Qualitative and Multi-Method Research 2018, Vol. 16, No. 1 https://doi.org/10.5281/zenodo.2562153

Beyond Mill: Why Cross-Case Qualitative Causal Inference Is Weak, and Why We Should Still Compare

Jason Seawright
Northwestern University

Qualitative cross-case comparisons were once widespread and respected enough to be described as "the comparative method." However, the current wave of research on qualitative methods has seen cross-case controlled comparisons fall substantially in esteem. Early criticisms based on selection bias by Geddes (1990) and King, Keohane, and Verba (1994) have been disputed and no longer receive sustained attention in the qualitative methods literature. A more recent argument is that qualitative comparison fails for purposes of causal inference because the required assumptions are simply implausible and because statistical methods are superior tools for the same purpose. Sekhon (2004) argues that comparisons based on Mill-type methods will always be susceptible to probabilistic alternative hypotheses, which generally cannot be reasonably evaluated using qualitative cross-case comparisons. George and Bennett (2004, 151–79) argue at length that "practically all efforts to make use of the controlled comparison method fail to achieve its strict requirements," and that various within-case qualitative methods are simply more usable than qualitative cross-case comparisons. Collier, Mahoney, and Seawright (2004) characterize many forms of qualitative cross-case comparisons as a form of "intuitive regression" that acts inferentially as a weaker and problem-laden equivalent of statistical analysis. Seawright (2016, 107–9) argues briefly that a potential-outcomes formulation makes evident that qualitative comparisons are exceptionally weak tools for causal inference.

This critical tradition coexists with sustained use of paired and otherwise grouped comparison in qualitative research, as Slater and Ziblatt (2013, 1302–3, 1307–10) demonstrate. This essay argues that the existing critical literature has been insufficiently attentive to the range of justifying assumptions that qualitative scholars might make in thinking about qualitative cross-case comparison, but also that careful consideration reinforces the view that such assumptions are generally implausible. It then goes on to argue that cross-case controlled comparisons in qualitative research have real value for other methodological objectives, value that has not been fully articulated or respected in the existing literature.

Comparison for Causal Inference

For many contemporary definitions of causation, there is no inherent, logical connection between the method of comparison across cases and the goal of causal inference. Some traditions of thought about causation, dating at least back to Hume and including a sequence of thinkers up to Baumgartner's (2008) contemporary work, focus on regularities across cases, arguing that causation is nothing but a certain pattern of predictable relationships between variables across a population of cases. Given this definition, it is clear that methods for causal inference would need to rely centrally on comparison. Indeed, it seems that such a view would rule out any methods other than comparison, as any kind of within-case analysis seems to be, at most, weakly related to the existence of reliable cross-case patterns.

However, influential alternative perspectives are available. Quantitative and statistical thinking about causation in the social sciences is currently dominated by a synthesis of counterfactual- and manipulation-based approaches (Rubin 1974; Holland 1986; Woodward 2003). Here, causation is not inherently about differences across cases, and cross-case comparison is at most a contingent tool for causal inference rather than part of the definitional core of the concept. Instead, causation is ultimately, if perhaps unobservably, about what would have happened within a single case had the treatment (or main independent variable) of interest been manipulated to take on a different value than it, in fact, did. That is, causation is always about the difference between what happened and what would have happened had a particular, well-defined choice been made differently.

A common mathematical notation has emerged around this way of thinking. The inherently counterfactual nature of causation in this framework is captured by the creation of multiple versions of the dependent variable for each case. If case *i* receives the treatment (which is an often arbitrarily chosen value from the range of the main independent variables), the dependent variable that occurs is $Y_{i,T.}$ On the other hand, assuming a binary independent variable for simplicity, if case *i* receives the control, the observed value of the outcome is $Y_{i,C}$. The true causal effect of the main independent variable for case *i* is, therefore, $Y_{i,T} - Y_{i,C}$.

Differences that Balance within a Group

The qualitative controlled comparison is sometimes analogized with statistical matching methods for causal inference (Seawright and Gerring 2008; Nielsen 2016). Hence, it is worth looking closely at the assumptions for causal inference made when using these methods. Are they a viable justification for qualitative cross-case comparison?

Matching methods, like most statistical techniques, rely on an analogy to experimental research designs for causal inference. In experiments, the combination of random assignment and the law of large numbers guarantees that the average value of $Y_{i,T}$ in the treatment group will be very similar to the average (unobservable) value of $Y_{i,T}$ in the control group. The same logic holds true for $Y_{i,C}$. Hence, the difference between the observed group average values of $Y_{i,T}$ and $Y_{i,C}$ is close to the true causal effect between the two groups.

The comparability of the treatment and control groups within a social-science experiment does not arise because each case in the experiment is interchangeable. Individuals, who are usually the cases in experiments, obviously differ from each other in unlimited ways. These differences are accommodated because they balance out on average. Causal inference works because random assignment, on average, balances individual differences within the treatment and control groups. Furthermore, significance testing of various kinds offers a framework for handling the inevitable real-world imbalances that arise in experiments with finite sample sizes.

Matching methods cannot appeal to random assignment to guarantee that differences across cases will balance out within the treatment and control groups. Instead, scholars using matching make a brute-force assumption that, after creating group balance across a fixed set of control variables, all other differences will balance out within each group. This assumption is difficult to justify; unmeasured or neglected variables seem to routinely fail to balance. Nonetheless, causal inference is possible with these methods as long as all relevant differences either (a) are included in the set of measured control variables for matching, or (b) happen to balance out within the treatment group and within the control group.

This logic, fragile as it is, is all but unavailable to qualitative scholars. In a paired comparison, the treatment and control groups each consist of a single case. Obviously, nothing can "balance out" statistically within a single instance. If the treatment case is, for example, unusually liberal, then the treatment group will simply be unusually liberal.

Thus, paired comparisons must seek some other justification for causal inference. Small-group qualitative comparisons are likewise obliged unless the cases under comparison are exceptionally simple. Finally, significance testing cannot offer to mitigate these problems, given that small-N controlled comparisons virtually never feature enough cases for such tests.

Differences that Balance within a Case

How, then, might scholars justify such qualitative comparisons? If the treatment case is just irreducibly different from the control case in ways other than the main independent variable, is causal inference possible?

In a paired comparison, causal inference revolves around $Y_{1,T} - Y_{2,C}$, where case 1 is the treatment case and 2 is the control. This quantity will correctly describe causal matters for case 1 if $Y_{1,T} - Y_{2,C} = Y_{1,T} - Y_{1,C}$. Thus, the inference requires that $Y_{2,C} = Y_{1,C}$. This is a strikingly stringent requirement: the two cases simply cannot differ from each other in terms of the outcome they would experience under the control. Indeed, while various quantitative approaches to causal inference require assumptions that are difficult to meet, or that are even implausible, this assumption is more restrictive than those required for any widely-used quantitative technique. Regression analysis, for example, allows for causal inference in the face of random measurement error or omitted cases that are not confounders; this assumption cannot succeed in the face of either of these problems. Thus, controlled comparison requires the same kind of assumption as regression, but a much stronger version of it.

One assumption that might meet this condition is that the differences between the treatment and control cases balance each other out within each of the two cases. If one unusual feature of case 1 adds, say, three points to $Y_{1,T}$ and $Y_{1,C}$, but the second (and only other) unusual feature subtracts three points, then taken as a whole, case 1 poses no problems for causal inference. Of course, there is no special reason to expect that differences within a case will tend to balance, as opposed to accumulate.

Differences that Barely Matter

Perhaps the most frequent informally expressed justification for causal inference via qualitative comparison involves the idea that two cases may not be identical, and the distinctive features of each case may not internally balance, but that causal inference will still work if the consequences of the differences are modest enough. Perhaps $Y_{2,C}$ does not equal $Y_{1,C}$, but the causal inference will still be acceptable if $Y_{2,C}$ is very close to $Y_{1,C}$—a condition that is met if the causal effects of differences between the cases on the outcome of interest are all quite small. What counts as "very close" is relative to the size of the true causal effect: any other differences must have effects that are a tiny fraction of the effect of interest, or the inference will be meaningfully distorted.

This setup can seem reasonable. Surely case experts are likely to focus on important differences, and may well have the knowledge necessary to pair up broadly similar cases. Nevertheless, this argument results in disturbingly fragile causal inferences. Because the causal inference is only approximately correct if the causal effect in question is large and the effects of all other differences between cases are small, it will only be persuasive to scholars who are already firmly convinced that the main independent variable is the biggest cause of the outcome. Any readers who are instead open to the alternative hypothesis that there are some other causes of comparable importance to the main independent variable cannot avoid worrying that the causal inference is biased by the differences between the cases.

No Differences

Finally, and most starkly, qualitative causal inference will succeed if there are simply no differences between the cases under comparison. If the treatment case and the control case are exactly identical in every way that is causally relevant to the outcome, then the causal inference will succeed. This condition, known as causal homogeneity, seems to capture a common interpretation of what J.S. Mill intended with the method of difference. In some kinds of physical science laboratories, careful procedure can more or less achieve exact interchangeability between a treatment and a control sample, allowing a direct pairwise comparison to support causal inference. It seems self-evidently problematic to identify a pair of human beings, let alone any larger social aggregate or institution, as comparably interchangeable.

Billionaires and the Causal Role of Public Pressure

To illustrate the problems with these four assumptions justifying qualitative comparison for causal inference, consider a paired comparison between two politically conservative American billionaires, David Koch and John Menard, Jr.[1] In a broad perspective, these individuals have an enormous amount in common. They are immersed in the shared political culture of the 21st-century United States. They share an elite socioeconomic position. They have overlapping social networks and quite convergent political views, as evidenced by Menard's participation in a series of seminars sponsored by Koch and his brother.

These and many more similarities notwithstanding, Menard and Koch have crucial differences that matter for understanding American billionaires' participatory strategies. While both Menard and Koch are heavily invested in conservative economic politics, Koch's views have received substantially more public attention and indirect defense through his foundations, support for scholarship, and even a handful of public statements. Menard's political perspectives, by contrast, have not been given deliberate public airing—and instead have emerged via investigative journalism and legal action. What explains this contrast in the two billionaires' willingness to engage in stealth politics (Page, Seawright, and Lacombe 2018), i.e., participatory strategies that evade public scrutiny and offer little or no deliberative defense of one's policy preferences?

One interesting explanatory possibility is that the difference is accounted for by the extent to which the two billionaires' wealth depends on public-facing businesses and brands. Menard's wealth is founded in the success of his eponymous chain of home improvement stores. As such, publicly visible political action might carry the risk of boycotts or general consumer distaste of a sort that could hurt Menard economically. Koch, by contrast, draws his wealth in substantial part from the energy industry, but also from a number of behind-the-scenes investments. Because Koch's wealth mostly comes from industries that sell to other industries, and depends little on his personal brand, he has limited economic exposure to boycotts or other forms of consumer rejection. Hence, it may be unsurprising that Koch is willing to participate in more potentially visible ways than Menard: Koch simply has less to lose.

The key issue for this essay is, of course, not whether this explanation is true or false, but rather whether any of the four assumptions characterized in the previous

[1] Material in this section draws on Page, Seawright, and Lacombe (2018).

section are applicable and can justify causal inference with this comparison. It should be immediately clear that the first assumption, that differences will balance within groups, is not applicable. There is only one billionaire within each group—and while it would certainly be possible to expand the analysis to include more billionaires, reaching the sample sizes that would justify use of the law of large numbers essentially precludes qualitative treatment of the comparison.

What of the assumption that differences balance within cases? Even evaluating this assumption would require exceptional prior causal knowledge. Potentially relevant contrasts between Menard and Koch are numerous and varied. The two billionaires differ in terms of family backgrounds, with Koch coming from a successful family with an established (if not yet world-dominant) business, while Menard was born to solidly middle class parents. They differ in birthplaces, as well, although perhaps in ways that will only be legible to Midwesterners: Menard was born in Eau Claire, Wisconsin, while Koch comes from Wichita, Kansas. They depart substantially in terms of their current places of residence and cultural interests. Menard still lives in Wisconsin, and has long sponsored a team that competes in the Indy Racing League. Koch, by contrast, lives on 740 Park Avenue in Manhattan, and is famous for his substantial philanthropical gifts to support cancer research, the New York and Washington, D.C., arts and museums scenes, and a public broadcasting foundation.

Do these and other differences between the billionaires exactly cancel out? While it would be fortuitous if they did, the fact is that even determining the answer would require remarkable prior causal knowledge. Does an affluent as opposed to middle class origin predispose a billionaire to greater public political visibility, or does the causal effect run in the other direction? Would Koch's philanthropical efforts create connections with Manhattan social life that increase the potential costs of visible political activism, or would philanthropy create a buffer against criticism? The issues involved in even deciding whether the assumption is met are immense and probably, at present, insurmountable.

The exact same challenge destroys any potential applicability of the third assumption, that the differences between these two billionaires barely matters. I might invite the reader to believe that public- versus industry-facing primary sources of wealth have much more powerful effects on political participatory strategy than do social networks, family histories, and so forth. Yet what of the inevitable reader who disagrees? For any reader who sees social networks as potentially as important, the comparison crumbles. A scholar might respond by selecting a comparison between billionaires with similar social networks—but this problem will remain as long as any difference whatsoever persists between the billionaires. At the ultimate limit, imagine a pair of identical-twin billionaires raised in the same household, and residing in the same condominium building, but with one of them for some reason heavily invested in consumer-facing enterprises and the other not. Even though these hypothetical billionaires are similar to the point of fantasy, it nonetheless remains certain that they will have subtle, but potentially relevant, differences. Their social networks will not be identical. They will have slightly divergent sets of politically relevant information. Some life experiences will not be shared, and so forth. To claim that these differences *must* have smaller effects on participatory strategies than the public- vs. industry-facing contrast is to assert *a priori* that the causal effect of interest is relatively large. Thus, this assumption becomes uncomfortably close to circular.

Finally, the discussion over the last paragraphs should absolutely suffice to reject any notion that Menard and Koch are identical in all ways other than whether their businesses face the consumer public. Billionaires are not chemical samples, and there is simply no prospect for interchangeability. Thus, it would seem that prospects for causal inference from a paired comparison between Menard and Koch are grim. One might tinker at the margins by selecting slightly more similar pairs of billionaires, but the basic issues encountered here will persist.

Yet if qualitative comparisons are a hopeless strategy for causal inference about billionaires, prospects are surely grim for virtually any other application of this design in the social sciences. Among the overall population of humans, after all, American billionaires are a remarkably homogeneous group with exceptional similarities in culture, class, and context. Simply put, qualitative comparison appears to have little to offer as a tool of direct causal inference because the required assumptions are implausible, at best, in the social sciences.

The Value in Comparison

Of course, it certainly does not follow from this argument that qualitative comparison is useless, or that existing qualitative work featuring controlled comparison needs to be discarded altogether. Rather, the value of comparison arises from goals other than direct causal inference. Here, I will highlight three ways that qualitative comparisons make social-scientific contributions. Such comparisons are valuable because they: sharpen conceptualizations and measurement; allow exploration of the prevalence of causal capacities; and provide raw materials for the construction of theories of causal moderation. It makes sense to reread existing studies of this sort along such lines, even perhaps against their authors' intentions, and it is emphatically reasonable to design future qualitative comparisons with these alternative goals in mind.

To begin with, as Slater and Ziblatt (2013, 1312) note, comparison facilitates conceptualization and measurement by providing empirical content and grounding for theoretical contrasts. There is value in using qualitative comparisons to understand the meaning and real scope of possible different outcomes with respect to a dependent variable. Qualitative comparison can allow inductive discovery related to that scope in ways that are hard to replicate with other methods.

Slater and Ziblatt (2013) argue at length for another advantage of qualitative comparison: providing external validity for causal inferences based on (qualitative or quantitative) single-country analysis. Their argument demonstrates decisively that scholars routinely use qualitative comparison for this purpose, and that such studies are often well received by their respective research communities. Yet there are certain tensions involved in the discussion that result from conceptual messiness related to the idea of external validity.

External validity is sometimes discussed in terms of sample-to-population statistical inference. It is deeply unclear that qualitative comparison could ever provide external validity in a statistical sense. Qualitative cases are rarely randomly sampled, and even if they were, it would be exceptional for qualitative analysis to include enough cases to acquire attractive statistical properties.

In order to speak of external validity in the context of qualitative comparison, a reframing is needed. Slater and Ziblatt (2013, 1314) helpfully reformulate the concept: "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." Here, external validity becomes a sliding scale: an argument scores higher to the extent that it applies to more cases, and also presumably to the extent that it throws a brighter "explanatory light" on each case.

What does "explanatory light" consist of? By usage in the quoted passage, and by Slater and Ziblatt's deployment of related terms throughout, "explanatory light" appears to be a relation between an explanation and a case. At one point, they gloss this feature as involving "verisimilitude on causal mechanisms." Unfortunately, this might mean a number of different things: highly detailed theories of causal pathways; extensive and persuasive pattern-matching (Campbell 1966); evidence that a given case exemplifies a theorized causal pathway; or evidence justifying an overall causal inference regarding an entire theoretical model's causal correctness vis-a-vis a given case (Waldner 2015), to name a few.

On the supposition that Slater and Ziblatt are referring to causal pathways, i.e., sequences of variables with causal linkages that may serve to fill in steps between the treatment and the outcome within a given case, there is another difficulty. Two cases might have very different results even if they experience the same causal pathways in the sense that each is affected by genuine causal linkages from the treatment variable, through one or more shared mediator variables, to the outcome. This is because the size of the causal effects involved in each relation need not be constant across cases. Background facts about a given case may render it more or less susceptible to a particular causal effect, and thus may change the magnitude of causal patterns without altering their form. Such changes in magnitude should probably be seen as altering the degree to which a theory explains a given case, as would differences in the nature of the observed mediator variables. But such sequential causal inferences are difficult at best. It is not clear that qualitative research is a powerful tool for quantifying causal magnitudes, and it is also unclear how one might trade off between magnitudes of effects versus identities of mediators in evaluating explanatory fit.

Ultimately, I am unpersuaded that there is value in applying the concept of external validity to qualitative cross-case comparisons. If a causal theory is correct for a given case, then it cannot be made untrue by results in another case; nor can a theory that is false for a given case be made true by its performance in other cases. Consideration of a theory's explanatory value outside the core case or cases of interest is thus neither a matter of statistical generalization, nor of testing the theory's validity.

Contemporary theories of causation provide a helpful set of tools for reframing this issue in ways that are arguably compatible with, although more extensive than, the potential-outcomes framework adopted earlier (e.g., Seawright 2016, 29–30; Cartwright 2009). Cartwright (2007), in particular, pushes us to think of causation as context-specific arrangements of objects, institutions, actors, and so forth in such a way that the well-known capacities of each specific entity interact to generate the outcome of interest. Because configurations of entities and their causal capacities differ across contexts as a brute fact, there is no reason why valid causal arguments should be expected to be universal, or indeed even to be valid more than once. Thus, in Cartwright's view, good causal theories are those which identify relevant entities and correctly describe their causal capacities and arrangements. These micro-components of a causal explanation can be real and theories about their capacities can be true, but because of the prevalence of difference in arrangements across cases, generalized causal theories or laws are false (Cartwright 1983). Rather, the kinds of causal findings that result from an experiment, a case study, or most other approaches are true or false relative to some given "nomological machine" or specific arrangement of causal capacities (Pemberton and Cartwright 2014).

From such a position of central concern for entities and their causal capacities, Slater and Ziblatt's concern for external validity can be helpfully recharacterized as understanding the prevalence of the causal capacities central to a theory across a range of cases. In-depth qualitative and/or quantitative analysis of a single case might give us good reason to believe that a particular causal arrangement is operative in that case, and no evidence from other cases need ever trouble that conclusion. Yet it remains instructive to ask whether there are other cases in which the same entities demonstrate the theorized causal capacities. While a causal explanation can be valid and nonetheless unique to a single case, there is an obvious gain to credibility when explanations involve common, easily demonstrated capacities. For this task, qualitative comparison can contribute.

Of course, social scientists are rarely satisfied to note that one set of cases is simply causally different from another. If causal capacities are arranged one way in a first domain, and another way in a second, it is reasonable and perhaps compelling to ask why the domains differ. This kind of second-order problem of causal theory involves the project of understanding relations of causal moderation, i.e., the background variable or variables that cause cases to differ in terms of the main theory's network of causal relations and capacities. Building (and perhaps to some extent testing) theories about moderation is in itself a highly valuable goal.

Furthermore, it is easy to interpret much of the use of qualitative comparison in the comparative-historical literature in political science and sociology as carrying out (to varying degrees) careful within-case causal inference and then using comparison to structure theory-building about moderation. Consider, for example, Collier and Collier's (1991) study of labor, parties, and regimes in early- to mid-20th-century Latin America, a classic that serves as one of Slater and Ziblatt's motivating examples. In that study, the treatment variable is labor incorporation (i.e., the inclusion of labor unions as part of the legal political system), and the outcomes involve patterns of party-system formation and certain trajectories of regime dynamics. The central argument of the volume is, in fact, that labor incorporation, a shared event across the eight countries in the study, does *not* have the same effects or activate the same causal dynamics across the region. Instead, Collier and Collier argue that certain background characteristics have a tendency to cause countries to incorporate labor in different ways, and that these different modes of incorporation produce divergent arrangements of causal capacities. The bulk of the study consists of careful within-case analysis that attempts to establish the actual causal effects for each case, and the cross-case comparisons that frame the volume can easily be understood as building a theory of causal moderation (a theory which is, in part, also tested using the within-case analysis).

Conclusions

This essay has argued that none of the assumptions which could justify causal inference via paired or more elaborately grouped qualitative controlled comparison are likely to be even remotely plausible in social-science applications. Thus, it is a mistake to attempt to justify a qualitative comparative research design by claiming that the design will achieve causal inference via control, correspond with Mill's Methods, or even meaningfully rule out a given explanation (which might after all be probabilistic or interact in complex ways with background variables).

Yet the conclusion is not that qualitative researchers should abandon comparison. Such research designs make real contributions in terms of conceptualization and measurement, exploring the prevalence of causal capacities, and building theories of causal moderation. These contributions all retain their value, independently of the methods used for within-case causal inference. The optimal comparative designs for these purposes should be a lively topic for future research within the qualitative methods community.

References

Baumgartner, Michael. 2008. "Regularity Theories Reassessed." *Philosophia* 36 (3): 327–54. https://doi.org/10.1007/s11406-007-9114-4.

Campbell, Donald T. 1966. "Pattern Matching as an Essential in Distal Knowing." In *The Psychology of Egon Brunswick*, edited by Kenneth R. Hammond, 81–106. New York: Holt, Rinehart and Winston.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Oxford University Press.

———. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.

———. 2009. "What is This Thing Called 'Efficacy'?" In *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice,* edited by C. Mantzavinos, 185–206. Cambridge: Cambridge University Press.

Collier, Ruth Berins, and David Collier. 1991. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton: Princeton University Press.

Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* 2 (1): 131–50. https://doi.org/10.1093/pan/2.1.131.

George, Alexander L., and Andrew Bennett. 2004. *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. https://doi.org/10.1080/01621459.1986.10478354.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Nielsen, Richard A. 2016. "Case Selection via Matching." *Sociological Methods & Research* 45 (3): 569–97. https://doi.org/10.1177/0049124114547054.

Page, Benjamin I., Jason Seawright, and Matthew J. Lacombe. 2018. *Billionaires and Stealth Politics*. Chicago: University of Chicago Press.

Pemberton, John, and Nancy Cartwright. 2014. "Ceteris Paribus Laws Need Machines to Generate Them." *Erkenntnis* 79 (Supplement 10): 1745–58. https://doi.org/10.1007/s10670-014-9639-4.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701. https://doi.org/10.1037/h0037350.

Seawright, Jason, and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61 (2): 294–308. https://doi.org/10.1177/1065912907313077.

Sekhon, Jasjeet S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics* 2 (2): 281–93. https://doi.org/10.1017/S1537592704040150.

Slater, Dan, and Daniel Ziblatt. 2013. "The Enduring Indispensability of the Controlled Comparison." *Comparative Political Studies* 46 (10): 1301–27. https://doi.org/10.1177/0010414012472469.

Waldner, David. 2015. "What Makes Process Tracing Good? Causal Mechanisms, Causal Inference, and the Completeness Standard in Comparative Politics." In *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel, 126–52. Cambridge: Cambridge University Press.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.