcomes with contentual similarity to the theory (e.g., moral communities explaining moral attitudes, religious networks explaining recruitment, etc.). From a theoretical point of view, there are good reasons to apply the theories to larger numbers of outcomes if the underlying mechanisms are related to the outcomes. As mentioned above, cross-cultural differences in the density of religious networks could explain differences in the strength of social support that individuals gain from religious participation. Before running new studies for testing the theories, however, scholars should overcome limitations of existing research that might produce the contradictory finding revealed in this review.

5 Limitations of the Existing Research

The most important conclusion from the review is that the associations between religiosity and social outcomes are not universal. Whether religiosity influences moral attitudes, pro-social orientations, social capital, subjective wellbeing, and health depends on the attributes of the religious context.

The advent of multilevel analysis stimulated research on how religious context interacts with individual religiosity in influencing attitudes and behavior. However, the existing research yields contradictory findings, thus precluding unambiguous conclusions about the validity of the theoretical frameworks used in comparative analyses. Thus, doubts about the mechanisms underlying the correlations between religiosity and attitudes and behavior persist. Several limitations of the existing research might explain why the results systematically contradict each other.

A recurrent limitation of the studies reviewed above is that *effects are tested and interpreted without a theoretical rationale*. This limitation concerns above all the interaction effects between religious contexts and individual religiosity that are reported as byproducts of studies primarily concerned with the direct effects of the contextual variables (Traunmüller 2011). In some cases, the authors test all possible interactions (Doebler 2015a) or suggest hypotheses for some indicators of religiosity but not for others, without justifying the selection theoretically (Dingemans and Van Ingen 2015). Another limitation is that the *operationalization of religiosity and religious contexts is not always based on the assumptions of the theory that is being tested*. Of course, cross-national survey programs do not include a wide choice of indicators covering all dimensions of religiosity (with the notable exception of the European Values Study). Indicators such as the "importance of religion for the respondent's life," which has been used in several studies, might be commonly

available in comparative survey programs. However, only a few studies discuss whether the question is an appropriate measure of religious beliefs or practice. In particular, the moral communities hypothesis should be tested using indicators for ritual participation, such as worship attendance or equivalent indicators. Church attendance is available for almost all social surveys, but has been criticized for its lack of conceptual equivalence across countries (Lüchau 2007).

Regarding religiosity, operationalizations of religious contexts should be tied up with theory. The religious networks thesis, for example, argues that individuals attending church have more chances to meet other religious people if attendance is higher on average. Measuring context religiosity using aggregate values of the "importance of God" does not therefore cover the theoretical content. Moreover, the moral community thesis assumes a supportive social context for religiosity to influence individuals' attitudes and behaviors. This cannot be appropriately measured using indicators of very private beliefs (e.g., self-assessed religiosity).

The theoretical approaches presented in the papers generally assume *strong homogeneity across the teachings of different religions*. They expect more or less similar effects for all religious individuals, independently of the denomination or congregation to which they belong. Similar effects at the individual level can only be expected in the case of convergent religious teachings and rules for all the religions represented in the samples. Studies based on large-scale comparative survey programs cover a wide range of different religious cultures, and assume that the teachings are the same across all religions, such as the promotion of pro-social attitudes and behaviors (Stavrova and Siegers 2014) or restrictive moral attitudes (Finke and Adamczyk 2008).

Differences in the individual-level effects of religiosity might, however, also depend on differences among religious traditions. Almost two decades ago, Stark (2001) showed that the link between religiosity and moral attitudes is conditional on a specific image of God, i.e., a "conscious, powerful, morally concerned being" (Stark 2001). For Eastern religions, Stark did not find an association between religiosity and moral attitudes because, he argues, they lack a transcendent power concerned with morality that would be comparable to the God of the monotheistic religions. Even regarding Christian Orthodoxy, Stark argues that it underscores the importance of ritual participation without having the same role for sustaining the moral order as the Protestant and Catholic Churches do (Stark 2001). Although the Orthodox Churches present themselves as defenders of the traditional social order, the absence of an association between religiosity and moral (or other) outcomes in Christian Orthodox cultures has been confirmed by several other studies (Halman and Van Ingen 2015; Prutskova 2013, 2015). Prutskova (2015) suggests an alternative explanation for this phenomenon. She argues that the current generations of believers were socialized during the period of the Communist repression of religion. Individuals did not therefore internalize basic moral teachings. The correlations do, however, exist in Catholic and Protestant countries from the former Warsaw Pact. The absence of a correlation between religiosity and moral and social outcomes in Orthodox countries is a research puzzle that merits more in-depth research.

Comparative research about religion must carefully evaluate whether the hypotheses regarding the moderation of individual religiosity by religious contexts are valid

for all religions included in the analysis. If this is not the case, the significant interactions might result from differences between religious cultures. In the future, the estimation of cross-level interactions should include controls for religious traditions to rule out the possibility that the effects of religious context mask denominational differences in the effects of religiosity. Researchers have to be careful when suggesting a hypothesis based on Protestant and Catholic theologies whilst including samples that are not of the Protestant or Catholic traditions in the analysis. For example, approximately a quarter of the countries covered by the most recent wave of the European Values Study (2008–2010) are Christian Orthodox.[8] It is not certain that including as many countries as possible is the optimal strategy to obtain the most valid results. Excluding countries or contexts where the theoretical assumptions cannot be assumed to be valid will improve the reliability of the results in many cases.

Another issue is that *case selection for comparative research requires more attention from researchers.* More than a decade ago, Ebbinghaus (2005) criticized quantitative comparative research for selecting the cases unsystematically. Almost all authors used the data available within the respective survey programs, and only excluded cases for reasons of missing data. However, the test statistics are only meaningful if the sample of countries was randomly drawn from a population of countries. The comparative survey studies are therefore likely to suffer from important selection bias (Schmidt-Catran 2019). Research in political science has extensively discussed the necessity of systematic case selection if a random sample cannot be drawn (Lauth et al. 2015). Selection bias can be modeled if relevant control variables are included in the models. There are, however, practical limitations because the number of contextual cases is often limited. Especially for the cross-level interactions, no study has reported results including relevant context-level controls for the interaction terms (religious cultures for example). Directly related to the problem of selection bias is the fact that *a majority of the studies cited here are based on only four international survey programs.* Most studies use data from the European Values Study or the World Values Study, and a few use data from the International Social Survey Program. Only two studies use data from the European Social Survey. The latter includes only a few indicators of religiosity (denomination, church attendance, and religious self-description), which does not allow for a differentiated operationalization of religiosity. If the selection of countries in international survey programs results in a bias of the contextual effects or the moderation of individual religiosity by religious context, additional studies using the same data will tend to reproduce the same bias. Contradictions between the studies might result from the fact that every database has a specific selection bias.

A method to avoid selection bias from cross-cultural studies is to complement the evidence from the comparative studies with data from a single country or groups of countries that are more homogeneous regarding the religious culture (Stavrova 2015; Traunmüller 2011). If the same effects as in cross-country comparisons are

---

[8] More information about the data from the European Values Study are available from the project website, www.europeanvalues.eu.

also found in regional analyses of religiously homogeneous contexts, there is more reason to trust the results.

6 Recommendations for Further Research

Although the paper does not allow for clear conclusions about the validity of theories regarding how religious context moderates the effects of individual religiosity, this review does enable some recommendations to be made for improving future research:

1. Papers should always report the correlations between individual religiosity and the outcomes for each contextual unit. This information helps identify regional or cultural patterns potentially confounding with indicators for religious contexts used in the study.
2. All studies should provide clear theoretical arguments for testing cross-level interactions. They should not be reported as byproducts of papers focusing on the main effects of contextual predictors because this increases the risk that studies report randomly significant estimates.
3. The operationalization of religiosity and religious context should refer to the theoretical framework. It should clearly distinguish between religious practice and/or religious beliefs. If composite scores for religiosity are used, scholars should also report the results for the single indicators. Before computing composite scores, they should check whether the effects of the single indicators have the same direction, or they should demonstrate the validity of the measurement using psychometric techniques (such as confirmatory factor analysis).
4. The rationale for selecting cases should be stated in the papers, and the risk of selection bias should be discussed. Appropriate robustness checks should be applied. The basic theoretical assumptions should be valid in all religious cultures covered by the empirical analysis.
5. The main effects and the cross-level interaction effects should be estimated with relevant controls for the denominational composition of the sample and/or other relevant cultural attributes. Studies including Orthodox or Far Eastern countries should control for these cultural affiliations because there are considerable doubts about whether religiosity is at all correlated with social and moral outcomes in these cultures.
6. Finally, cross-national studies are at risk of reporting interaction effects that are confounded with other contextual attributes. A method for solving this problem is to complement cross-national studies with regional studies from single countries or groups of similar countries, thus limiting cultural and denominational diversity. A possible means to do this is to use the regional information available from international survey programs for a subsample of culturally homogeneous countries (e.g., Protestant countries). Providing both cross-cultural and sub-national evidence of the moderation of religiosity by religious context would increase the trustworthiness of the results.

7 Conclusion

The correlations of individual religiosity with attitudes and behaviors are not universal. They vary across contexts, and there is evidence that the religious context is important for shaping the effects of religiosity. Studying how the religious context moderates individual religiosity might reveal the mechanisms underlying the correlations between religiosity and social and moral outcomes. The evidence published to date does not allow one to generalize the findings, because too much contradicting evidence is available.

Comparing the results of different papers is difficult because they study different outcomes and use different measurements for religiosity and different measures of religious contexts. Moreover, the authors naturally chose their theoretical framework depending on the outcome that they are studying.

The review of existing research does not therefore provide conclusive evidence for deciding which theoretical framework offers the best explanation of the systematic variation of the individual-level effects, even for single groups of dependent variables. Instead, I pointed to contradictory findings to identify major limitations in existing studies, and suggested recommendations to improve future research. Such recommendations can help address limitations in existing studies in order to improve future research on how religious contexts shape the effects of religiosity. Most importantly, the cultural and denominational specificities require greater consideration in cross-national studies because religiosity seems to be unrelated to attitudes and behavior in specific cultural settings. More studies are needed to determine which contextual factors moderate individual religiosity before conclusions can be generalized to enrich theorizing about religion.

References

Adamczyk, Amy. 2008. The effects of religious contextual norms, structural constraints, and personal religiosity on abortion decisions. *Social Science Research* 37:657–672.

Adamczyk, Amy, and Cassady Pitt. 2009. Shaping attitudes about homosexuality: The role of religion and cultural context. *Social Science Research* 38:338–351.

Adamczyk, Amy, Katharine A. Boyd and Brittany E. Hayes. 2016. Place matters: Contextualizing the roles of religion and race for understanding Americans' attitudes about homosexuality. *Social Science Research* 57:1–16.

Allport, Gordon W., and J. Michael Ross. 1967. Personal religious orientation and prejudice. *Journal of Personality and Social Psychology* 5:432–443.

Boyd, Katharine A., and Hyewon Chung. 2012. Opinions toward suicide: Cross-national evaluation of cultural and religious effects on individuals. *Social Science Research* 41:1565–1580.

Bruce, Steve. 2006. Secularization and the impotence of individualized religion. *The Hedgehog Review* 8:35–45.

Dingemans, Ellen, and Erik Van Ingen. 2015. Does religion breed trust? A cross-national study of the effects of religious involvement, religious faith, and religious context on social trust. *Journal for the Scientific Study of Religion* 54:739–755.

Doebler, Stefanie. 2015a. Relationships between religion and two forms of homonegativity in Europe—A multilevel analysis of effects of believing, belonging and religious practice. *Plos One* 10, https://doi.org/10.1371/journal.pone.0133538.

Doebler, Stefanie. 2015b. Love thy neighbor? Relationships between religion and racial intolerance in Europe. *Politics and Religion* 8:745–771.

Dülmer, Hermann. 2014. Modernization, culture and morality in Europe: Universalism, contextualism or relativism? In *Value contrasts and consensus in present-day Europe. Painting Europe's moral landscapes,* eds Wil Arts and Loek Halman, 251–276. Leiden: Brill.

Durkheim, Emile. 1897. *Le suicide: étude de sociologie.* [Suicide: a sociological study] Paris: Alcan.

Durkheim, Émile. 1912[1968]. *Les formes élémentaires de la vie religieuse. Le système totémique en Australie.* [The elementary forms of religious life. The totem system in Australia] Paris: Presses Universitaires de France.

Ebbinghaus, Bernhard. 2005. When less is more—Selection problems in large-N and small-N cross-national comparisons. *International Sociology* 20:133–152.

Finke, Roger, and Amy Adamczyk. 2008. Cross-national moral beliefs: The influence of national religious context. *The Sociological Quarterly* 49:617–652.

Gnaldi, Michela, Venera Tomaselli and Antonio Forcina. 2018. Ecological fallacy and covariates: New insights based on multilevel modelling of individual data. *International Statistical Review* 86:119–135.

Goldberg, Andreas C. 2014. The impact of religion on voting behvior—a multilevel approach for Switzerland. *Swiss Political Science Review* 20:305–329.

Halman, Loek, and Erik Van Ingen. 2015. Secularization and changing moral views: European trends in church attendance and views on homosexuality, divorce, abortion, and euthanasia. *European Sociological Review* 31:616–627.

Halman, Loek, and John Gelissen. 2019. Values in life domains in a cross-national perspective. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00602-0.

Hayward, R. David, and Marta Elliott. 2014. Cross-national analysis of the influence of cultural norms and government restrictions on the relationship between religion and well-being. *Review of Religious Research* 56:23–43.

Headey, Bruce, Hoehne Gerhard and Gert G. Wagner. 2014. Does religion make you healthier and longer lived? Evidence for Germany. *Social Indicators Research* 119:1335–1361.

Hox, Joop. 2002. *Multilevel analysis. Techniques and applications*. Mahwah/London: Lawrence Erlbaum Associates.

Huber, Stefan. 2007. Are religious beliefs relevant in daily life? In *Religion inside and outside traditional institutions,* ed. Heinz Streib, 211–230. Leiden: Brill.

Huijts, Tim, and Gerbert Kraaykamp. 2011. Religious involvement, religious context, and self-assessed health in Europe. *Journal of Health and Social Behavior* 52:91–106.

Inglehart, Ronald. 2006. Mapping global values. *Comparative Sociology* 5:115–136.

Inglehart, Ronald, and Wayne E. Baker. 2000. Modernization, cultural change, and the persistance of traditional values. *American Sociological Review* 65:19–51.

Inglehart, Ronald, and Christian Welzel. 2005. *Modernization, cultural change, and democracy*. Cambridge/New York: Cambridge University Press.

Kelly, Jonathan, and Nan Dirk De Graaf. 1997. National context, parental socialization, and religious belief: Results from 15 Nations. *American Sociological Review* 62:639–659.

Lauth, Hans-Joachim, Gert Pickel and Susanne Pickel. 2015. *Methoden der vergleichenden Politikwissenschaft*. [Methods of comparative political science] Wiesbaden: Springer VS.

Lüchau, Peter. 2007. By faith alone? Church attendance and christian faith in three European countries. *Journal of Contemporary Religion* 22:35–48.

Luria, Gil, Ram A. Cnaan and Amnon Boehm. 2017. Religious attendance and volunteering: Testing national culture as a boundary condition. *Journal for the Scientific Study of Religion* 56:577–599.

Nicholson, Amanda, Richard Rose and Martin Bobak. 2009. Association between attendance at religious services and self-reported health in 22 European countries. *Social Science & Medicine* 69:519–528.

Norris, Pippa, and Ronald Inglehart. 2004. *Sacred and secular. Religion and politics worldwide*. Cambridge/New York: Cambridge University Press.

Oarga, Cristina, Olga Stavrova and Detlef Fetchenhauer. 2015. When and why is helping others good for well-being? The role of belief in reciprocity and conformity to society's expectations. *European Journal of Social Psychology* 45:242–254.

Olson, Daniel V. A., and Miao Li. 2015. Does a nation's religious composition affect generalized trust? The role of religious heterogeneity and the percent religious. *Journal for the Scientific Study of Religion* 54:756–773.

Pargament, Kenneth I., Bruce W. Smith, Harold G. Koenig and Lisa Perez. 1998. Patterns of positive and negative religious coping with major life stressors. *Journal for the Scientific Study of Religion* 37:710–724.

Pickel, Gert. 2010. Säkularisierung, Individualisierung oder Marktmodell? Religiosität und ihre Erklärungsfaktoren im europäischen Vergleich. [Secularisation, individualisation or a market model? Religiosity and its explanatory factors in a European comparison] *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62:219–245.

Pollack, Detlef, and Gergely Rosta. 2015. *Religion in der Moderne: ein internationaler Vergleich*. [Religion in modern times: an international comparison.]

Prutskova, Elena. 2013. Religioznost' i ee sledstviia v tsennostno-normativnoi sfere. *Sotsiologicheskii zhurnal* 2:72–88.

Prutskova, Elena. 2015. Sviaz' religioznosti i tsennostno-normativnykh pokazatelei: faktor religioznoi sotsializatsii. *Vestnik PSTGU. Series I: Theology* 59:62–80.

Ruiter, Stijn, and Nan Dirk De Graaf. 2006. National context, religiosity, and volunteering: Results from 53 countries. *American Sociological Review* 71:191–210.

Saroglou, Vassilis. 2011. Believing, bonding, behaving, and belonging: The big four religious dimensions and cultural variation. *Journal of Cross-Cultural Psychology* 42:1320–1340.

Scheepers, Peer, Manfred Te Grotenhuis and Frans Van Der Silk. 2002. Education, religiosity and moral attitudes: Explaining cross-national effect differences. *Sociology of Religion* 63:157–176.

Schmidt-Catran, Alexander W., Malcolm Fairbrother and Hans-Jürgen Andreß. 2019. Multilevel models for the analysis of comparative survey data: Common problems and some solutions. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00607-9.

Snijders, Tom A., and Roel J. Bosker. 1999. Multilevel Analysis. An introduction to basic and advanced multilevel modeling. London: Sage.

Spenkuch, Jörg L., and Philipp Tillmann. 2018. Elite influence? Religion and the electoral success of the Nazis. *American Journal of Political Science* 62:9–36.

Stack, Steven, and Augustine J. Kposowa. 2011. The effect of survivalism-self-expressionism culture on black male suicide acceptability: A cross-national analysis. *Social Science & Medicine* 72:1211–1218.

Stark, Rodney. 1996. Religion as context: Hellfire and delinquency one more time. *Sociology of Religion* 57:163–173.

Stark, Rodney. 2001. Gods, rituals, and the moral order. *Journal for the Scientific Study of Religion* 40:619–636.

Stark, Rodney, and Roger Finke. 2000. *Acts of faith. Explaining the human side of religion*. Berkeley: University of California Press.

Stark, Rodney, Daniel P Doyle and Lori Kent. 1980. Rediscovering moral communities: Church membership and crime. *Understanding Crime: Current Theory and Research* 18:43–52.

Stavrova, Olga. 2015. Religion, self-rated health, and mortality: Whether religiosity delays death depends on the cultural context. *Social Psychological and Personality Science* 6:911–922.

Stavrova, Olga. 2019. How much do sources of happiness vary across countries? A review of the empirical literature. In *Cross-national comparative research – analytical strategies, results and explanations*. Sonderheft *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. Eds. Hans-Jürgen Andreß, Detlef Fetchenhauer and Heiner Meulemann. Wiesbaden: Springer VS. https://doi.org/10.1007/s11577-019-00612-y.

Stavrova, Olga, and Pascal Siegers. 2014. Religious prosociality and morality across cultures, how social enforcement of religion shapes the effects of personal religiosity on prosocial and moral attitudes and behaviors. *Personality and Social Psychology Bulletin* 40:315–333.

Stavrova, Olga, Detlef Fetchenhauer and Thomas Schlosser. 2013. Why are religious people happy? The effect of the social norm of religiosity across countries. *Social Science Research* 42:90–105.

Storm, Ingrid. 2016. Morality in context: A multilevel analysis of the relationship between religion and values in Europe. *Politics and Religion* 9:111–138.

Taylor, Charles. 2007. *A secular age*. Cambridge, MA: The Belknap Press of Havard University Press.

Tittle, Charles R., and Michael R. Welch. 1983. Religiosity and deviance—toward a contingency theory of constraining effects. *Social Forces* 61:653–682.

Traunmüller, Richard. 2011. Moral communities? Religion as a source of social trust in a multilevel analysis of 97 German regions. *European Sociological Review* 27:346–363.

Verbakel, Ellen, and Eva Jaspers. 2010. A comparative study on permissiveness toward euthanasia. *Public Opinion Quarterly* 74:109–139.

Voas, David, Daniel V.A. Olson and Alasdair Crockett. 2002. Religious pluralism and participation: Why was previous research wrong? *American Sociological Review* 67:212–230.

Weaver, Andrew J., Kenneth I. Pargament, Kevin J. Flannelly and Julia E. Oppenheimer. 2006. Trends in the scientific study of religion, spirituality, and health: 1965–2000. *Journal of Religion and Health* 45:208–214.

Welch, Michael R, Charles R Tittle and Thomas Petee. 1991. Religion and deviance among adult catholics: A test of the "moral communities" hypothesis. *Journal for the Scientific Study of Religion* 30:159–172.

Zuckerman, Phil. 2012. Contrasting irreligious orientation: atheism and secularity in the USA and Scandinavia. *Approaching Religion* 2:8–20.

Pascal Siegers 1980, Dr. rer. pol., head of the Research Data Center German General Social Survey (FDZ ALLBUS) and team leader of National Surveys at the GESIS Data Archive in Cologne. Areas of Research: sociology of religion & morality, quantitative religious research, new data types in social research, research data infrastructures. Publications: Religious indifference and religious rites of passage. Cham 2017; The religious and spiritual underpinnings of party choice in Christian Europe. Electoral Studies 44, 2016 (with S. Franzmann and M. Hassan); Alternative Spiritualitäten: neue Formen des Glaubens in Europa: eine empirische Analyse. Frankfurt a. M. 2012.