

The Enduring Indispensability of the Controlled Comparison

Comparative Political Studies
46(10) 1301–1327
© The Author(s) 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0010414012472469
cps.sagepub.com



Dan Slater¹ and Daniel Ziblatt²

Abstract

Do controlled comparisons still have a place in comparative politics? Long criticized by quantitatively oriented methodologists, this canonical approach has increasingly been critiqued by qualitative methodologists who recommend greater focus on within-case analysis and the confinement of causal explanations to particular cases. Such advice accords with a welcome shift from a combative “tale of two cultures” toward mutual respect for research combining qualitative and quantitative methods in the simultaneous pursuit of internal and external validity. This article argues that controlled comparisons remain indispensable amid this “multimethod turn,” explicating how they too can generate both internal and external validity when their practitioners (a) craft arguments with general variables or mechanisms, (b) seek out representative variation, and (c) select cases that maximize control over alternative explanations. When controlled comparisons meet these standards, they continue to illuminate the world’s great convergences and divergences across nation-states in a manner that no other methods can surpass.

Keywords

controlled comparisons, comparative methods, qualitative methods, mixed methods, fascism, authoritarianism, Southeast Asia

¹University of Chicago, Chicago, IL, USA

²Harvard University, Cambridge, MA, USA

Corresponding Author:

Dan Slater, Department of Political Science, University of Chicago, Chicago, IL 60637, USA.

Email: slater@uchicago.edu

Unfortunately, practically all efforts to make use of the controlled comparison method fail to achieve its strict requirements. This limitation is often recognized by investigators employing the method, but they proceed nonetheless to do the best they can with an admittedly imperfect controlled comparison. They do so because they believe that there is no acceptable alternative and no way of compensating for the limitations of controlled comparison.

—Alexander George and Andrew Bennett (2004, 152)

Do controlled comparisons still have a place in comparative politics? At first glance this appears a peculiar question. After all, the logic of controlled comparisons has been one of the defining methodological orientations of comparative politics for generations. Nearly all graduate courses on comparative politics commence with a discussion of Mill's (1843/2002) classic methods of "difference" and "agreement" as well as their further elaboration by Lijphart (1975) and Skocpol and Somers (1980) on the comparative method. Even a cursory examination of leading journals and university presses reveals the enduring ubiquity of controlled comparisons, in which a researcher strategically selects cases for analysis that either exhibit contrasting outcomes despite their many otherwise similar characteristics, or similar outcomes despite their many otherwise contrasting characteristics.

Indeed, it is the puzzling and dazzling array of divergences and convergences across nation-states in the modern world that has long drawn scholars to the craft of comparative politics in the first place. These puzzles make controlled comparisons not merely an impassionate method of scientific inquiry, but a comparativist's labor of love. Why did economic modernization pave a pathway to fascism in Germany and Japan, liberal democracy in England and the United States, and communism in China (Moore, 1966)? Why do Latin American countries exhibit such dramatic variation in their patterns of political representation and levels of economic development, despite their common Spanish colonial heritage (R. B. Collier & Collier, 1991; Mahoney, 2010)? Why did the most powerful territorial unit in Germany unify the nation-state through federalism, whereas its Italian counterpart crafted a unitary nation-state (Ziblatt, 2006)? Why did India emerge from the British colonial raj as a durable democracy, unlike Pakistan and Burma (Callahan, 2003; Tudor, *in press*)? Why did geopolitical competition give rise to patrimonial states in some parts of Europe, and bureaucratic states in others (Ertman, 1997)? Why have developmental states emerged in some but not all Asian countries, and in no Middle Eastern countries whatsoever (Waldner, 1999)? Why did some

former communist parties in Eastern Europe regenerate themselves from the ashes of Soviet occupation and domination after the Berlin Wall came down, whereas others became obsolete (Grzymala-Busse, 2002)? And why have so many authoritarian regimes in the postcolonial world survived the end of the Cold War whereas many others have collapsed (Brownlee, 2007; Levitsky & Way, 2010; Slater, 2010)?

Besides such great historical divergences, controlled comparisons also seek to apprehend modernity's great *convergences*, in which highly disparate cases experience surprisingly similar macro-processual outcomes. Why did great social revolutions from below occur in cases as dissimilar as 18th-century France, interwar Russia, and post-World War II China (Skocpol, 1979)? Why have revolutions occurred, quite contrarily, "from above" in cases ranging from Japan to Turkey to Peru (Trimberger, 1978)? Why do states have so much trouble asserting authority and establishing social control in every diverse corner of the postcolonial world, from India to Egypt to Sierra Leone (Migdal, 1988)? Why and how has nativist exclusion displaced more inclusive imperial forms of identity politics in cases as radically different as Iraq, Mexico, and Switzerland (Wimmer, 2002)? Why did democracy similarly emerge, yet also similarly falter, in Iran, China, Russia, Portugal, Mexico, and the Ottoman Empire at the turn of the 20th century (Kurzman, 2008)? And why did "insurgent transitions" forge a far sturdier form of popular democracy in late-20th-century El Salvador and South Africa, despite those cases' countless obvious differences (Wood, 2000)?

It is one thing to be ubiquitous, however, and quite another to be indispensable. And controlled comparisons have found themselves in an oddly lonely position in recent years, their enduring value to political science sharply challenged from two very different sides.¹ First, they have been criticized by quantitatively oriented scholars, quite fairly at times, as insufficiently rigorous or for requiring such restrictive epistemological assumptions that they do not justify their authors' grandest ambitions of generalizability (Geddes, 2003; King, Keohane, & Verba, 1994; Sekhon, 2004). More surprisingly, the method has also increasingly been criticized by a new generation of qualitative methodologists. These critiques fit within an emergent "multi-method turn" in comparative politics that entails growing endorsement among both qualitative and quantitatively oriented scholars for a particular methodological formula: Deep studies of particular country cases for purposes of establishing "internal validity," combined with broader large-*N* analysis to ensure "external validity."

In the following we argue that this methodological mix can offer an excellent way, but by no means *the only* way, to generate both internal and external

validity in comparative politics. We see the multimethod turn as offering a useful opportunity—and inciting an important challenge—to reconsider and elaborate the value of controlled comparisons, and to elaborate systematically how this research design can generate both internal and external validity under certain restrictive conditions. In so doing, our goal is emphatically to expand and not to constrict the repertoire of viable approaches to causal analysis in comparative politics. More than 150 years after Mill introduced their logic and more than three decades since Lijphart (too vaguely) defended their value, an updated explication of the indispensable place of the controlled comparison in the rapidly evolving discipline of political science is long overdue.

Such an intervention is necessary because, in our view, existing writings either mischaracterize the purpose and practice of controlled comparisons or sell the method short by failing to appreciate (much less explicate) its potential for building generalizable claims. For instance, George and Bennett's (2004) recent influential methodological work criticizes controlled comparisons for simply showing correlations across a small set of cases, without reference to the rich narrative and process tracing that in practice *always* accompanies the best work in this tradition. Hence they mistakenly consider process tracing an "alternative" to controlled comparisons rather than one of its defining features. According to this caricatured but commonly held view, controlled comparisons are nothing more than a weak form of statistical inference—a view Lijphart effectively debunked back in 1975²—rather than a distinctive method with distinct advantages. Indeed, George and Bennett do not discuss any strategies or principles for pursuing external validity in qualitative research whatsoever. Similarly, Tarrow's (2010) qualified defense of what he calls "paired comparisons" does not portray or position this method as a viable alternative to multimethod approaches for generating both internal and external validity. In part this is because Tarrow seems to accept the assumption that controlled comparisons entail nonrepresentative sampling, rather than considering how case selection might be done more systematically to improve prospects for generalizability.

No work to date has systematically explicated the best practices through which controlled comparisons advance portable causal claims. This is the primary task for our analysis to follow. First, we question the presumed elective affinity between types of analysis and types of validity, suggesting that large-*N* analysis may not be as indispensable for attaining externally valid results—or even as geared for doing so—as commonly assumed. Second, we explicate the potentialities of controlled comparisons to generate causal arguments that are at once internally and externally valid, absorbing the insights of recent work in qualitative methodology to reassert the enduring indispensability of

the controlled comparative method. We then illustrate our argument by exhibiting the impressive portability of the controlled comparison offered by Luebbert (1991) on the origins of fascism in interwar Europe. In sum, we argue that controlled comparisons remain indispensable in political science amid the multimethod turn because our defining goal as social scientists should be to capture multiple types of validity (i.e., internal and external), not merely to deploy multiple types of methods (i.e., qualitative and quantitative).

Multiple Validities Versus Multiple Methods

The vast majority of political scientists would presumably agree that, at its best, causal inference produces empirical results that are both internally and externally valid. Any causal argument should ideally promise to tell us something new about the wider world—or at a minimum inspire us to ask new questions or think differently about the wider world—while doing justice to the particularities of the case or cases directly under study. The tension between these goals is especially acute for comparativists, who tend to face greater reproach for ignoring case particularities when crafting theoretical claims than their colleagues studying American politics or international relations.

Before going further, we should be clear about how we define internal and external validity. In the sciences generally, the definitions for these terms are uncontroversial. As Gerring (2007) usefully details,

Internal validity refers to the correctness of a hypothesis with respect to the sample (the cases actually studied by the researcher). External validity refers to the correctness of a hypothesis with respect to the population of an inference (cases not studied). The key element of external validity thus rests upon the representativeness of the sample. (p. 217)

These basic definitions could be fruitfully employed in any field of scientific study, from astronomy to zoology. Internal validity refers to the robustness of the analyst's causal inferences within a sample, external validity refers to inferential robustness in that sample's broader population, and the representativeness of one's sample is of the essence in determining the external validity of one's hypothesis. No reasonable definition of internal and external validity can depart from these fundamental precepts.

Yet in the comparative politics subfield, it is important to recognize that a more colloquial understanding has also come to be associated with these two types of validity: namely, as the validity of a causal argument within a

single country case (i.e., internal), as opposed to beyond that country case (i.e., external). Hence when we say in this essay that a hypothesis enjoys external validity, we mean *both* that it holds true in more than one country case, *and* that the additional country case(s) played no role in helping to generate the hypothesis in question. By this understanding, when a hypothesis generated in part through the study of empirics from a single-country case study proves valid in any other country case study, it enjoys at least some degree of external validity. To be sure, subnational evidence can often establish the external validity of a hypothesis in its standard sense, without requiring cross-national evidence. Yet external validity as discussed here requires both that an argument prove valid in multiple countries (i.e., the colloquial comparativist understanding) and that at least some of those cases have “out-of-sample” status (i.e., the standard definition of external validity). Assessing external validity in a controlled comparison thus requires transparency on the part of the researcher as to whether particular case studies preceded or followed the generation of a theoretical argument.³

For reasons that have been much discussed elsewhere, comparative political research involves a tension between the pursuit of internal and external validity. To construct a causal argument that does justice to a particular case, many variables are typically required. Who, after all, believes that there was one and only one reason why Nazism arose in Germany, or China had a communist revolution? Yet when crafting a hypothesis that applies to many cases, “parsimony” rather than “accuracy” is of the essence (Przeworski & Teune, 1970, p. 17). In Sartori’s (1970) terms, work that prioritizes internal validity tends to make use of concepts that sit quite low on the “ladder of generality,” whereas the key concepts in work aiming at external validity must sit higher.

The problem with this metaphor is its implication that a causal argument cannot be two things—or be in two places on the same ladder—at once. This skeptical tone has surprisingly been picked up in recent leading works defending the practice of qualitative methods. Multiple such works now suggest that qualitative analysis is ill-suited for constructing externally valid causal claims. In their depiction of qualitative and quantitative research as “a tale of two cultures,” Mahoney and Goertz (2006) conclude that researchers in these two traditions harbor distinctive explanatory goals. Although “quantitative researchers” aim to estimate the average causal effect of variables across a wide population of cases, “qualitative scholars” hope to “explain individual cases.” Similarly, Gerring suggests that internal validity is a characteristic strength of qualitative analysis, whereas quantitative analysis is ideally suited for external validity. Drawing on the qualitative/quantitative distinction, Gerring (2007) writes, “It seems appropriate to regard the trade-off between

external and internal validity, like other trade-offs, as intrinsic to the cross-case/single case choice of research design" (p. 43).

Although we certainly accept that tensions exist between internal and external validity, the claim that types of analysis and types of validity have such strong elective affinities does not hold up very well in practice in comparative politics. In many major recent multimethod comparative works, in fact, the qualitative and quantitative components perform *the opposite* tasks that Gerring, Goertz, and Mahoney ascribe to them. We will elaborate in the third section how small-*N* controlled comparisons can best aim to generate external validity. However, before doing so we think it worthwhile to provide some significant examples of multimethod research in which quantitative work ironically helps explain "particular outcomes,"⁴ whereas small-*N* comparisons generate a study's external validity by *explaining variation in outcomes across closely matched cases rather than an individual outcome in a single case*. Only after the purported elective affinity between type of analysis and type of validity is problematized will the ground be cleared for our argument that controlled comparisons can generate explanations with both internal *and* external validity.⁵

The consummate example in our view of a book that defies this elective affinity is the one that, more than any other, marked the shift in comparative politics toward studies combining multiple methods in the intensive study of a single country: Putnam's (1993) *Making Democracy Work*. Indeed, Putnam could justly be considered the progenitor of the "multiple methods, single-country case" approach that increasingly predominates in comparative research. As any attentive political scientist must already know, Putnam's central argument is that social capital and civic engagement profoundly influence government performance. The empirical setting for his theoretical argument is Italy, and Putnam adroitly combines qualitative and quantitative methods to establish social capital's effect on quality governance.

The elegance and logic of Putnam's theory explain why readers were *initially captivated* by his argument. Yet theory alone cannot explain why so many readers were *ultimately convinced* of its empirical validity. Based on the arguments of Gerring, Goertz, and Mahoney, one would suspect that Putnam's qualitative case analysis offered ample evidence of internal validity within Italy (i.e., his country of expertise and original sample), whereas his expert regression analyses must have established external validity (i.e., in the broader global population beyond Italy). Yet this is not how Putnam's multimethod analysis works. In fact, there is nothing in Putnam's quantitative analysis that establishes the external validity of his argument beyond the Italian case because his large-*N* empirics are generated entirely within Italy.

His quantitative analysis establishes the internal validity of his argument—the exogenous impact of Italian social capital on Italian government performance, independent of alternative explanations such as socioeconomic modernization. If Putnam had presented only his statistical findings, however, readers would have as much reason to question the cross-country portability of his argument as any single-country argument supported only by qualitative evidence (i.e., a paradigmatic “case study”).

Why, then, did scholars immediately pick up on the portability of Putnam’s argument? A primary reason, we submit, is that *Making Democracy Work* is an exemplary work of controlled comparison. The goal of its qualitative components is not to explain Italian government performance as an “individual outcome,” but to explain *variation* in government performance across closely matched cases: the provinces of northern and southern Italy. As Putnam (1993) himself explains the variation-driven character of his causal enterprise:

Some regions of Italy, we discover, are blessed with vibrant networks and norms of civic engagement, while others are cursed with vertically structured politics, a social life of fragmentation and isolation, and a culture of distrust. . . . The powerful link between institutional performance and the civic community leads us inevitably to ask *why some regions are more civic than others* [italics added]. (p. 15)

As we explore in the third section, what makes this research design powerful is its *theoretically informed combination of control and variation*. By working subnationally and quantitatively, Putnam controls for a wide array of national-level factors (e.g., Catholicism, parliamentarism, fascist legacies) that could plausibly influence government performance in the Italian context. Even more important for purposes of portability,⁶ Putnam broadly compares two parts of Italy whose variation in outcomes is so vast that it approximates the full range of variation in industrialized democracies; in other words, northern Italy is nearly as well run as any country or subnational region in the OECD, whereas southern Italy is among the worst managed. With no quantitative finding in Putnam’s book providing any information on social capital and government performance beyond Italy’s shores, it is above all his controlled comparison—with its combination of sophisticated theoretical argumentation, meticulous control over alternative explanations, and representative range of empirical variation—that best explains the portability of *Making Democracy Work*. Putnam’s quantitative analysis convinces us that social capital shapes government performance in Italy (internal validity), whereas his qualitative

controlled comparison raises the tantalizing prospect that his explanation for dramatic variation across Italian regions might shed light on the similarly dramatic variation in government performance that we witness around the world (external validity)—as indeed it has, though hardly without challenge or refinement (Tarrow, 1996; Tsai, 2007; Varshney, 2002).

Putnam is far from the only example of a major “multiple methods, single country” book in comparative politics in which the quantitative analysis lends confidence in internal validity whereas qualitative comparisons provide evidence of external validity. Wilkinson (2004) arrays impressive quantitative evidence in support of his parsimonious theoretical argument that electoral incentives provide a more powerful and efficient explanation for ethnic riots in India than other prominent explanations, such as intercommunal associations and consociational institutions.⁷ In a nutshell, Wilkinson theorizes and demonstrates that Muslims will be safest from Hindu-led riots in Indian states where they are an electorally pivotal constituency, because this prompts state-level politicians’ timely deployment of the state police for their protection. As with Putnam, Wilkinson’s sophisticated theory serves to captivate and his amassing of evidence serves to convince. Yet also as with Putnam, Wilkinson’s statistical evidence is entirely drawn from a single country case. His primary goal is to explain violence in the “specific case” of India, and this internal validity is generated largely through quantitative tests—again, the inverse of what Gerring, Mahoney, and Goertz assert the characteristic strength and purpose of statistical analysis to be.

Why are so many scholars convinced that Wilkinson’s argument resonates beyond the Indian subcontinent? Again as with Putnam, Wilkinson’s qualitative controlled comparisons provide more direct evidence of external validity than his statistical findings. His goal is to explain why some Indian states are so much more riot-prone than others, not simply why a particular riot broke out. Variation is again of the essence, and control is again critical in its analytical pursuit. For instance, Gujarat experiences levels of violence that rival those of highly violent nation-states, whereas Tamil Nadu is as peaceful as highly stable nation-states, capturing a range of variation in Wilkinson’s sample that broadly mirrors variation in the global population.

Wilkinson does not limit his qualitative attentions to India, either. In his concluding chapter, he considers the external validity of his argument beyond India through additional qualitative comparisons—*not* through additional quantitative tests. Through in-depth analysis of three cases strategically chosen for their variation in theoretically relevant factors besides electoral incentives (Romania, Malaysia, and Ireland), Wilkinson deploys Mill’s “method of agreement” to increase the reader’s confidence that his argument is externally valid.

One need not even look beyond Cambridge University Press's prestigious Comparative Politics series to find additional examples. Tsai's (2007) impressively multimethod treatment of China in *Accountability Without Democracy*—which establishes the importance of solidary groups in pressuring local officials to provide public goods despite any lack of electoral control over their behavior—contains no data from outside China. Yet the volume opens with Tsai's depiction of dramatic variation in public goods provision in two otherwise similar Chinese villages—that is, with a controlled comparison. Similarly, Kalyvas's (2006) *The Logic of Violence in Civil War*, rightly renowned for theorizing and analyzing civil war violence as a distinct phenomenon from civil war writ large, confines its statistical analysis to a single country case: Greece. It is Kalyvas's sophisticated theoretical framework and case material drawn from other countries that likely convinces readers that the book's arguments on the rational banality and parochial character of civil war violence have external validity beyond Greece's shores, and less the quantitative analysis *inside* the "particular case" of Greece. The same can be said of Posner's (2006) award-winning study of institutions and ethnic politics in Africa, which conducts its quantitative analyses subnationally (for internal validity within Zambia) and its concluding qualitative analyses cross-nationally (for external validity across Africa), when demonstrating his influential argument that electoral arithmetic shapes the shifting relative salience of tribal and linguistic cleavages. Finally, recent books by Scheiner (2004), on the durability of a democratic dominant party in Japan, and by Magaloni (2006) and Greene (2007), on a durable authoritarian ruling party in Mexico, duplicate this multimethod pattern: Quantitative analysis is entirely focused on the single country case under examination, whereas separate, entirely qualitative chapters and chapter sections are devoted to exploring the external validity of their explanations in multiple countries (e.g., Malaysia and Taiwan for Greene and Magaloni, Austria and Italy for Scheiner).

Whether a single-country case study primarily generates the internal validity of its causal inferences through qualitative process tracing or the statistical analysis of quantitative evidence, achieving external validity always requires "out of sample" tests. In each of the cases referenced above, these tests happened to be (but did not have to be) qualitative rather than quantitative. To be sure, to know whether an argument holds across a very wide population, one of the most powerful tools will always be to use a range of possible statistical techniques to establish "average causal effects" in a broader selection of cases (Mahoney & Goertz, 2006). Yet any work that demonstrates via a carefully selected "out of sample" test that an argument travels, even if that test is qualitative, demonstrates some degree of external validity, or what Lincoln and

Guba (1985) more precisely label “transferability.” Even though Wilkinson’s study of Romania and Scheiner’s analysis of Austria literally provide internal validity in those specific cases, these authors are conducting these case studies as qualitative tests of their arguments’ portability beyond the “sample” of their original country cases: that is, external validity.

In sum, this section has unsettled the purported elective affinity between types of analysis and types of validity—quantitative provides external, and qualitative provides internal—making the case that this accepted wisdom is overdrawn and at times misleading. Comparative scholars routinely limit their quantitative data-intensive analyses to their single, primary country under investigation, while using qualitative comparisons to speak to a broader population of country cases. Having established that many leading scholars indeed use qualitative analysis as a tool in the pursuit of external validity, we now argue that there is a pressing need to be more self-conscious about this strategy, and propose a series of “best practices” that make this difficult translation from specific cases to general patterns more systematically attainable.

Controlled Comparisons and External Validity: Variation, Theory, and Control

How exactly can controlled comparisons best generate cross-case and out-of-sample (i.e., external) validity? In this section we argue that the most portable controlled comparisons are those that meet three criteria. Though these criteria may not be controversial, more controversial is our claim that controlled comparisons can actually fulfill them.

The first is that the guiding research puzzle and reported findings should always be expressed in terms of *general variables or mechanisms*, not terms that are completely context specific. We thus endorse Przeworski and Teune’s (in)famous invocation that comparativists should “eliminate proper names” only in part—so long as “eliminate” means “reduce” and not “eradicate” altogether. Mahoney and Goertz are correct that qualitative researchers retain an active and abiding interest in particular cases; but this should not be taken to mean that researchers who know their cases well care *only* about explaining those particular cases, and not about using those cases to elaborate more general, portable theoretical claims.

The second principle for gaining external validity is to capture *representative variation*. In our view, outstanding comparative work is more often driven by a desire to explain puzzling variation in outcomes than particular cases per se. Consider the long list of controlled comparisons that we offered in our essay’s introduction. Such empirical works are most likely to generate

externally valid findings when the *variation in the sample broadly mirrors variation in some broader and explicitly defined population of cases*. As we saw above, Gerring goes so far as to *define* a sample's external validity in terms of its representativeness. Contra the frequently misunderstood implications of the recommendation to avoid selecting cases "on the dependent variable," we suggest that strategically choosing cases in search of representative variation can be one effective way to avoid the trap of selection bias. It can also allow qualitative researchers to go beyond specifying necessary conditions—as Dion (1998) convincingly argued such nonvariant research designs can accomplish—and make the case for conjunctural causal sufficiency through a controlled comparison alone.⁸

Any careful reader will immediately raise the key question of how one knows whether one's sample of cases mirrors some larger population. Indeed, this problem bedevils all social-scientific research, not only controlled comparisons. Yet it is here where we see the advantage of controlled comparisons and their attentiveness to typological conceptualization. As Collier, LaPorte, and Seawright (2012) have argued, when one is working on a topic in which previous scholarship had laid the groundwork for a rich conceptual environment, especially where outcomes are ordinal or nominal, one can identify outcomes via a strategy of what we call *typological representativeness*. Based on deep knowledge of cases and the categories scholars have used to array them, one can identify the relevant range of outcomes *ex ante* using well-accepted typologies that by definition specify mutually exclusive outcomes that also are *exhaustive* of all empirical variations.

Again, consider the list of controlled comparisons we offered in our essay's introduction—Moore's contrasting outcomes of communism, fascism, and democracy; Ertman's outcomes of bureaucratic versus patrimonial state building; or Ziblatt's contrasting outcomes of federalist versus unitary state formation. In each instance, the range of variation is established through systematic and logical process of conceptualization. Of course, this is not to say existing typologies are unchallengeable; this is precisely one appropriate task of scholarly interchange and knowledge accumulation. Nonetheless, what we dub a strategy of typological representativeness is a particularly effective way of assuring that a controlled comparison's outcomes reflect a broader population's, especially when a study's main outcomes take on nominal and ordinal values (Adcock & Collier, 2001).

A different strategy is typically required, however, when the scholar finds herself or himself in either a new empirical domain where prior conceptual work is absent, or when one is dealing with a continuous range of outcomes. Here one can begin by conducting the kind of "brush-clearing" cross-national

statistical analysis Lieberman (2005) proposes as a first step of a “nested research design” to locate individual cases in a broader universe. With this model, one can supplement and situate one’s case-specific knowledge by analyzing descriptive statistics on a broader set of cases to identify the full range of actual outcomes within a population.⁹ Once the researcher has identified the relevant variation of outcomes as well as the “scope conditions” for testing a theory, she or he can proceed with a case-selection strategy that aims at representative variation.

If variation is one blade of the scissors of controlled comparisons, the other is control. It is on this dimension where an overly literal interpretation of Mill’s methods of difference and agreement threatens to take the greatest toll. If one believes that controlled comparisons are viable only when selected cases are truly “most similar” or “most different” on every possible dimension, it is no wonder that pessimism has at times dominated discussions of this method. In our view, controlled comparisons need not meet the standard of “natural experiments,” but they do require *intense theoretical engagement* to generate external validity. The reason is not because theory is interesting for its own sake, but because theory serves an essential methodological purpose—namely, guiding case selection. Although simply choosing the cases that substantively interest one most is perfectly valid for comparativists pursuing internal validity, the pursuit of external validity requires that cases be selected precisely *to control for existing rival hypotheses*. This kind of “folk Bayesian” (McKeown, 1999) perspective pushes scholars to choose cases whose variation simply cannot be accounted for by extant hypotheses, rather than seeking the chimerical goal of a perfectly paired comparison.

In what sense do controlled comparisons guided by considerations of variation, theory, and control produce external validity? Obviously no within-sample data can ever itself confirm the out-of-sample validity of a hypothesis. We are not suggesting that controlled comparisons can perform inferential miracles. What we are suggesting is that controlled comparisons offer direct evidence of limited transferability and the theoretical foundations for wider transferability, while remaining true to what we believe has long been the guiding goal of most comparative work, especially in its classic, macro-historical vein: capturing variation in outcomes, rather than single outcomes or average, partial causal effects.

Perhaps controversially, moreover, we challenge the notion that small-*N* controlled comparisons necessarily lack external validity until a cross-national large-*N* test is conducted. In our view, such cross-national tests do not *produce* external validity; they *confirm* it. For instance, in recent work on oil wealth and authoritarian breakdown, Smith’s (2007) original argument

derived from a controlled Indonesia–Iran comparison proved to hold true in a quantitative test of 107 developing countries over a 40-year period. This test did not *make* Smith's argument externally valid, but rather *showed* that his immaculately crafted controlled comparison had produced externally valid findings before a single regression analysis was run.

In a political world marked by equifinality and multiple causation, probabilistic statistical significance simply cannot serve as the solitary viable standard for external validity. If an argument deriving from a controlled comparison is stated in terms of general variables and can be shown to shed explanatory light on specific cases outside the original sample, then the original argument can be said to enjoy external validity. Naturally, transferability is a matter of degree. An argument that holds true only in one additional case, or fails to apprehend causal mechanisms in those cases where it seems to be confirmed, is less externally valid than one that explains dozens of cases or enjoys impressive verisimilitude on causal mechanisms. Nevertheless, we find it too restrictive to deny external validity to causal claims whenever they are not supported by cross-national large-*N* tests. We also find the use of qualitative cross-case comparisons amenable to the practice of a kind of comparative politics that treats both internal and external validity as equally important, rather than privileging either type of validity to the detriment of the other. The next section provides a powerful example of the kind of research we have in mind.

Putting Controlled Comparisons to the Transferability Test

We now turn our attention to a concrete example of a controlled comparison that proves useful in explaining variation in outcomes beyond an original sample. Luebbert's controlled comparison capturing representative regime variation in interwar Western Europe proves extraordinarily useful for explaining the rise of reactionary mass movements and murderous right-wing regimes in Cold War Southeast Asia.

Interwar European Fascism and Cold War Asian Pogroms: The Urban Left in the Countryside

Luebbert's (1991) *Liberalism, Fascism, or Social Democracy* is an illuminating exemplar of representative variation, deploying a controlled comparative method that generates a theory with impressive theoretical reach. In a single class-coalitional framework, Luebbert sets out to explain why

liberal democracy survived in three interwar European cases (Britain, France, and Switzerland), while transforming into social democracy in four (Denmark, the Netherlands, Norway, and Sweden) and collapsing into fascism in three others (Germany, Italy, and Spain).¹⁰

Luebbert not only generates an argument that appears to establish a convincing causal argument for the cases at hand (i.e., internally valid) but also proposes an argument that, as we will see below, is generalizable (i.e., externally valid) insofar as it offers causal arguments that travel remarkably well to different places and times. How does he achieve this balance in a single study? Of course writing long before the “multimethod turn,” in comparative politics, Luebbert does not deploy the now paradigmatic mix of a single national case study and large-*N* cross-national study. Indeed, had he used the latter method, it is hard to imagine he would have developed the innovative argument that he does. Instead, he adopts the core elements of the simple but powerful arsenal we have identified as defining the classic controlled comparison.

First, to a greater extent than Moore, Luebbert self-consciously crafts his argument using general variables and mechanisms that abstract away from case particularities to identify similarities and differences across cases.¹¹ He searches for “master variables” (p. 5) that include the nature of, and barriers to, working class–middle class coalitions (what he calls Lib-Lab coalitions in the European context), and relations between urban elites and a society’s rural population—all categories that not only work in his own cases but also can apply to any society undergoing modernization. Second, using a self-consciously robust typological strategy of conceptualization based on a close knowledge of cases and their historiography, he identifies four distinct outcomes for the period he studies. By applying his arguments not only to the Western European cases of fascism, social democracy, and liberal democracy but also to a fourth, logically consistent outcome of Eastern European cases of traditional dictatorship, Luebbert attains sample variation approximating the full population of European cases.

In addition, Luebbert establishes the greater internal and external validity of his argument vis-à-vis competing explanations through careful process tracing and original cross-national comparisons. For example, he notes the weakness of one common hypothesis, that landed elites were in and of themselves an automatic barrier to democracy, by examining where within countries fascists received their most votes. He finds that it was first in northern Italy and Spain and western Germany, not where landed estates were dominant but rather where family farms were, that fascists made their greatest inroads, thereby challenging conventional accounts (p. 309). This potent combination of within-case and cross-case evidence simultaneously bolsters

both the internal and external validity of his claims, and serves as an example of the controlled comparison at its best.

For our purposes, the most revealing of Luebbert's causal arguments relates to the *causal mechanism* through which fascism originated as a regime type distinct from traditional dictatorship. According to Luebbert, a fascist outcome was not the result of highly concentrated landholdings or labor-repressive agriculture, but was in large part a right-wing reaction to efforts by urban socialists to mobilize the masses in the countryside (pp. 295-303). Indeed, it was precisely this failure of a "Red-Green" coalition in Germany that led to Germany's divergence from the stable Scandinavian social democratic interwar outcome.

Perhaps surprisingly, this seemingly region-specific explanation resonates in the geographically and temporally distant context of Cold War Southeast Asia. There, as in interwar Europe, fledgling electoral democracies crumbled between the mid-1950s and early 1970s, primarily to be replaced by right-wing, military-dominated successors. Yet also as in interwar Europe, democratic breakdown occurred quite differently in what would otherwise appear to be quite similar cases. Of most interest to our discussion of Luebbert here, violent right-wing regimes rose to power with the direct and murderous assistance of reactionary elements in the countryside in response to a burgeoning leftist threat (i.e., Luebbert's fascist pathway) in two specific cases: Indonesia (1965–1966) and Thailand (1976). Interestingly, neither of these cases is characterized by highly concentrated landownership, lending added credence to Luebbert's argument as opposed to Moore's.

Indonesia

Indonesia's era of "Guided Democracy" (1959–1965) was marked by a tri-cornered political struggle among the military (ABRI), the noncommunist world's largest communist party (the PKI), and the charismatic figure of President Sukarno, who attempted to balance these powerful forces against each other. This era would be marked by the explosive mobilization of the PKI and its affiliated mass organizations in both urban and rural areas, as well as significant countermobilization by anticommunist forces in the military and the Javanese countryside. "By the late 1950s the PKI had a firm organized base among city workers, estate laborers, and squatters on forestry lands," notes Mortimer. "As the cast of national politics grew more authoritarian, however, this base became more vulnerable. Accordingly, the party's leaders decided to *enlarge and intensify their work among the peasant population* [italics added]" (Mortimer, 1974, p. 278).

As Luebbert argued in the interwar European context, concerted efforts by urban leftists to mobilize the rural proletariat tend to invite the coalescence of a reactionary alliance, willing to use violence to forestall the radicalization of the countryside. The PKI crossed this revolutionary Rubicon by making increasingly strident demands for land redistribution in late 1963 and 1964. "By *radical* land reform, [PKI leader D.N. Aidit] meant *confiscation* of all landlord holdings and their distribution *free* to landless and poor peasants" (Mortimer, 1974, p. 297). Also consistent with Luebbert's analysis, the PKI was not strictly a homegrown rural movement, but represented the kind of populist urban-rural alliance that rural elites fear and resent most. As the PKI launched a program of "unilateral actions" to seize property from landed elites in the early to mid-1960s, more than half of its estimated 20 million members remained located in urban areas (Mortimer, 1974, p. 366).

Even as harsh a critic of the Indonesian "New Order" that emerged from these conflicts as Lev was thus led to conclude that the rise of the PKI in the early to mid-1960s "threatened not only the other parties but the entire traditional elite," and indeed "threatened the entire social and political order" (quoted in Mortimer, 1974, p. 373). Although the coup and countercoup that signaled the fall of Sukarno and rise of Suharto's rightist regime had much to do with military factionalism, the killing campaign that ushered in the New Order was a direct product of rural antisocialist animus. "Anti-communism provided a basic ideological reference point from which the Army was able to organize a diverse coalition of anti-left wing groups," argues Goodfellow. "This alliance had one common denominator. It shared a mutual desire to see an end to the political and economic upheaval for which the PKI was held solely responsible" (Goodfellow, 1995, p. v).

Although this anticommunist alliance clearly transcended the urban-rural divide, it would be in the countryside—as Luebbert argues of Europe's fascist cases—where the rightist popular reaction was most virulent. "The followers of the main Islamic organizations provided the bulk of the masses in the demonstrations in Jakarta and other cities, and Islamic youth participated directly in the eradication of the Left in rural areas" (Aspinall, 1996, p. 216). The carnage against suspected communists reached genocidal proportions, as "Muslims joined forces with the conservative army leadership to destroy the Communist Party," writes Hefner; "as many as half a million people died. Muslim organizations sacralized the campaign, calling it a holy war or *jihad*" (Hefner, 2000, p. 16).

The PKI's strenuous mobilization of a mass following in a concerted effort to overturn the economic and religious order in the countryside was crucial in triggering Indonesia's rightist authoritarian response. Absent this

move to the countryside, it is likely that turn would have more closely resembled the kind of low-violence, demobilizing transition to "traditional dictatorship" seen by Luebbert in Eastern Europe, as opposed to the kind of high-violence, mass-mobilized reaction that Luebbert associates with Western Europe's fascist cases.

Thailand

Alongside fascism, a second authoritarian outcome in interwar Europe identified by Luebbert (1991, pp. 258-266) was "traditional dictatorship." Unlike Indonesia, Thailand's recurrent coups since the overthrow of the absolute monarchy in 1932 consistently installed this form of nondemocratic rule, in which both bloodshed and mass participation were strictly limited. Only on one occasion did democratic collapse in Thailand occur via violent mass mobilization on a "fascist" model; and *it was precisely the efforts by urban socialists to mobilize the countryside* that seems to have instigated such a violent popular response among reactionary, royalist rural Thais in the country's relapse to military rule in 1976. Once again, the external validity of Luebbert's interwar European arguments appears to be confirmed by their transferability to postwar Southeast Asian empirics.

In a region renowned during the Cold War for the robustness of radical leftist movements, Thailand has historically stood apart for its relatively low levels of class-based mobilization. As of the early 1970s, while communist insurgencies raged in nearby Cambodia, Laos, and Vietnam, "[c]ommunism has aroused little response within Thai society, mainly because of the widespread ownership of land, the lack of severe economic misery, popular devotion to the monarchy, the tolerance of Buddhism, and the absence of a Western colonial background" (Darling, 1971, p. 237). The countryside remained politically quiescent and marginal, as Thailand's only significant civilian party under military rule had still "not formulated lucid or coherent plans to benefit the under-privileged rural population" (Darling, 1971, p. 237). In Luebbert's terms, traditional dictatorship continued to characterize the Thai polity.

This would change in 1973, when massive student-led urban protests and the supporting hand of Thailand's charismatic king forced the military from power and inaugurated the country's first genuine experience with representative democracy. "Politicizing the masses and encouraging them to participate politically through protest groups became the students' priority," writes Prudhisan. "In early 1974, they went on democracy propagation campaigns in the provinces. Through it, the villagers became politicized more for pressure politics than for parliamentary politics" (Prudhisan, 1992, p. 68). According to

Morell and Samudavanija, urban leftists' growing role in agrarian mobilization had effects very much in keeping with the Luebbert model:

In June 1974, when the textile workers' strike occurred in Bangkok, *the students went to the villages and mobilized many Central Plain farmers to lend support to the mass demonstrations*. This was the first explicit formation of what they called samprasan, a tripartite alliance of students, farmers, and workers. This type of political coalition, unprecedented in Thailand, caused much alarm among authorities in Bangkok, especially those responsible for the government's counterinsurgency operations. To these officials, such an alliance looked like the basis for implementation of a communist strategy of *inciting urban riots supported by an organized farmers' uprising*. The evidence suggests that it was at about this time—midsummer of 1974—that elements within the rightist establishment . . . decided that firm action against this communist menace was imperative, forming a number of new groups responsive to their own direction. (Morell & Samudavanija, 1981, pp. 159-161; [italics added])

The most important of these new groups was the notorious Village Scouts. Founded as something of a fringe group in 1971 amid growing concerns with the Communist Party of Thailand's strengthening rural presence, the Scouts became a much more powerful force in the wake of the 1973 democratic transition. "The Village Scout movement, directly linked to the Border Patrol Police and the all-powerful Ministry of Interior, had a twofold purpose," notes Bowie.

Viewed broadly, the movement sought to inoculate the Thai body politic against communism by injecting its citizenry with a dose of nationalism. Viewed more immediately, the movement served to intimidate anyone critical of the government into silence and to inform on those who refused to be intimidated. (Bowie, 1997, p. 2)

Though on a more limited scale than in Indonesia, the urban left's rural penetration in Thailand was nonetheless critical to the severity of the reactionary backlash. "Students and others who traveled to rural areas with every intention of helping villagers became embroiled in the political battle between the illusive 'communist terrorist' and the very real Village Scouts" (Bowie, 1997, p. 3). Even more eerily consistent with Luebbert's arguments, the Scouts increasingly took on a troublingly familiar appearance. "With their

uniforms and massive rallies, and by the extensive reliance on emotional ideology, the Village Scouts by 1976 clearly had many attributes of a fascist organization" (Morell & Samudavanija, 1981, p. 244).

Political polarization came to a violent head in October 1976, amid growing accusations of communist sympathies and *lèse-majesté* against student activists at Bangkok's Thamassat University. Military-linked radio stations called on patriotic rural Thais to flood into Bangkok to confront the students; an estimated 200,000 Village Scouts answered the call (Bowie, 1997). When a military assault on leftist students broke out, thousands of Scouts were on hand to engage in grotesque acts of mob killing. The subsequent coup marked the first time in Thai history that fascist-style reactionary mass violence would accompany a military takeover. Not coincidentally, it was also the first Thai coup in which urban leftists' "engagement in agrarian class conflict" played a role in the backlash.

Beyond Indonesia and Thailand, can the Luebbert hypothesis explain variation in the fascistic versus traditional mode of democratic breakdown in Southeast Asia more generally? Although this is not the place to make this case in depth, it is worthwhile to briefly contrast the Indonesian and Thai cases with the Philippines, particularly the declaration of martial law by President Ferdinand Marcos in September 1972. Although Marcos's self-coup was immediately preceded by an upsurge in leftist student activism, Philippine students diverged from their Thai counterparts in making no substantial effort at mobilizing the countryside (Slater, 2010, chap. 5). As Luebbert's argument would suggest, Marcos's coup was marked by no mass violence or rural involvement. For all the considerable *state* violence against dissidents that characterized Marcos's brutal rule (McCoy, 1999), *societal* violence played no role in his seizure of autocratic powers. The Philippine case lends especially strong support for Luebbert's hypothesis as opposed to Moore's alternative approach; although the Philippines clearly possessed more of an entrenched labor-repressive landed elite than either Thailand or Indonesia, it would be in Thailand and Indonesia—and not in the Philippines—where democratic breakdown would most chillingly resemble the popularly violent dynamics of European fascism.

Thus, we see the striking analytical power and theoretical transferability generated by Luebbert's analysis of a distinct place and time. By undertaking a controlled comparison of a set of cases exhibiting representative variation, Luebbert developed a theoretical framework for explaining interwar European political outcomes. Yet his general argument focusing on varying threats

from urban–rural leftist alliances also appears to help explain variation in authoritarian regime onset in distant cases of Cold War Southeast Asia.

Conclusion

One of the enduring trade-offs that scholars of comparative politics face is that of offering general statements that have wide applicability across a broad population of countries and cases, and crafting deeper but narrower arguments that identify the causal process or causal pathway in a smaller number of countries and cases. The tension between these two goals is very real and is often described as the trade-off inherent in making externally or internally valid causal claims. A variety of proposals have been made by scholars over the years to navigate this tension, but currently in comparative politics, a “multimethod turn” has emerged. Driving this turn are two propositions: first, that large-*N* analysis is better at establishing general claims and case studies are better at tracing discrete causal processes; and, second, that the combination of these two methods allows scholars to sidestep the respective pitfalls of each approach when deployed alone. Whether and where the classical controlled comparison still fits in comparative research amid this multimethod turn has remained unclear and hence uncertain.

We wholeheartedly agree that the multimethod solution just detailed has great benefits in principle, and applaud the advances that comparative research in this multimethod vein has established. Yet it is also vital to distinguish what an approach does in principle from what it does in practice. And so far in practice, many of the most impressive and prominent applications of this large-*N*/small-*N* research strategy have ironically been deployed for the opposite purposes usually ascribed to qualitative and quantitative methods. The small-*N* controlled comparison in multimethod designs is often the foundation for externally valid claims, whereas statistical analysis of many subunits *within* the country case in question is often utilized, counterintuitively, to establish internally valid causal relationships (i.e., within the original “sample” of a country case).

This strongly suggests that the standard “single country, multiple methods” research design is not the only mitigating strategy for one of social science’s core dilemmas. Indeed, obscured in this discussion is that controlled comparisons—which in practice have always contained process-tracing evidence and have often contained quantitative evidence—can boast a substantial record of generating externally valid theory that remains sensitive to internal validity. We have argued that these two goals are most likely to be achieved under three specific conditions, that if adopted by scholars hold the promise of revitalizing this canonical but recently much-maligned method:

comparisons that operationalize their chief subject of concern in terms of *general variables or mechanisms*, that seek out *representative variation* that attempts to mirror a broader population, and that engage with theory to select cases that *maximize control*.

Controlled comparisons thus remain a powerful approach for scholars seeking to juggle the dueling demands of internal and external validity within a single study. It is a strategy that allows scholars to generate theory that travels across time and space but remains sensitive to the challenge of establishing precise causal relationships among variables. The multimethod turn does not negate the strengths of the controlled comparison, but ironically underscores them. It is precisely the careful attention to external validity and internal validity within a single study that the multimethod turn has instigated that has offered us an ideal opportunity to develop a robust and more self-conscious elaboration of the classical controlled comparison research strategy.

In the final analysis, it is the fascinating substance that controlled comparisons reveal, and not simply their methodological features, that makes the approach enduringly indispensable. If empirical variation is the lifeblood of comparative politics, scholars can clearly find more of it across cases than within any one. An exclusive focus on variation within countries has a major limitation if never combined with cross-national comparisons: We necessarily overlook the causal importance of national-level attributes in conditioning relationships among subnational variables. Unless one believes that the political features of entire nation-states (e.g., electoral institutions, party systems, state capacity, revolutionary histories) are irrelevant to the outcomes comparativists care about most, controlled cross-national comparisons remain indispensable to the craft of comparative politics.

Acknowledgments

This manuscript has benefited from comments from James Caporaso, Richard Doner, John Gerring, Rich Nielsen, and two anonymous reviewers at Comparative Political Studies, as well as from audiences at Emory University's Department of Political Science, the World Congress of the International Institute of Sociology in Budapest, and the Social Science History Association in Long Beach.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. A third line of critique, emanating primarily from historical sociologists and interpretivist political scientists, takes controlled comparisons to task for their quasiexperimental pretensions. Since this disagreement is ontological as much as methodological, it would require another article at a minimum to address it adequately.
2. Controlled comparisons “differ fundamentally” from statistical inference because the former method “focuses on the problem of many variables rather than on the problem of small N” (Lijphart, 1975, p. 163). By this understanding, which we share, the “most different systems” design that Przeworski and Teune (1970) so famously endorsed “does not fit this definition of the comparative method and should be assigned to the category of statistical analysis” (Lijphart, 1975, p. 164).
3. For an example, see Slater (2010, p. 30, n. 77).
4. The same is more obviously true of experimental research designs. Although the claim that experimental research prioritizes internal over external validity is hardly novel, our claim that even the large-*N* statistical work in many leading comparative works is not geared toward establishing external validity is, to our knowledge anyway, entirely original.
5. We do not suggest that this elective affinity between types of method and types of analysis is necessarily presumed by the authors reviewed below.
6. Variation is more important in subnational analyses than control in that it is the harder source of inferential leverage to gain. Finding subnational regions that are similar in potential explanatory factors is relatively easy; finding subnational regions that vary substantially in outcomes is more difficult.
7. As with Putnam, we by no means imply that Wilkinson’s arguments are beyond critique and reproach, even within India. See Varshney and Gubler (2012).
8. Two examples include Waldner (1999) and an article written in response to Waldner’s book: Doner, Ritchie, and Slater (2005).
9. A prime application of this method is Lieberman (2003). See also Slater (2010, p. 9), which uses Lieberman’s descriptive statistics for similar purposes. It has also been suggested that comparativists might use statistical matching techniques to identify representative cases for controlled comparisons (Nielsen, 2012).
10. Belgium and Czechoslovakia are borderline cases but generally followed the pattern of Luebbert’s liberal and social democratic cases.
11. Though certainly inspiring multiple important efforts to test (and clarify) its main causal claims, Moore’s *Social Origins* (1966) suffers from some core ambiguities because of inattention to delineating its core argument in terms of clearly defined variables and mechanisms.

References

- Adcock, Robert, & Collier, David. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95, 529-546.
- Aspinall, Edward. (1996). The broadening base of political opposition in Indonesia. In Garry Rodan (Ed.), *Political oppositions in industrializing Asia* (pp. 175-195). New York, NY: Routledge.
- Bowie, Katherine A. (1997). *Rituals of national loyalty: An anthropology of the state and the village scout movement in Thailand*. New York, NY: Columbia University Press.
- Brownlee, Jason. (2007). *Authoritarianism in an age of democratization*. New York, NY: Cambridge University Press.
- Callahan, Mary. (2003). *Making enemies: War and state building in Burma*. Ithaca, NY: Cornell University Press.
- Collier, David, LaPorte, Jody, & Seawright, Jason. (2012). Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political Research Quarterly*, 65, 217-232.
- Collier, Ruth Berins, & Collier, David. (1991). *Shaping the political arena: Critical junctures, the labor movement, and regime dynamics in Latin America*. Princeton, NJ: Princeton University Press.
- Darling, Frank. (1971). Political parties in Thailand. *Pacific Affairs*, 44, 228-241.
- Dion, Douglas. (1998). Evidence and inference in the comparative case study. *Comparative Politics*, 30, 127-145.
- Doner, Richard, Ritchie, Bryan, & Slater, Dan. (2005). Systemic vulnerability and the origins of developmental states: Northeast and Southeast Asia in comparative perspective. *International Organization*, 59, 327-361.
- Ertman, Thomas. (1997). *Birth of the Leviathan: Building states and regimes in medieval and early modern Europe*. New York, NY: Cambridge University Press.
- Geddes, Barbara. (2003). *Paradigms and sand castles: Theory building and research design in comparative politics*. Ann Arbor: University of Michigan Press.
- George, Alexander, & Bennett, Andrew. (2004). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.
- Gerring, John. (2007). *Case study research: Principles and practices*. New York, NY: Cambridge University Press.
- Goodfellow, Rob. (1995). *Api Dalam Sekam: The new order and the ideology of anti-communism* (Working Paper No. 95). Clayton, Australia: Monash University Center for Southeast Asian Studies.
- Greene, Kenneth. (2007). *Why dominant parties lose: Mexico's democratization in comparative perspective*. New York, NY: Cambridge University Press.

- Grzymala-Busse, Anna. (2002). *Redeeming the communist past: The regeneration of communist parties in East Central Europe*. New York, NY: Cambridge University Press.
- Hefner, Robert. (2000). *Civil Islam: Muslims and democratization in Indonesia*. Princeton, NJ: Princeton University Press.
- Kalyvas, Stathis. (2006). *The logic of violence in civil war*. New York, NY: Cambridge University Press.
- King, Gary, Keohane, Robert O., & Verba, Sidney. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Kurzman, Charles. (2008). *Democracy denied, 1905–1915: Intellectuals and the fate of democracy*. Cambridge, MA: Harvard University Press.
- Levitsky, Steven, & Way, Lucan. (2010). *Competitive authoritarianism: Hybrid regimes after the cold war*. New York, NY: Cambridge University Press.
- Lieberman, Evan. (2003). *Race and regionalism in the politics of taxation in Brazil and South Africa*. New York, NY: Cambridge University Press.
- Lieberman, Evan. (2005). Nested analysis as a mixed method strategy for comparative research. *American Political Science Review*, 99, 435–452.
- Lijphart, Arend. (1975). The comparable-cases strategy in comparative research. *Comparative Political Studies*, 8, 158–175.
- Lincoln, Yvonne, & Guba, Egon. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Luebbert, Gregory. (1991). *Liberalism, fascism, or social democracy: Social classes and the political origins of regimes in interwar Europe*. New York, NY: Oxford University Press.
- Magaloni, Beatriz. (2006). *Voting for autocracy: Hegemonic party survival and its demise in Mexico*. New York, NY: Cambridge University Press.
- Mahoney, James. (2010). *Colonialism and postcolonial development: Spanish America in comparative perspective*. New York, NY: Cambridge University Press.
- Mahoney, James, & Goertz, Gary. (2006). A tale of two cultures: Contrasting quantitative and qualitative research. *Political Analysis*, 14, 227–249.
- McCoy, Alfred W. (1999). *Closer than brothers: Manhood at the Philippine military academy*. New Haven, CT: Yale University Press.
- McKeown, Timothy. (1999). Case studies and the statistical worldview. *International Organization*, 53, 161–190.
- Migdal, Joel. (1988). *Strong societies and weak states: State-society relations and state capabilities in the third world*. Princeton, NJ: Princeton University Press.
- Mill, John Stuart. (2002). *A system of logic: Ratiocinative and inductive*. Honolulu: University Press of the Pacific. (Original work published 1843)
- Moore, Barrington. (1966). *Social origins of dictatorship and democracy: Lord and peasant in the making of the modern world*. Boston, MA: Beacon.

- Morell, David, & Samudavanija, Chai-anan. (1981). *Political conflict in Thailand: Reform, reaction, revolution*. Cambridge, UK: Oelgeschlager, Gunn & Hain.
- Mortimer, Rex. (1974). Traditional modes and communist movements: Change and protest in Indonesia. In John Wilson Lewis (Ed.), *Peasant rebellion and communist revolution in Asia* (pp. 99-124). Stanford, CA: Stanford University Press.
- Nielsen, Rich. (2012). *Case selection via matching* (Working paper). Cambridge, MA: Harvard University, Department of Government.
- Posner, Daniel. (2006). *Institutions and ethnic politics in Africa*. New York, NY: Cambridge University Press.
- Prudhisan, Jumbala. (1992). *Nation-building and democratization in Thailand: A political history*. Bangkok, Thailand: Chulalongkorn University Social Research Institute.
- Przeworski, Adam, & Teune, Henry. (1970). *The logic of comparative social inquiry*. New York, NY: Wiley-Interscience.
- Putnam, Robert. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.
- Sartori, Giovanni. (1970). Concept misinformation in comparative politics. *American Political Science Review*, 64, 1033-1053.
- Scheiner, Ethan. (2004). *Democracy without competition in Japan*. New York, NY: Cambridge University Press.
- Sekhon, Jason. (2004). Quality meets quantity: Case studies, conditional probability, and counterfactuals. *Perspectives on Politics*, 2, 281-293.
- Skocpol, Theda. (1979). *States and social revolutions: A comparative analysis of France, Russia, and China*. New York, NY: Cambridge University Press.
- Skocpol, Theda, & Somers, Margaret. (1980). The uses of comparative history in macro-social theory. *Comparative Studies in Society and History*, 22, 174-197.
- Slater, Dan. (2010). *Ordering power: Contentious politics and authoritarian levitans in Southeast Asia*. New York, NY: Cambridge University Press.
- Smith, Benjamin. (2007). *Hard times in the lands of plenty: Oil politics in Iran and Indonesia*. Ithaca, NY: Cornell University Press.
- Tarrow, Sidney. (1996). Making social science work across space and time: A critical reflection on Robert Putnam's *Making Democracy Work*. *American Political Science Review*, 90, 389-397.
- Tarrow, Sidney. (2010). The strategy of paired comparison: Toward a theory of practice. *Comparative Political Studies*, 43, 230-259.
- Trimberger, Ellen Kay. (1978). *Revolutions from above: Military bureaucrats and development in Japan, Turkey, Egypt, and Peru*. New Brunswick, NJ: Transaction.
- Tsai, Lily. (2007). *Accountability without democracy: Solidary groups and public goods provision in rural China*. New York, NY: Cambridge University Press.

- Tudor, Maya. (in press). *Twin births, divergent democracies: The social and institutional origins of regime outcomes in India and Pakistan*. New York, NY: Cambridge University Press.
- Varshney, Ashutosh. (2002). *Ethnic conflict and civic life: Hindus and Muslims in India*. New Haven, CT: Yale University Press.
- Varshney, Ashutosh, & Gubler, Joshua. (2012). Does the state promote communal violence for electoral reasons? *India Review*, 11, 191-199.
- Waldner, David. (1999). *State building and late development*. Ithaca, NY: Cornell University Press.
- Wilkinson, Steven. (2004). *Votes and violence: Electoral competition and ethnic riots in India*. New York, NY: Cambridge University Press.
- Wimmer, Andreas. (2002). *Nationalist exclusion and ethnic conflict: Shadows of modernity*. New York, NY: Cambridge University Press.
- Wood, Elisabeth Jean. (2000). *Forging democracy from below: Insurgent transitions in South Africa and El Salvador*. New York, NY: Cambridge University Press.
- Ziblatt, Daniel. (2006). *Structuring the state: The formation of Italy and Germany and the puzzle of federalism* (Princeton, NJ: Princeton University Press, 2006).

Author Biographies

Dan Slater is associate professor of political science at the University of Chicago. He is the author of *Ordering Power: Contentious Politics and Authoritarian Leviathans in Southeast Asia* (Cambridge Studies in Comparative Politics, 2010) and coeditor of *Southeast Asia in Political Science: Theory, Region, and Qualitative Analysis* (Stanford, 2008). His new book projects address the questions of why some authoritarian regimes concede democratization from a position of strength, how accountability does or does not develop in young democracies, and how formal institutions and informal practices interact to shape regime outcomes. He has also published articles in journals such as *American Journal of Political Science*, *American Journal of Sociology*, *Comparative Politics*, *Comparative Political Studies*, *International Organization*, and *Studies in Comparative International Development*, among others.

Daniel Ziblatt is professor of government at Harvard University and is the author of *Structuring the State: The Formation of Italy and Germany and the Puzzle of Federalism* (Princeton University Press, 2006). He is also coeditor of a 2010 special issue of *Comparative Political Studies* titled "The Historical Turn in Democratization Studies." He is completing a book titled *Conservative Political Parties and the Birth of Modern Democracy in Europe* that offers a new interpretation of the general causes of settled democratization by analyzing the historical rise of democracy in Europe. Other new work analyzes how 19th-century bond markets reacted to democratization. Recent articles have appeared in many publications including *American Political Science Review*, *Comparative Political Studies*, and *World Politics*.