

Comparative Analysis for Theory Development

Mala Htun

University of New Mexico

Francesca R. Jensenius

University of Oslo

Norwegian Institute of International Affairs

This is a pre-copyedited version of a chapter in the book manuscript *Rethinking Comparison*, edited by Erica S. Simmons and Nicholas Rush Smith (under review with Cambridge University Press).

Introduction

Comparison generates a big payoff for political science. Comparing generates insights not visible when we focus on a single institution, group, event, or context, enabling us to discover what is unusual and what is more widely shared. The practice of comparing requires that we clearly define and operationalize our concepts, and sharpens awareness of the assumptions that underlie our theories. In addition, comparing often changes what questions we ask.

Whereas the advantages of comparing are many, our field's most widely-taught qualitative methods texts tend to orient research designs toward only one of comparison's many purposes—testing causal theories via controlled comparisons across a small number of cases, as the introduction to this volume points out. But in practice, comparisons often have other goals—such as developing theoretical arguments, identifying key variables, exploring scope conditions, or probing mechanisms (see Eckstein, 1975)—particularly so if they form part of a multi-method project. As this suggests, there is a discrepancy between the types of research designs scholars typically teach and the type of research they actually conduct.

In this chapter, we identify five misperceptions about comparative research design that we have encountered among our students and which are likely to be more widely held. These misperceptions pertain to the goal, nature, logic, and justifications of comparative research design, and include the belief that the primary goal of comparative research is to test theories; that a research project has to have only one, static, research design; that all qualitative projects must be shoe-horned into a narrow understanding of a “case study”; that theory primarily derives from epiphanies, identifying holes in existing literature, or close scrutiny of single experiences or cases; and that it is “unscientific” to take personal or practical concerns into account when choosing what to study.

We then discuss why we consider these ideas to be misperceptions. First, we hold that comparative research has multiple goals besides theory testing, such as theory development, exploration of mechanisms, and so forth, and that we therefore need different types of comparative research design for different projects. Second, we argue that research design should evolve as projects evolve and that different parts of a project might need a different design. And, we hold that, as a general rule, a comparative study should be designed in a way that creates interesting variation in whatever one seeks to learn more about with a part of the study. Third, rather than selecting “cases” to test theories, we propose that projects aimed at theory development actually tend to focus on selecting diverse *contexts* in which to explore ideas. This is particularly when researchers are unsure what the phenomenon they are studying is “a case of,” or what the most important variables in their argument will be. Fourth, we explain why we consider comparative research to be a great source of theoretical insight. Fifth, we argue that personal and practical concerns should guide, but not constrain, our choice of field sites. To be sure, none of these arguments are new. Rather, we are reiterating perspectives that are vital to our own methodological understanding but that are scattered across the literature and rarely emphasized, and not often taught, as part of training in research design.

After analyzing and pushing beyond the five misperceptions we have encountered, we discuss how the five arguments pertain to the evolution of one of our own ongoing projects on women’s empowerment. Our study started out as a cross-national study of the relationship between gender discriminatory laws and women’s economic empowerment but, as our research

progressed, new questions emerged. We became more interested in how the concept “empowerment” has different meanings and implications in different parts of the world. As a result, different research designs became more appropriate. To probe the concept’s differing reference points and boundaries, we conducted fieldwork in three contexts: Norway, Japan, and the United States. These countries are not “cases” for us—understood as specific instances of a clearly defined class of events (George and Bennett 2005)—but “contexts” that provide interesting variation in the key concept we wish to learn more about. Our “conceptual comparison” across three very different contexts does not allow us to draw strong inferences about the associations between state action and societal outcomes. But the conceptual comparison has made us think differently about the cultural and historical backdrops that shape state action and people’s responses, and changed our theoretical thinking about how to study the effects of state action on society. Finally, we offer examples of how our multi-cultural research team has enabled us to build local knowledge and leverage insider and outsider advantages across different contexts. Our experiences are suggestive of a potentially liberating approach to research design for qualitative comparative politics.

Five misperceptions about research design for qualitative comparative studies

Theory testing is not the only goal of comparative research

The publication of *Designing Social Inquiry* in 1994 (King et al 1994, henceforth KKV) cast a long shadow over much of the methodological training received by graduate students in political science. The book sparked an important discussion about research design and scientific rigor, and offers useful advice. It focuses primarily on how to use observational data to produce causal inferences. The main take-away message is that a larger N is better and that it is difficult to draw causal conclusions from the comparison of a small number of observations. For “small-N” research, the book promotes controlled comparisons, preferably with observations selected on the explanatory variable. Researchers interested in something other than causal inference are referred to as “interpretivists” (King et al 1994, p. 37) and the book has little guidance to offer in terms of research design for their studies. In fact, as they reiterated in King et al 1995 (p. 475),

their goal was to provide guidance on research designed aimed to evaluate theories, not to develop theory.

KKV presume that the general goal of most scholars conducting comparative work is to test theories about a causal relationship among variables, and they offer research design advice accordingly. Many later books on qualitative methods for political science follow KKV by focusing overwhelmingly on comparative research designs aimed at theory testing, either explicitly (e.g. Gerring 2001, Brady and Collier 2010) or more implicitly (e.g. George and Bennett). George and Bennett, for example, argue that case studies in general are good for fostering new hypotheses, but they refer to *comparative* research designs as “case study methods that attempt to approximate the conditions of scientific experiments” (2005, p. 151), which affirms KKV’s idea that comparison is geared toward causal inference. These widely taught text books thus emphasize controlled comparisons across small numbers of units, or variations of Mill’s methods of agreement and difference (see Mill 1884, Teune & Przeworski 1970). By engaging with these texts, many students acquire the misperception that theory testing is the primary goal of comparative studies. And methods texts insist that research designs aimed at testing theories should be carefully designed to avoid endogeneity and various forms of bias, leading us to the recommendation of controlled comparisons. However, controlled comparisons are not the only way to design comparative research, as this volume argues. In fact, controlled comparisons are arguably a poor way to test causal argument (Seawright, this volume).

Though methods texts focus on designs for theory testing, most empirical studies using comparative methods are aimed primarily at theory development. As a result, as George and Bennett (2005, p. 10) point out, there is a lack of alignment between the goal presumed by methods texts and the actual goal of students conducting comparative work. In contrast to theory testing, which is rigorous and transparent, texts tend to portray theory development as a more casual or idiosyncratic process, usually happening through the study of single country experiences, and as a “lesser cousin” to the real work of theory testing. The methods training students receive may thus seem irrelevant to them or even prevent them from developing an appropriate research design, as they try to force their projects into a design that is inappropriate for their research goals (see also Soss, this volume).

Another danger of having little concrete advice about how to conduct research aimed at theory development is that we may fall in the trap of thinking that anything goes. However, we should reflect on and defend our research design for theory development in a similar way as research designs aimed at theory testing. As we argue below, scholars should think carefully about how to best design their study to achieve their particular goals, whether that is interpreting, building causal intuition, elaborating concepts, identifying patterns of regularity, exploring mechanisms or something else.

Variation and iteration in research design

The second misperception we often encounter is the idea that a research project needs to have one static research design, an idea pushed in part by the growing emphasis, in recent years, on the importance of rigor and transparency in research, and a trend towards pre-registration of empirical studies. For research aimed at testing clearly laid out theoretical claims, these are exciting developments, which can help to prevent p-hacking and constrain scholars tempted to adjust their research question after the fact to conform to the patterns they actually found. Yet the trend toward pre-registration, which presumes a fixed design, is not a helpful approach for *all* types of research, nor for all *parts* of a research project. Different parts of a project, for example, might have different goals, require different designs, and involve comparing different things. Even if your overarching goal is to develop a causal argument, you may not want to design the qualitative part of your study as a quasi-experiment. Collier et al (2004, p. 101), for example, defend “no-variance designs,” which involves the comparison of units with the same value on the dependent variable. Though King et al (1995), Geddes (1990), and others warn against this type of design, Collier et. al. say that such a design can help to frame the analytic problem, suggest causal ideas, refine conceptualization and measurement (including stimulating the creation of typologies), allow for valuable descriptive inferences, and provide exploratory tests. In other words, since different parts of a study play different roles, it is ok that they are designed differently.

What is more, a research design can and should evolve with the theoretical thinking. If your theoretical thinking changes, so should your research design. George and Bennett (2005, p. 73) note that one might need “some iteration” in the process of designing and implementing

qualitative research. This is an under-statement. Very often, researchers start out with a research design based on their theoretical priors, but as they start conducting their work, their understanding of key concepts, important variables, and even what the research question should be, changes. This is partly because of the realities of fieldwork. As McKeown (2004, p. 164) notes, during fieldwork we often find that some of our key assumptions going into the study are wrong, that the data-gathering techniques we envisioned are not feasible, or that people we thought would be key informants are actually irrelevant. Changes are also simply a part of the intellectual process of discovery, and the path to knowledge is rarely as straight-forward as published articles and books make it seem. As Lund (2014, p. 231) puts it: “the orderliness of one’s method is easier to establish in hindsight as futile detours can be erased to make it look more coherent and neat than it felt and was at the time.” As this suggests, it is important to start out with a plan and to provide a sensible justification for that plan, but it is also fine to modify that plan as one learns more. An overly static approach may stifle the intellectual contributions from a project focused primarily on theory development.

Existing literature offers few recommendations about how to select cases for theory development. As Levy (2008) points out, the main recommendation that is made is to pick cases that are extreme or deviant with respect to the main dependent or independent variables (see also Van Evera 1997). If you have a clear sense of what those variables are, this makes sense. But many people engaged in the early stages of projects aimed at theory development may not know what their main variables are. In fact, a great deal of qualitative work occurs *prior to* gathering information on relevant variables. And even if you do have a sense of key variables it is often counterproductive to follow this strategy unless the goal is to observe whether different values in the explanatory variables are associated with different outcomes on the outcome variable. You are better off with a design better aligned for your research goal.

Gerring (2011, 41) includes a comparative design that is not aimed at theory testing in his overview of “case-selection strategies and criteria.” When the goal is primarily “descriptive inference,” he recommends selecting a diverse set of cases based on their “descriptive features.” This is sound advice, but it underplays the utility and importance of the comparative research design. Theory development is not the same as “description”: it is a broad category of activities

with different goals, including identifying possible variables, developing concepts, probing causal mechanisms, exploring scope conditions, and so on.

Generalizing from both Levy and Gerring, a more general approach is to choose units that offer *interesting variation in whatever we wish to understand more about* (the possible key variable, concept, mechanisms, or conditions). This *variation* is what will provide all the benefits of a comparative design and yield insights that can enhance our understanding and make it clearer what exactly we are studying. In other words, we should select objects of study that differ in whatever phenomenon we aim to learn about: If the goal is to understand the politics of racial identity, we should look at contexts where racial identities assume different forms. If our interest is in scope conditions of an argument, we should look for places where we think it is more and less likely that the argument will hold. When probing for mechanisms of an observed pattern we should pick cases where our suspected mechanisms may or may not be in place. And when we want to understand more about the meaning of a concept in different circumstances (like the example we give below), we select contexts where our main concept of interest (“empowerment”) seems to be used in different ways.

Interestingly, the approach of choosing interesting variation in whatever you want to know more about might actually bring us back to controlled comparison, as selecting on potential dependent or independent variables may be one way of creating interesting variation. In fact, as Van Evera (1997, p. 68) points out, controlled comparison can be great for theory development if one has a fairly good idea about key variables. But again, the purpose of this selection strategy is then not to be able to draw causal conclusions based on the study of two or more places—which, as Seawright (this volume) makes very clear, is a terrible way of trying to make causal inferences—but to gain comparative insight and develop theory.

Selecting contexts not cases

In our experience, many students struggle with the idea that their qualitative studies have to fit into the framework of “cases” and “cases studies”, and that a case selection has to be defended on the basis of variation in key variables. The standard questions: “what is this a case of?” and

“aren’t you selecting on your dependent variable?” often result in a moment of existential crisis and the realization that providing a clear answer to skeptics is not so easy.

Confusion arises partly from contradictory usage of the term “case” in different texts and even within the same text. A common understanding is that cases can be understood as “specific instances of a clearly defined class of events” (George and Bennett 2005)—the French revolution is a case of a revolution, while the adoption of the Colombian constitution of 1991 is a case of a massive legal change. Yet as Eckstein (1975, p. 85) explains, a study of six elections in England can be a study of $N=1$, $N=6$, or $N=120,000,000$, depending on whether the subject of study is electoral systems, elections, or voters. To avoid confusion, Eckstein suggests using the term “case” to refer to a phenomenon for which only a single measure is made. King et al (1994, pp. 52-53) observe, however, that Eckstein’s solution goes against common usage. They suggest using “observation” to refer to a single measure and “case” more loosely to refer to a single unit for which a large number of observations is usually made.

If your goal is to test a theory, you must be clear about what your unit of analysis is and what universe of cases you are selecting from. However, clarity in defining cases is not as important for theory development. For example, our research is often inspired by novel political phenomena that we feel compelled to study. At the outset, we are usually not sure what it is a case of, nor what the universe is, because it is not yet clear what we are studying (see Ragin and Becker 1992). As Soss (this volume) explains, qualitative fieldwork is more about “casing a study” than “studying a case.” Gaining awareness of what the object under study is a case of, and how it fits into existing bodies of knowledge, is a central part of the research process itself.

How then should we talk about the work that we do? Levy (2008, p. 2) defines a case study as an “attempt to understand and interpret a spatially and temporally bounded set of events.” This comes close to how many qualitative researchers seem to use the term and closer to our own intuitive understanding. However, Levy’s definition does not provide a solution to the main challenge. Do we know, and are we able to define, the nature, most salient elements, and boundaries of the set of events *before* we undertake the study, or do we know this only *during* or *after* we are progressing in the research?

In practice, we often choose to study something because it seems inherently interesting and important, even if it is not entirely clear what it is a case of. One way of thinking about this is that “case-oriented studies start with the seemingly simple idea that social phenomena in like settings [...] may parallel each other sufficiently to permit comparing and contrasting them” (Ragin 2004, p. 125). Similarly, we may look for “resonance with other cases in different localities or different times” (Lund 2004, p. 226). These are looser understandings of case selection than what is commonly taught to students, but they are closer to the actual research process of most qualitative scholars.

Later, as we gather more information about the topic, we are in a better position to say what it is “a case of”—how it fits into existing theory and understanding—a process Soss (this volume) describes as “casing a study.” Often, we need to embark on and become deeply involved in a study before we are able to identify the universe of cases to which it belongs or what variables to focus on. That is why, as Van Evera (1997, p. 78) says eloquently, decisions on case selection are premature before you know what you want to know.

In our experience, when we are in the situation of approaching a study without a clear sense of the universe of cases it is often more productive to think of selecting *contexts* rather than *cases*. Particularly if the goal of the project (or a part of the project) is to get a sense of relevant variables, understand more about a certain concept (as our example below), explore the scope conditions of an argument, probe for mechanisms, or even to figure out what research questions are worth asking. Comparing the phenomenon of interest across various contexts has the benefit of making it clearer what is unique or puzzling in one place compared to another.

The sources of theoretical inspiration

The previous discussion brings us to the fourth misperception we often encounter among students: the perception that theory development is something that comes mainly from moments of inspiration, from discovering inconsistencies in the literature, or from fieldwork in a single case—not from comparative work. These are, of course, great potential routes to theory development. For example, Seawright and Gerring (2008) provide an excellent overview of how single case studies can be used to achieve various goals that form part of developing theoretical

insights. This advice can also be extended to comparative work. Comparative studies are, in fact, great for theory development. They are excellent for generating hypotheses (Lijphart, p. 691), for improving conceptualization and measurement (Collier et al 2004, p. 100), or identify antecedent conditions for one's theories (Van Evera, 1997, p. 55).

In our experience, comparative fieldwork is actually one of the best ways to develop theoretical intuition—understanding what questions to ask, which concepts and variables are important, how they may hang together under what circumstances, and perhaps what the phenomenon under study is a “case of”. When trying to understand a new phenomenon, it may be hard to grasp what is interesting and unique about the circumstances in which it occurred unless you have a good point of comparison (see Schaffer, this volume, on perspectival comparison). As mentioned above, exploring the same phenomenon across multiple contexts helps you identify what features are important. When you look at one context from the perspective of other contexts, some things stand out, others don't. Things that people take for granted in one place are seen as shocking in another.

Comparative race relations provides an example of the theoretical payoff of comparative work. To many people raised in the United States, the U.S. system of racial identity seems normal and obvious. But when we look at how American society classifies people racially—who it considers Asian, black, or white, for example—from the perspective of Brazil or Colombia, the U.S. system seems anomalous. Many people who identify as “white” or “mixed” in Brazil would be classified as “black” in the U.S, and when they immigrate they feel misidentified. Observing the logic and mechanisms of racial classification from the perspective of different countries in the Americas helps identify key theoretical features of a system of racial identity that can guide further comparison. Is classification based on descent, or on appearance? Are mixed people classified with the subordinate status group, or are they included in intermediate categories? Is it only the government and census, or also other social organizations—hospitals, workplaces, schools, universities—that identify and organize people by race?

Comparative work aimed at theory development starts out with theoretical intuitions and ideas—perhaps even with fairly strong conjectures—but these may change as the project progresses. By contrast, comparative work intent on theory testing tends to establish the

parameters of comparison ahead of time, and risks missing important dimensions of variation. For example, large comparative datasets on ethnic diversity established by Alesina et al assume, rather than show, that ethnic identity is based on descent, and classify groups accordingly in order to generate data about their number and distribution (see discussion in Htun 2016). Yet in most of Latin America, ethnic identity, as well as racial identity, is based not on descent but on appearance and behavior. Though 44 percent of Peruvians stated in the PERLA survey that they had an indigenous ancestor, only 20 percent identified with one of the country's indigenous groups, and only 10 percent identified with the category "indigenous" (Telles et. al. 2014). This example suggests that it may be wise to conduct a deep dive into the subject of investigation before establishing the parameters of comparison and operationalizing important concepts.

Balancing theoretical and practical concerns

Finally, we turn the question of how to balance theoretical and practical concerns when choosing what to study. Koivu and Hinze (2017) emphasize the lack of attention, in methods texts, to the "human element" of selecting what to study based on a researcher's prior knowledge and skill-set. They point out that people see personal reasons for case selection as almost unprofessional. Texts, advisors, and colleagues seem to say that our motivation for picking cases to study should be based on theoretical concerns alone.

The human element is crucial, including language skills, local knowledge, data access, or a passionate interest in the topic of study. As pointed out by King et al (1994, p. 37), for people engaging in fieldwork, knowing a great deal about a culture and context prior to formulating research questions is important. They direct this advice toward "interpretivist" work, but the same holds for most research, and certainly for most theory development. How can someone with little knowledge of a context say something important and persuasive about it? Theories can of course travel, but, in the study of politics, contextual idiosyncracies may still be important. And when conducting fieldwork, the insider perspective may be crucial to understand nuances in actions and responses, or norms that are taken so much for granted that they remain unstated.

At the same time, there is great learning in being exposed to new ideas and important advantages to having an outsider's perspective. It is also problematic to limit yourself to studying

the same place, or the place you are from, again and again because of ease of access and cultural comfort. Sometimes theoretical intuition can guide us to new places that seem to offer a valuable contrast to that which we are familiar. The ideal, in our minds, is therefore to try to find a good balance between theoretical and practical considerations, to neither default to the same locality nor to ignore the value of language skills, contextual knowledge or other personal or practical concerns.

An evolving research project on “Empowerment”

In one of our ongoing research projects we aim to understand more about the effects—and limitations—of different types of state action on women’s empowerment. As a result of our prior research, we were somewhat disenchanted with research focused on top-down laws and policies alone. Htun’s work on the “rights revolution” for women, for example, left unanswered whether policy changes to combat violence and harassment reduce discrimination at work, promote equality in the family, or improve public support for caregiving actually produced changes on the ground (Htun and Weldon 2018). Jensenius’ work on marginalized communities in India shows that, despite decades of quotas in politics, educational institutions, and jobs, and a slew of programs aimed at improving their socio-economic status, historically stigmatized groups are still disadvantaged both socially and economically (Jensenius 2017). These combined experiences make it clear that participation in political parties and elected office, formal laws ensuring equal rights, and other state efforts to change entrenched social inequalities may be necessary but are far from sufficient to deliver inclusion and justice to disadvantaged groups.

What does it take for legal changes to lead to empowerment? Existing research has shown that a principal driver of women’s empowerment is not the law but rather women’s economic agency, or their disposition and capacity to make autonomous economic choices. Economic agency enables women to contest oppressive gender norms and change gender relations from the ground up. For example, women’s labor force participation increases their political participation, shifts the division of labor in the household, and improves fertility rates in the Global North, while reducing them in the Global South. Women who control property can

exit, or threaten to exit, abusive relationships, and exert more control over institutions that shape gendered ideologies, such as schools, the media, and religious organizations (Agarwal 1994; Hashemi et al. 1996; Hakim 1996; Agarwal 1997; Iversen and Rosenbluth 2008; Rosenbluth 2006; Okin 1989; Duflo 2012; United Nations 2015; Panda and Agarwal 2005).

An important question, then, is how laws can enhance women's economic agency. To explore this, we worked with the Women, Business, and the Law dataset developed by the World Bank, which includes information about a large variety of legal provisions in 173 economies (World Bank 2015). Following the multi-dimensional approach to gender outlined in Htun and Weldon (2018), we developed indices on constraining and enabling laws, including restrictions on women's legal capacity, discrimination in the workplace, and the extent to which the state promotes work-life balance (Htun et al. 2019). Using these indices, we explored the correlation between legislative choices and various macro-level indicators of women's economic agency: access to bank accounts, participation in firm ownership, participation in the labor force, share of women workers in the informal sector, and the gender wage gap. This analysis reveals, not surprisingly, that countries with fewer restrictions on women's legal capacity tend to have higher numbers of women with bank accounts, more firms where women participated as owners, and higher female labor force participation (Htun et al. 2019). However, the associations between our indicators of agency and laws regulating women's work and publicly-paid parental leave are weaker. Moreover, the great variation in women's agency among countries with similar legal environments makes it clear that laws tell only a small part of the story. A great deal of action shaping patterns of women's agency takes place *within* countries, and is hard to see when we conduct studies *across* countries. We needed to conduct within-country research to understand this variation better.

Our cross-country work thus led us toward a different line of investigation, focused on understanding more about the effects of legal changes and policy interventions on gender norms, and women and men's behavior within countries. But this raised new questions. The experience of rich countries today shows that even when liberal laws and access to resources bring about considerable economic agency, there are still many challenges. Across the Global North, wealthy, elite women suffer from sexual harassment and abuse, less pay for the same work,

biased and discriminatory treatment, and a glass ceiling blocking their rise to leadership, and trouble juggling the demands of family and care work. Moreover, there are many ideas of what an “empowered” life looks like, and considerable inequality among women. Rich women “lean on” poor women by outsourcing their care work (see Fraser 2016), and many women on the lower end of the distribution of income and wealth are consequently locked in a struggle to provide for themselves and their families.

If this is empowerment, is it so great? And if not empowerment, what is empowerment anyway? Does empowerment imply that women occupy the same roles as men? That households conform to a dual-income, co-participation in care work model? Is empowerment about everyone doing what they want, even if it means adhering to strict and traditional gender roles? If empowerment is not about equality, why do we care about legal reform, political participation, social mobilization, and other mechanisms to promote it? These questions led us to decide that we wanted to do more work on the concept of empowerment, what it means in different contexts, and how these different ideas animate the ways that states and societies put empowerment into practice. In other words, although our overarching research interest is to study the role of the state in bringing about empowerment, this particular part of our research aims to understand more about variation in the *meaning* of empowerment, since these different meanings may shape state interventions and their effects.

Comparing to Explore the Boundaries of a Concept

To explore differences in the meaning of women’s empowerment—which may or may not matter for the main causal patterns we are interested in—we chose a comparative approach. We decided to take a deeper look at three wealthy countries that have relatively few formal restrictions on women’s legal capacity and little state-sanctioned discrimination in the workplace and where most women seem to be fairly economically empowered, but with seemingly different discourses about empowerment: Norway, Japan, and the United States. Each country holds a particular appeal. Following much of the gender and politics literature, Htun had long seen Norway as the “paradise” of gender and social equality, due to its extensive social provision, relatively high degree of class equality, and the widespread commitment of virtually all political

actors to gender justice. The United States is compelling due to its many puzzles and contradictions, including a stratification of gender equality by class groups (Putnam 2016; Esping-Andersen 2009) and inconsistent progress toward gender justice by issue area (Htun and Weldon 2018). Japan stands out among the rich countries for the resilience of traditional social norms in the face of policy and rhetorical changes to promote women's labor force participation, men's caregiving, and greater work-life balance for all (Estévez-Abe 2013; Nemoto 2016).

Some people may say our study resembles a controlled comparison of similarly-placed countries that differ in their state approach to empowerment. However, we did not choose to study these three countries because we wanted to make a causal argument about the effects of one state's approach compared to the others'. Nor do we believe they are similar in all other ways besides their legal approach to the empowerment of women. Following the framework of Gerring (2017, 41), the "case-selection strategy" that most resembles our approach is the analysis of a set of cases that are diverse with respect to their descriptive features. However, consistent with our discussion above, the "cases" in our study are not units from a well-defined universe of a class of events. We picked them because they are vastly different contexts which offer interesting variation that furnish insights into women's empowerment. Our goal is theory development, not "description." Further, we did not choose our diverse contexts on the basis of values on a single descriptive feature. Our prior knowledge of the many characteristics of these countries made us think that they would provide us with interesting stories about the content and boundaries of the empowerment concept. As such, our choices fit somewhat uneasily within this common framework describing different types of "case selection." We also do not think of this study as a comparative "case study", but rather as a conceptual comparison across three contexts.

Since our goal is theory development, and to understand more about the concept empowerment in particular, we sought interesting variation in how empowerment seems to be understood and used to inform our theoretical thinking. Each context we study in this project has helped to bring out particular and unique features of the others, while revealing connections within contexts we thought we already knew well. By conducting fieldwork in each context, we have gained a better sense of the nature and degree of social contestation over women's empowerment. We have seen how gender, class, and race equality relate to one another in

different ways. We have learned that people's views on women's empowerment are connected to their views of the good life, as well as how they evaluate the proper role of the state in citizen's lives and as a mechanism for social coordination. We have traced the connection between government goals vis-à-vis women's liberation and other state priorities, and observed how these goals sometimes compete and sometimes complement each other. In addition, the fieldwork has alerted us to our own cultural biases related to women, gender, the state, and the economy (more on this below).

Multi-cultural Research Teams and Collaborative Fieldwork

Our selection of three contexts was not based purely on theoretical concerns. We also consciously designed our study to maximize our insider and outsider advantages. In the United States, we focused on New Mexico, where Htun has spent much of her life. Jensenius has spent considerable time in the United States, but is quite unfamiliar with the New Mexico context, and her accent makes it evident that she is a foreigner. In Norway, we worked in Oslo which was a new context for Htun but where Jensenius grew up. In Tokyo, Jensenius was clearly an outsider, Htun was somewhat more acculturated due to a one-year fellowship in 2006-2007, and we recruited a Ph.D. student from Yokohama, Melanie Sayuri Dominguez, as a collaborator.

By conducting fieldwork collaboratively, we could leverage both an insider and an outsider advantage in real time. As insiders, we had easier access to sources, command of the native language, and greater understanding of subtle cultural cues. As outsiders, we were able to pose out-of-the-box questions and notice patterns and particularities that an insider rarely thinks of.

Including both insider and outsider perspectives on the same team, at the same time, allowed us to push further in interviews and in participant-observation situations than we would have been able to do on our own. It made us familiar with the narratives that people commonly choose to share with dissimilar others, while at the same time allowing us to achieve a level of intimacy that made respondents more open and candid as the conversation progressed.

What is more, this research technique helped us become more aware of and challenge some of our own cultural biases that shape our research. We were, for instance, fascinated to

discover the extent to which Norwegians (including Jensenius) take the role of the state for granted when it comes to upholding a common image of the good life and structuring people's lives. Htun, while enamored with the generous welfare policies for working parents, reacted intuitively with more skepticism to the ways in which the state's one-size-fits-all policy solutions limit individual choice. Meanwhile, Jensenius questioned the ideal of a leaning-in form of feminism that Htun takes more for granted, as well as the dichotomous choice—between career and care work—many women face. For Jensenius, an ideal of “empowerment” that implies outsourcing care work to other women, such as low-income immigrant women, seemed unattractive and unjust.

Concluding Thoughts

Since the publication of KKV's *Designing Social Inquiry* in 1994, much of the research design advice passed to graduate students as part of their methodological training has presumed that their general goal is to test theories about a causal relationship among variables. It is critical to design research aimed at theory testing carefully to avoid bias. However, most empirical studies involving the in-depth comparison of a few cases or contexts have the goal of theory development rather than theory testing. And theory development consists of many different stages and parts, including building intuition and contextual knowledge, conceptualizing and operationalizing key variables, exploring causal mechanisms, scope conditions of arguments, gaining insight from deviant cases, and so on. In our view, comparing generates important perspectives for all of these stages and parts, and we therefore need to think carefully about research designs for multiple aspects of a project. There is no one-size-fits all solution, but rather some rules of thumb: align your research design with your research goals, be explicit about the choices you make, and aim to capture some interesting variation in whatever you want to understand better.

Clarity and transparency about these thought processes will help others see the source of your insights, many of which will likely stem from systematic research and not epiphanies. What is more, scholars don't need to know everything about their theory in advance, and so research

designs may evolve as thinking develops. It can be rigorous, transparent, and scientific to learn as you go along. In the final analysis, comparing is useful for almost any part of a research project. Comparative work helps open your mind and extend your horizons. Let its insights enrich your theory.

References

- Agarwal, Bina. 1994. *A Field of One's Own: Gender and Land Rights in South Asia*. Cambridge: Cambridge University Press.
- Agarwal, Bina. 1997. "'Bargaining' and Gender Relations: Within and Beyond the Household." *Feminist Economics* 3 (1): 1–51.
- Brady, H. E., & Collier, D. (Eds.). 2010. *Rethinking social inquiry: Diverse tools, shared standards*. Rowman & Littlefield Publishers.
- Collier, D., Mahoney, J., & Seawright, J. 2004. Claiming too much: Warnings about selection bias. *Rethinking social inquiry: Diverse tools, shared standards*, 85-102.
- Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature* 50 (4): 1051–79.
- Esping-Andersen, Gøsta. 2009. *Incomplete Revolution: Adapting Welfare States to Women's New Roles*. Cambridge: Polity Press.
- Estévez-Abe, Margarita. 2013. "An International Comparison of Gender Equality: Why Is the Japanese Gender Gap So Persistent." *Japan Labor Review* 10 (2): 82–100.
- 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: MIT Press.
- Gerring, John. 2017. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Hakim, Catherine. 1996. *Key Issues in Women's Work: Female Heterogeneity and the Polarisation of Women's Employment*. London: The Athlone Press.

- Hashemi, Syed M., Sidney Ruth Schuler, and Ann P. Riley. 1996. "Rural Credit Programs and Women's Empowerment in Bangladesh." *World Development* 24 (4): 635–653.
- Htun, Mala, Francesca R. Jensenius, and Jami Nelson-Nuñez. "Gender-discriminatory laws and women's economic agency." *Social Politics: International Studies in Gender, State & Society* 26, no. 2 (2019): 193-222
- Htun, Mala, and S. Laurel Weldon. 2018. *The Logics of Gender Justice: State Action on Women's Rights Around the World*. New York: Cambridge University Press.
- Iversen, Torsten, and Frances Rosenbluth. 2008. "Work and Power: The Connection between Female Labor Force Participation and Female Political Representation." *Annual Review of Political Science* 11: 479–495.
- Jensenius, Francesca R. 2017. *Social Justice through Inclusion: The Consequences of Electoral Quotas in India*. Oxford: Oxford University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- King, G., Keohane, R. O., & Verba, S. 1995. The importance of research design in political science. *American Political Science Review*, 89(2), 475-481.
- Koivu, Kendra L. and Hinze, Annika Marlen. 2017. "Cases of Convenience? The Divergence of Theory from Practice in Case Selection in Qualitative and Mixed-Methods Research." *PS: Political Science & Politics* 50 (4): 1023-1027.
- Lund, C. 2014. Of what is this a case? Analytical movements in qualitative social science research. *Human organization*, 73(3), 224-234.
- McKeown, T. J. 2004. Case studies and the limits of the quantitative worldview. *Rethinking social inquiry: Diverse tools, shared standards*, 139-167.
- Mill, J. S. (1884). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation* (Vol. 1). Longmans, green, and Company.
- Nemoto, Kumiko. *Too few women at the top: The persistence of inequality in Japan*. Cornell University Press, 2016
- Okin, Susan Moller. 1989. *Justice, Gender, and the Family*. New York: Basic Books.

- Panda, Pradeep, and Bina Agarwal. 2005. "Marital Violence, Human Development and Women's Property Status in India." *World Development* 33 (5): 823–850.
- Putnam, Robert D. 2016. *Our Kids: The American Dream in Crisis*. New York: Simon & Schuster.
- Ragin, Charles C. 2004. Turning the tables: How case-oriented research challenges. *Rethinking social inquiry: Diverse tools, shared standards*, chapter 8, pages 123-138.
- Ragin, C. C., & Becker, H. S. (Eds.). 1992. *What is a case? Exploring the foundations of social inquiry*. Cambridge University Press.
- Rosenbluth, Frances M. 2006. *The Political Economy of Japan's Low Fertility*. Redwood City: Stanford University Press.
- United Nations. 2015. "Progress of the World's Women 2015-2016: Transforming Economies, Realizing Rights." <http://progress.unwomen.org/en/2015/>.
- Van Evera, S. 1997. *Guide to methods for students of political science*. Cornell University Press.
- World Bank. 2015. *Women, business, and the law 2016: Getting to equal*. Washington, DC: World Bank.