Conclusion: Lessons for the dialogue between theory and data

Gschwend, Thomas; Schimmelfennig, Frank

12

Conclusion: Lessons for the Dialogue between Theory and Data

Thomas Gschwend and Frank Schimmelfennig

In the introduction, we categorized research designs along two dimensions. One dimension classifies them according to the focus of research as *factor-centric* or *outcome-centric;* on the other dimension, we distinguished *large-n* and *small-n* research designs according to the number of observations. Yet we also claimed that, no matter which research design we use, we all face the same set of core research design issues: Defining the research question and problem, specifying concepts and theory, operationalizing and measuring them, selecting cases and observations, controlling for alternative explanations, and drawing theoretical conclusions from the empirical analysis. Each of the preceding chapters then took on one of these issues and explicated the challenges, and also provided some hands-on advice on how to deal with these challenges.

What are the lessons to be learned from comparing the challenges across all types of research design? The results here seem to be unequivocally clear. It does not matter whether you care about outcomes or causal factors nor whether you can leverage a few or many observations. We do in fact share the very same research design problems. We can identify a set of questions which help to increase the relevance of our research both in the scientific community and beyond (Lehnert, Miller and Wonka, Chapter 2). If your theoretical concepts are fuzzy, your research cannot yield valid inferences – no matter how many observations you can leverage on or the type of inferences in which you are primarily interested (Wonka, Chapter 3). Moreover, measurement as a process of attributing 'values to observations according to pre-defined rules' (Miller, Chapter 5, p. 84) is a challenge irrespective of the number of observations you measure and whether your main theoretical focus is on an independent or a dependent variable. Whether you select a few or many observations, selection bias is always looming large (Thiem, Chapter 7; Leuffen, Chapter 8; Geddes, 1990). Likewise, the decision as to which variables to include in a quest for explanation, and which to control for is tricky in any type of research design (Sieberer, Chapter 9; Dür, Chapter 10). Finally, a potential reformulation of the theory which started a dialogue with the data is an issue in every empirical research process (De Bièvre, Chapter 10).

While all types of research face the same problems and challenges, to what extent do they also lend themselves to common solutions? The answer from comparing the guidelines that are offered in each chapter seems to suggest that we should not expect to find a cookie-cutter approach 'out there' to solve all research design problems for us in the same mechanical way. Surely this does not come as a big surprise. Otherwise our distinction of research designs along two different dimensions would be just one more attempt to clutter the literature with yet another piece of jargon. Rather, the preceding chapters play variations of a common theme: *Different research designs offer and require different solutions to the very same challenges, each of which produces specific trade-offs*. The evaluation of these trade-offs should ideally determine the research design you choose. This fact, we think, has not been appreciated enough in the discussion about unified logics and common standards of good research design.

Relevance

The only exception may be seen at the very beginning of the research process. For one, the social or theoretical relevance of the research question does not appear to be systematically related to the number of observations or factor- versus outcome-centric designs. A single case study can be just as (ir)relevant as a global survey. Both knowledge of the causal effects of a single factor and knowledge of the multiple determinants of a specific outcome can or cannot meet the standards of relevance. At first sight, outcome-centric research – for example, on the conditions of wars, effective institutional reform, or electoral success – may seem more relevant. However, we do not see why this should not be the case for factor-centric research on the causal effects of peacekeeping activities, constitutional designs, or electoral systems.

Concept specification

Clearly specified theories and concepts are indispensable for all types of research design. On the one hand, as Wonka (Chapter 3) notes, the specification of concepts needs to follow the theoretical interests of the researcher or the study rather than the selected research design. On the other hand, however, the extension of a concept must also be commensurate with the object of research. Whereas the 'Cold War' will hardly qualify for a large-n study, 'international rivalry' does. According to Wonka (see also Rathke, Chapter 6), decreasing the intention of a concept to widen its empirical applicability involves raising the level of abstraction and possibly shedding context specificity (for instance from 'Cold War' to 'international rivalry'), which is likely to blur conceptual boundaries and reduce analytical leverage. Hence, the common perception of 'qualitative', small-n researchers that 'quantitative', large-n researchers often work with extremely thin concepts, which neglect important real-world variations and are used out of context. By contrast, large-n researchers may find that many concepts used in small-n research are so 'thick' and overloaded with context-specific attributes that they are not only hard to measure, let alone quantify, but also stand in the way of comparative research and general knowledge. Whereas, in the first case, the analytic leverage of the concept derives from context specificity, in the second case it comes from its general applicability and context independence. This trade-off applies to concept-specification in large-n and small-n studies regardless of whether they are factor-centric or outcome-centric (Table 12.1)

Measurement

Rathke (Chapter 6) brings up a general measurement issue that researchers are confronted with when devising measurement strategies based on secondary data from various sources. Not only may the measures used in different data collection projects be incomparable, but even if the measurement instruments are formulated identically, they may produce

Table 12.1 Research design and concept specification

	Factor-centric	Outcome-centric
Large n	Abstraction, context independence, 'thinness'	
Small n	Concreteness, context specificity, 'thickness'	

Table 12.2 Research design and measurement

	Factor-centric	Outcome-centric
Large n	Variance-based validity and reliability	
Small n	Case-based validity and reliability	

incomparable results because of the different political or cultural context in which they are applied. Researchers must therefore check for and ensure conceptual equivalence of the measures used. Although exemplified based on a large-n research design, the challenge she brings up is neither specific to the number of cases under investigation nor to the focus of research, be it factor-centric or outcome-centric.

Since measurement is intimately linked to concept specification, the conclusions from the book are similar. On the one hand, Miller (Chapter 5) argues that questions of measurement, and the problems of validity and reliability, apply to factors (independent variables) and outcomes (dependent variables) alike. Therefore, it does not make a difference for the choice of solutions to measurement problems whether the research design is factor-specific or outcome-specific. On the other hand, however, the measurement problems in small-n and large-n studies mirror those of the specification of concepts. As Miller points out, large-n studies are 'often said to be reductionist, based on inadequate indicators ... resulting in poor data quality', whereas small-n studies draw criticism for being prone to biases (Miller, Chapter 5, p. 84).

Put positively, small-n designs allow the researcher to become very familiar with the individual cases and to put high emphasis on the refinement of indicators and measures as well as the interpretation of data in order to improve case-based validity and reliability. By contrast, in large-n studies, the researcher is more likely to encounter the full range of variance on the independent and dependent variables, which is equally helpful in refining indicators and measures and improving their variance-based validity and reliability. However, the researcher will not be able to put the same effort into checking the validity and reliability of the measurement in the individual case. Thus, we end up with the familiar trade-off between depth and breadth (Table 12.2).

Case selection

Case selection is the issue on which the four research designs differ most obviously. On the one hand, large-n and small-n studies vary precisely on the number of cases selected for analysis. On the other hand, factor-centric

Table 12.3 Research design and case selection

	Factor-centric	Outcome-centric
Large n	Random selection (or universe of cases)	
Small n	Intentional selection on the independent variable, selection of crucial cases	Intentional selection on the dependent variable plus within-case analysis

research implies the selection of cases on the independent variable, whereas outcome-centric research intentionally selects cases on the dependent, 'outcome' variable. These differences of design translate into different specific problems, solutions and various trade-offs (Table 12.3).

According to Leuffen (Chapter 8) and Thiem (Chapter 7), selection bias potentially plagues all varieties of political science research. This is particularly true of 'real world bias' induced by history and political processes. Yet the extent to which selection bias looms large (and can be detected and corrected) varies across research designs. First, whereas large-n researchers usually select their cases randomly (if not the entire universe of cases), small-n research needs to start from the intentional selection of cases, because random selection would be likely to produce selection bias and thus reduce the validity of inferences (King, Keohane and Verba, 1994, pp. 125–7). Intentional selection is generally more prone to bias than random selection (if it is really random).

Second, if selection is intentional, the inferences drawn from factor-centric research are in general less affected by the selection rule than those drawn from outcome-centric research, because researchers will usually select their cases on the explanatory variable (King, Keohane and Verba, 1994, p. 137). Despite this, a selection that does not cover the full range of the explanatory variables will *a priori* obfuscate their potential impact on the dependent variable, and consequently seriously threatens the generality of the inferences. Leuffen points out that this limit can be addressed if the researcher selects a 'crucial' case study (Eckstein, 1975). A 'theory-confirming' (Lijphart, 1971) or hard case will demonstrate that if the theory holds in this case, it will also hold in most other cases. Conversely a 'theory-infirming or easy case shows that if the theory does not hold in this case, it will also not hold in most other cases (see also Flyvbjerg, 2006, p. 230).

Third, the evaluation of selection bias is easier in large-n than in small-n studies, and large-n studies are more likely to cover the full range of each variable than small-n studies. We can thus conclude that large-n designs are less likely to suffer from selection bias or limited generality, whereas particularly small-n outcome-centric studies are most affected by selection bias and limited generality.

This negative characterization may, however, miss the point of small-n outcome-centric studies, aka case studies (see Table 1.2). Outcome-centric studies often search for explanations of the specific cases they study rather than being inspired by the quest for generalization, and they usually employ within-case designs such as process-tracing to produce causal inferences (George and Bennett, 2005). Limited generality is thus not a major concern, and their case-specific inferences are not affected by selection bias (Collier, Mahoney and Seawright, 2004, pp. 95–7).

To sum up, selection bias is not a threat to your inference if you deliberately limit the scope of your research question. At the same time, however, the limitation in scope does not allow one to draw any generalizations above and beyond the case-specific ones. An intentional selection on the dependent variable does deliberately restrict the range of the dependent variable and, therefore, cannot tell us anything about how well the causal story travels to other cases.

Control

The issues of case selection and control are partially linked via the problem of determinacy. The general rule to avoid indeterminacy is simple. An increase in the number of variables should correspond to an increase in the number of cases. Conversely, as the number of cases decreases, the researcher is forced to be selective with regard to the variables included. This, however, increases the likelihood of an 'omitted variable bias', that is, the lack of control for variables that may be correlated with both the explanatory variable and dependent variable, as well as 'equifinality' (that is, the fit of two and more hypotheses with the evidence and the inability to disentangle them). Again, however, the extent of the problem and the proposed solutions vary across research designs, and both the large/small-n and the factor/outcome-centric dimensions are relevant (Table 12.4).

Table 12.4 Research design and control

	Factor-centric	**Outcome-centric**
Large n	Add independent variables to avoid omitted variable bias	Add independent variables to maximize explained variance
Small n	Use typologies and matching	Use process-tracing

In general, large-n studies achieve control by adding independent variables to the analysis, because they tend to have high degrees of freedom. By contrast, small-n studies achieve control by carefully delimiting and matching cases and by using within-case evidence. However, there are differences between factor-centric and outcome-centric designs as well. As Sieberer points out for large-n research, factor-centric designs should minimize the addition of those independent variables to what is strictly necessary in order to avoid omitted variable bias. By contrast, outcome-centric studies will include all theoretically relevant and consistent 'independent variables that allow you to capture additional variance in the dependent variable' (Sieberer, Chapter 9, p. 169). The trade-off here is that while maximizing the number of variables included in a multivariate analysis will decrease the likelihood of omitted variable bias, it will also decrease the quality of the inferences the researcher can draw on for any individual variable (Ganghof, 2005a, pp. 79–80; King Keohane and Verba, 1994, pp. 182–4).

For factor-centric small-n research, both Lehnert (Chapter 4) and Leuffen (Chapter 8) emphasize the usefulness of typologies. For Lehnert, 'typologies can serve as a remedy for indeterminacy because they combine several variables into broader concepts, thus reducing the number of variables to be integrated into a causal model' (Chapter 4, p. 67). For Leuffen, theory-guided typologies help the researcher control for alternative explanations and focus research on the theoretically most interesting cells. More generally, factor-centric small-n research relies on the careful matching or controlling of cases. Ideally, if the researcher is able to find cases that vary broadly on the explanatory variable(s) of interest but are constant with regard to all other potentially relevant independent variables, a high degree of control is achieved and the causal impact of factors can be validly assessed without increasing the number of cases.

Dür (Chapter 10) specifically targets the problem of control in outcome-centric small-n research. Just as in outcome-centric large-n research, researchers using this design seek a full explanation of their cases. In contrast with the large-n variety, however, they often combine too little variance on the dependent variable with too many independent variables, thus resulting in indeterminacy or overdeterminacy and, indeed, the inability to decide which explanation *really* works or works *best*. Dür mainly advises the researcher to further specify the causal mechanisms implied by her own and alternative theories and then conduct a process-tracing analysis of these causal mechanisms to discriminate between competing explanations.

Theoretical conclusions

The final issue is the theoretical conclusions which can be drawn from our research. Here it seems that De Bièvre's recommendation not to infer falsification from a single anomalous case indicates that we can draw stronger conclusions from large-n research than from small-n research. In addition, however, it is important to see that falsification follows a different strategy in both designs. In large-n research, the assessment of theories is based on regularity and generality. A theory is corroborated if it is consistent with empirical evidence across many observations, the entire universe of cases or a representative sample, and it is weakened or falsified if it fails to account for the average pattern of outcomes. Single or a few deviant cases or outliers will either 'disappear' in the overall pattern or be consciously disregarded by the researcher. By contrast, small-n research focuses on single, critical cases or observations or conducts intensive within-case analysis to assess theories and explanations (see the discussion of crucial cases above and by Leuffen, Chapter 8). The trade-off is obvious: whereas large-n research tends to 'overlook' and neglect deviant special cases, small-n research is likely to put too much emphasis on them in drawing theoretical conclusions from research. This holds for both factor- and outcome-centric research (Table 12.5).

In addition, the theoretical conclusions one draws obviously have to match the kind of theory addressed in these two kinds of research design. Whereas factor-centric research only tells us something about 'factor-oriented theory', outcome-centric research addresses 'outcome-oriented theory'. To give an example, modernization theory is a typical factor-oriented theory that stipulates socioeconomic development as the cause of various relevant political outcomes such as democracy, political culture, and political cleavages (Lipset, 1959). By contrast, an outcome-oriented theory of democracy might bring together various factors (such as wealth, education, international environment, export dependency, and civil society) to explain the variance in democratic stability as fully as possible. In general, however, De Bièvre's discussion of falsification and the reformulation of hypotheses as well as his practical

Table 12.5 Research design and theoretical conclusions

	Factor-centric	Outcome-centric
Large n	Conclusions based on average pattern of observations	
Small n	Conclusions based on critical observations	

guidelines apply to both factor- and outcome-centric research (De Bièvre, Chapter 11).

Two ways to analytical rigor: the logic of breadth and the logic of depth

In the preceding paragraphs, we have approached the variation of solutions across different types of research design individually for each design problem, from relevance to theoretical conclusions. That leads us to a bigger question: Is there a general logic behind these solutions, which cuts across the individual problems of research design? First, as reflected in the widely perceived cleavage between 'qualitative' and 'quantitative research', it appears from our collection of tables that the number of observations – that is, the divide between large-n and small-n – is the dominant dimension within the universe of research designs. For each design problem, the solutions for large-n research differed clearly from those used in small-n research, whereas the solutions for factor- and outcome-centric research only varied for the problems of control and, partially, for case selection.

In addition, there do seem to be unified logics of both large-n and small-n designs across design problems. In most simple terms, large-n research follows the *logic of breadth*, whereas small-n research follows the *logic of depth*. In the literature, we find other dichotomies that capture a similar distinction. The logic of breadth corresponds to an extensive or generalizing research strategy, whereas the logic of depth resembles an intensive or particularizing research strategy (Dessler, 1999, p. 129). According to the logic of depth, small-n research seeks to maximize leverage by extracting as much information as possible from the analysis of a single or a few cases studied in depth. This includes concrete, context-specific concept specification, the case-based improvement of the validity and reliability of measurement, the intentional selection of the 'right' case or cases, control through within-case analysis or careful matching, and theoretical conclusions based on critical observations. According to the logic of breadth, large-n research seeks to maximize leverage by increasing the number of cases. This entails abstract and context-independent concepts, the variance-based optimization of measurement, random selection, the achievement of control by adding independent variables, and theoretical conclusions based on a lot of observations. Pliny's '*multum non multa*', cited by Leuffen (Chapter 8, p. 152) as a maxim for case selection in small-n research, can be generalized to small-n research as a whole. Conversely, large-n research acts on the maxim of '*multa non multum*' (many, not much).

In comparison, the distinction of factor-centric and outcome-centric research seems to be second-order and less pervasive. Nevertheless, it is relevant for two crucial problems of research design: The selection of cases (in small-n research) and the control for, and discrimination among, alternative explanatory factors. The common feature of factor-centric research is *a priori selection*. Factor-centric large-n research is highly selective in adding control variables ahead of the analysis. Factor-centric small-n researchers try to find crucial, carefully matched, or typologically categorized cases before conducting empirical research. By contrast, outcome-centric research is characterized by *a posteriori discrimination*. Outcome-centric large-n researchers add all plausible independent variables to their models and see whether they turn out to be significant and relevant in the analysis. Outcome-centric small-n research relies on within-case analysis, in particular the process-tracing of causal mechanisms, to discriminate among alternative explanations (Scharpf, 1997).

In sum, while we have analyzed the different logics and solutions of alternative research designs in this concluding chapter, we do not wish to reify our typology of research designs. Instead we would like to stress two lessons we learned for the dialogue between theory and data that emanates from all this. First, as we have emphasized in the introduction, researchers are free to choose research designs. Nevertheless, *different research designs offer and require different solutions to the very same challenges, each of which produces specific trade-offs*. Second, there is no reason why researchers should not combine research designs or move from one design to the other to compensate for the weaknesses and limits of a particular design and to capitalize on the strengths of the other in a 'nested analysis' (Lieberman, 2005) if time constraints and the scarcity of other resources do not suggest otherwise. But in order to do so, they need to be aware of the logics of different research designs and cognizant of the solutions, guidelines, and trade-offs that any design choice entails. Hence our plea for your dialogue between theory and data: Get in sync with the opposite camps! Understand the conflicts. Make deliberate choices. End the confusion.