



Internationally Comparative Research Designs in the Social Sciences: Fundamental Issues, Case Selection Logics, and Research Limitations

Achim Goerres · Markus B. Siewert · Claudius Wagemann

Published online: 29 April 2019

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Abstract This paper synthesizes methodological knowledge derived from comparative survey research and comparative politics and aims to enable researchers 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

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 sub-national 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).

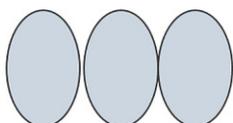
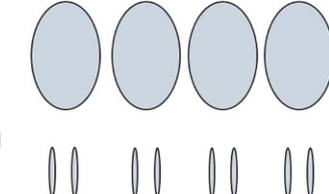
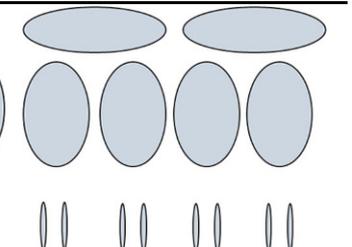
When speaking about comparative research, we thus quickly touch upon the debates between qualitative and quantitative methods, or more specifically between macro-level comparativists using (comparative) case study logics *versus* quantitative researchers who apply the large-*N* logic of individual analysis 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 seq.; 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 mid-sized 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 reunification 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 example 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 *independence* 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 white-washed 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 nonexistent) 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-Schlusser 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; Rohlfling 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.

Acknowledgements Equal authors listed in alphabetical order. We would like to thank Ingo Rohlfling, Anne-Kathrin Fischer, Heiner Meulemann and Hans-Jürgen Andreß for their detailed feedback, and all the participants of the book workshop for their further comments. We are grateful to Jonas Elis for his linguistic suggestions.

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*. Berkeley: 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. 2010a. 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 inquiry. *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.