

Theory, Comparison, and Design: A Review Essay

Thomas B. Pepinsky , Cornell University
pepinsky@cornell.edu

Theory and Credibility: Integrating Theoretical and Empirical Social Science. By Scott Ashworth, Christopher R. Berry, and Ethan Bueno de Mesquita. Princeton: Princeton University Press, 2021. 280p. \$95.00 cloth, \$35.00 paper.

The Logic of Social Science. By James Mahoney. Princeton: Princeton University Press, 2021. 416p. \$95.00 cloth, \$35.00 paper.

Rethinking Comparison: Innovative Methods for Qualitative Political Inquiry. Edited by Erica S. Simmons and Nicholas Rush Smith. New York: Cambridge University Press, 2021. 225p. \$105.00 cloth, \$34.99 paper.

Three recent books introduce political scientists to important debates and conceptual challenges for contemporary social science research. Rather than proposing new methods, these books propose new frameworks that their authors believe should guide current practice. The authors and editors seek to guide researchers through the research process, from asking questions and proposing theories to operationalizing concepts and executing comparative research designs. Reading the three books together will give researchers a comprehensive overview of the state of the discipline on important questions like how and what to compare, what social science concepts should refer to, and how to proceed when theory and empirics disagree.

The three books, however, have starkly different orientations toward the practice of political science research. And accordingly, they address different problems. In *Theory and Credibility: Integrating Theoretical and Empirical Social Science*, Scott Ashworth, Christopher Berry, and Ethan Bueno de Mesquita examine the relationship between formal theoretical models (usually expressed using mathematics) and empirical models (usually quantitative in nature), inviting readers to understand how these two abstract representations of important political phenomena can be mutually enriching. The contributors to *Rethinking Comparison: Innovative Methods for Qualitative Political Inquiry* advance a pluralist agenda that explicates and defends alternatives to what the volume's editors, Erica Simmons and

Nicholas Rush Smith, identify as a hegemonic emphasis on controlled comparisons in the Millian framework. In *The Logic of Social Science*, James Mahoney advances a distinctive set-theoretic approach to social science theorizing in which abstract categories are socially constructed yet amenable to scientific analysis.

In essence, each of these three books identifies and responds to a different problem facing social science research:

1. *Rethinking Comparison* addresses the problem of *are there alternatives to the standard template for qualitative comparisons?*
2. *The Logic of Social Science* addresses the problem of *what are social categories made of, and how can we study them?*
3. *Theory and Credibility* addresses the problem of *how can we reconcile theoretical and empirical models of the social world?*

Reviewing these three books alongside one another illustrates a good range of methodological orientations and substantive concerns within the discipline. It does not make for much cross-talk, as the books set out to do such radically different things that connections among them can be hard to see. A critical observer might conclude that the discipline is unmoored, at least when it comes to research design and political methodology. An optimistic alternative interpretation is that the

discipline retains the flexibility and intellectual capaciousness to recognize and respond to all manner of challenges in the research process. In what follows, I will seek to identify points of common interest among these three works, but my main objective is to show how each work addresses the problem it identifies.

Theory and Credibility

Ashworth et al. motivate their book by observing the growing gap between formal theory and the credibility revolution in the social sciences. They are not alone in identifying that these two approaches to social science research seem to be growing apart in political science (I do not think that it is correct that such a gap exists in economics), but I am not aware of any other work that focuses so closely on empirical methods for causal inference when proposing a solution. This is a specific instance of a general problem that has a long pedigree in the philosophy of science: if theories are simplifications of a complex world, designed to identify some part of how it functions, and empirical data also summarize some part of that complex world, what should we do when theory and empirics disagree? (On this question, see Clarke and Primo 2012.) On what basis could we say that an empirical model contradicts a theoretical one? Ashworth et al.'s focus on credible research designs is particularly welcome because empirical methods for causal inference have thoroughly transformed the practice of quantitative political science research in the past two decades (see Angrist and Pischke 2009). And, importantly, credible research designs estimate causal quantities that have a different epistemic status than conditional correlations and data summaries produced in other empirical work: these are, aspirationally at least, *causal* relationships.

Theory and Credibility makes a conceptual point about how to integrate formal theory and credible research designs for causal inference. The authors clearly have students in mind, and this book will serve well as an introduction to positivist quantitative social science that can complement other texts like *Mostly Harmless Econometrics*, now a standard reference for any graduate student working in the applied microeconomics tradition of empirical social science (Angrist and Pischke 2009). But Ashworth et al.'s contribution is a clear and novel framework for integrating formal theoretical models with empirical models. Building on the idea of models as simplifications, the authors posit that both empirical and theoretical models are simplifications of some social phenomenon (this is figure 1 on p. 13). We judge the theoretical model on its faithfulness to that social phenomenon, and we judge the empirical model in the same way. To the extent that these are each faithful simplifications of the same referent that produce pertinent theoretical implications and relevant empirical facts, we can say that theory and empirics are commensurable. When

developing theories to learn about the empirical world, a good objective is to have theoretical models and credible research designs that are commensurable. It is bad to abandon theory in response to empirics without checking first whether the theory's implications are commensurable with what we can learn from the data at hand.

The key to their argument is that both empirical and theoretical models make "all else equal" claims; importantly, credible research designs do this, but there are many empirical research designs that do not. It is safe to say that credible empirical research designs make commensurability more defensible, so the credibility revolution can strengthen the link between theory and empirics. Scholars should strive for the theory's all-else-equal claims to be reflected in the research design's all-else-equal claims. But even if we set commensurability as an objective, we need to also attend to the similarity between the empirical and theoretical models and the social phenomena being represented.

This is a sensible and productive way to conceptualize the relationship between empirical and theoretical social science research. One notes, however, just how important the pragmatic judgments of the researcher and audience are. Similarity, relevance, pertinence, and commensurability are the key ingredients for Ashworth et al.'s integration of theory and empirics, and each of them is a judgment, rather than a fact. There is no procedure to calculate a relevance statistic, nor an equation that proves a commensurability relationship. Instead, there are arguments. Skeptical readers will identify this as a weakness of Ashworth et al.'s approach, but I think it better to understand this as a clear statement of the limits of what any scientific research enterprise can hope to accomplish.

The first part of the book explicates their approach to integrating formal theory with the credibility revolution, and then introduces the reader to formal theories and credible research designs with this discussion in mind. The second part of the book discusses the practical way that a dialogical relationship between theory and research design can advance social scientific knowledge. I found this second part of the book the most interesting, as it is replete with illustrative examples of how theory and empirics inform one another. I found it helpful to think of reinterpreting, elaborating, distinguishing, and disentangling—their terms—as working within some part of the schematic representation of the research process in figure 1.

Beyond the concepts of commensurability and similarity, there are important practical implications to their argument. One such implication, which may not be the one they intended, is that theory is necessary to discipline credible causal inference. It is common for causally identified quantities to be misinterpreted as evidence in favor of a theory, or as "pure" estimates of a model parameter, but we usually have to write down the theory to know if

that is true. The book contains numerous examples, ranging from incumbency advantage to gender bias in political careers, where convincing empirical results (sometimes descriptive, others actually causal estimates of treatment effects) have entirely different meanings when interpreted using a simple formal model.

I would have preferred to see a more forceful discussion of this point in light of the growth of experimental methodologies in political science. Causal inference, after all, is atheoretical: the definition of a causal effect in the potential-outcomes framework is simply the difference in outcomes between treated and control state (see Rubin 2005). You can randomize anything without any theory of why an effect exists or what brings it about. This approach to scholarship cannot contribute to explanation in the social sciences—Ashworth et al.’s stated goal—without a theory of what that estimand represents. Too often, contemporary empirical practice in political science is to randomize something and then to speculate about some theory with which that design is commensurable. At the same time, however, many working within the causal inference tradition are not interested in explanation, but rather evaluation. Although the authors and I share a preference for explanation, and for theories that can help us to answer “why” questions, much of the credibility revolution has been inspired by the desire for convincing answers to “whether” questions.

This brings me to my final point, about formal languages for representing causal inference. The authors are most comfortable working in the potential-outcomes framework, and use the language of directed acyclic graphs (DAGs) only to illustrate complicated causal inference strategies visually. DAGs, though, are models of causal relationships, too. Do they have all-else-equal claims as well? Should they be understood as empirical models, or as theoretical models? The recent title *The Book of Why*, which introduces DAGs with the subtitle *The New Science of Cause and Effect*, suggests that DAGs are capable of answering those “why” questions (Pearl and Mackenzie 2018). It is an open question whether DAGs ought to be understood as a bridge between theoretical and empirical models—perhaps a visual representation of what needs to be true for the two to be commensurable—or as something else entirely. Reading *Theory and Credibility* does lead me to conclude that DAGs are not theoretical models in the sense that Ashworth et al. mean them.

The Logic of Social Science

A reader familiar with the language of Ashworth et al., but not with the philosophical fields of ontology or epistemology, will find the first two chapters of Mahoney’s new volume to be wholly foreign, even though both books are about how to ground the study of the social world in the procedures of science and logic. One way to see how they relate to one another is to examine one point of tangency:

concepts and measurement. For Ashworth et al., we may take as given the existence of certain objects in the world—things like *incumbent*, *woman*, or *conflict*. The task is to conceptualize these social objects correctly and measure them accurately. Or, we can posit them theoretically: there can be a concept like *candidate quality* that causes our theories to work in certain ways even if we never try to measure it. Mahoney is not willing to concede that these things exist without explaining first *what it means to say that they exist*.

Mahoney draws on philosophy, linguistics, and cognitive science to argue that the fundamental problem with contemporary social science research is that it is essentialist. It is hard to follow the argument and its implications, so I will spell out what I take the core issues to be. Essentialism describes the belief that classes of objects form *kinds*, or classes of objects, because they share certain properties, or *essences*. An apple is a member of the category of things that are apples because it has apple properties. Importantly, these exist regardless of what humans think about them. The number seven exists regardless of whether humans exist, or whether humans are counting, or how our minds represent it. Apples and seven are *natural kinds* because their essences exist independently of human thought, action, or existence.

Social scientists study *human kinds*, which lack any intrinsic essence that makes them kinds, and are instead constructed as such through the operation of the human mind. *Chief justice*, *purple*, *democratization*, and *population growth* are all human kinds. Human minds create categories—like *family*—whose members share no common constituent parts or even intrinsic features. Kinds vary continuously on a single dimension from purely natural to purely human (there does not seem to be a third alternative: partial membership in the category of natural kind requires partial membership in the category of human kind). Social science deals with kinds that are mostly or exclusively human kinds, which lack any intrinsic essence that constitutes their membership in that category.

Mahoney is a firm believer that reality *does* exist, in that human kinds are made up of natural kinds, but he argues that studying human kinds requires social scientists to abandon essentialism. This means more than conceding the point that social phenomena like *revolutions* or *parliaments* are constituted through human thought and action. It means also rejecting the “*property-possession assumption*—i.e., the assumption that social science categories, like natural kinds, possess hidden and causally efficacious powers” (p. 23, emphasis in original). Continues Mahoney, “our built-in essentialist orientation leads us to view as strange the argument that states are not entities that exist in the world with causal potentials ... [but] to understand any relationship between states and violence, we must acknowledge and somehow model the mind-dependent nature of the category *state*” (p. 31,

emphasis in original). The details of how to accomplish this occupy roughly two dozen pages, but the idea is that we social scientists must work against our inherent cognitive instinct to essentialize social phenomena. We can do so by embracing a form of constructivism that recognizes that categories like *Black* or *coupe* or *coup* or *black* are ultimately mental projections, conceptualized as sets (which are defined axiomatically), and using logic—which is “woven into the fabric of reality” (p. 40)—to organize and relate them with one another.

I suspect that no matter how clearly Mahoney has defended his objective of scientific analysis, his embrace of formal logic, and his belief that objective reality does exist, certain readers will conclude that he is rejecting science and reality by grounding his account in a mind-dependent approach to human kinds for social analysis. That is a mistake, for even if the heavy focus on cognitive science and philosophy proves unfamiliar to readers who are not used to engaging with such issues, there is no mistaking Mahoney’s ambitions for a science of society that comports with what we generally know about logic, the philosophy of science, and human cognition.

Most of the rest of the book is an exercise in elaborating a set-theoretic approach to social analysis. Much of this is familiar ground, although it is explained with fuller detail and with more citations to the relevant literature than in most other treatments of set-theoretic reasoning for empirical social scientists. Mahoney also offers us the most robust and philosophically grounded treatment of causality under an alternative to the counterfactual model that I have seen in social science methodology since Holland (1986). (Mahoney favors the regularity model, although I believe his dismissal of inherent causal powers operating through social aggregates is too hasty.) If you accept the premise that set-theoretic thinking is essential for a philosophically coherent social science, then the book’s early chapters set the foundation for the later ones. However, one may reject that premise and still make use of the later chapters. For example, even if one believed—against all that we know to be true—that *race* and *wealthy person* were natural kinds, that person could also use a set-theoretic analysis to query the relationships between sets of individuals with various degrees of membership in different racial and wealth categories. Set-theoretic reasoning does not depend on Mahoney’s articulation of scientific constructivism.

Might it be the case, though, that only set-theoretic reasoning can accommodate the anti-essentialist stance that Mahoney advocates? This is Mahoney’s big argument, but it rests uncomfortably on claims about how human cognition works that are themselves socially constructed under any possible definition of the term (and certainly under Mahoney’s). These involve statements about how the human mind represents categories, as well as which parts of the biochemical process of human cognition are

relevant. I do not know what to think about these claims, although they are all made with due reference to the appropriate literatures in cognitive psychology. Nor do I know how to think about variation across individuals in human cognitive styles, exceptions to general tendencies in one individual’s cognitive processes, or more generally, about errors, disagreements, interpretations, reflexivity, or positionality. I think that Mahoney would agree that when we talk about sex, we are talking about a mind-dependent category. On what basis, then, would we resolve a disagreement between me and someone else about what sex refers to without positing some mind-independent frame of reference?

It seems that Mahoney must posit a series of human cognitive universals to produce the conclusion that only set-theoretic reasoning can accommodate anti-essentialism. As it is, I found myself agreeing with just about every argument about human kinds and the social construction of categories, but not agreeing that this compels me to work differently than I do right now, and specifically as a set-theoretic social scientist.

What is missing from *The Logic of Social Science* is a discussion of the practical stakes of a set-theoretic approach to social science grounded in social construction instead of some other approach. Mahoney provides a strong analysis of how to work within this tradition, but not enough for me to know what I am likely to get wrong by not working in this tradition. Such lacunae notwithstanding, Mahoney’s intellectual project is refreshing in its ambitiousness, and discussions of “theory frames” and related metatheoretical and normative concerns over time and across disciplines make for compelling and entirely novel reading.

Rethinking Comparison

Read next to Ashworth et al. and Mahoney, *Rethinking Comparison* is an entirely different enterprise. Whereas the former two works seek to organize current practices in empirical research into a coherent framework, Simmons and Smith urge researchers to break free from a singular template of qualitative comparative research organized around the “controlled comparison.” A controlled comparison is a qualitative comparison of a small number of units designed to identify causal effects through an application of Mill’s method of difference (see Przeworski and Teune 1970). All actually existing qualitative comparative research designs are in some sense imperfect—there is no case of a country that shares all the features of China except for a different regime type—but the aspiration is to have chosen comparative cases that are so similar as to have eliminated most plausible confounding differences that might explain an outcome in question.

Should we ever take such qualitative comparative analyses seriously? An early chapter by Jason Seawright lays the point bare: qualitative comparisons plausibly identify

causal effects only under highly unrealistic assumptions of comparability. Comparisons across like units identify causal effects in quantitative work under well-known assumptions, but none of them are likely to obtain when comparing aggregate social entities like states or social movements, no matter how carefully the cases are chosen. There are three possible responses. One is that qualitative comparisons cannot identify causal effects. Another is that qualitative comparisons only help to identify causal effects when embedded in a larger multi-method research exercise that invokes data at different scales and temporalities. And the third—the focus of this volume—is to remember that qualitative comparisons have different purposes.

Simmons and Smith have assembled the best collection of arguments in favor of comparison on its own terms—to probe concepts, to give perspective, for discursive purposes, and others. Comparison is everywhere in our daily lives, and it is ubiquitous in social science (and, too, in other forms of intellectual inquiry, from biological systematics to literature). Nearly every small-N qualitative comparison in the history of social research is *not* a controlled comparison designed to identify a causal effect. We have learned from such noncontrolled comparisons for centuries. Simmons and Smith and their contributors have established that a discipline that eliminated any noncontrolled comparisons would be eliminating some of the most important sources of knowledge that we have. I doubt that anyone would disagree strongly with this conclusion, but as today's comparative methodologists continue to emphasize the essentially causal nature of many qualitative empirical claims, reminders of the alternative values of comparisons are essential.

As a comparativist myself, I found the most useful contributions to *Rethinking Comparison* to be those that walk us constructively through alternative modes of comparative work. Sarah Parkinson's chapter on network logics and comparisons stands out in this regard. Jillian Schwedler's analysis of "encompassing comparisons" (originally due to Charles Tilly) also shows how to think productively about large-scale process with local variations that themselves co-constitute the large-scale process in question. Yet others see themselves as working against the methodological status quo represented by the political science mainstream. Contributors like Nic Cheeseman, for example, celebrate the work of rule breakers like Benedict Anderson whose "comparisons were informal, uncontrolled, and at times empirically off the mark" (p. 65) while asking audiences "whether or not they trust the author who has written" (p. 72) the provocative comparative works that they are reading—even those that are empirically off the mark!

Simmons and Smith have more in mind, though, than explaining that there are alternative logics of comparison for qualitative researchers. They also want

to encourage more scholarship that employs these alternative logics, embracing the possibilities of comparison without bounds. Here, though, we encounter the challenge of arguments that are designed to open up new ways of thinking *against* a common standard rather than to bring researchers together into a common framework with mutually accepted terms of engagement. What is a *bad* comparison if there are so many different logics one might use to justify odd, surprising, or provocative comparisons? Are there any kinds of qualitative comparisons that scholars should avoid? What sorts of conclusions or inferences should scholars refuse to draw from others' work? These are not idle questions: no one holds that all comparisons are of equal value, a point repeated explicitly several times by chapter authors.

It seems to me that Frederick Shaffer's discussion provides the right answer—but that it is an unsatisfying one. Qualitative comparisons of any form are good just so long as they yield insights into how the world works or what it is made of (pp. 55–58). Sadly, we don't know how to identify those comparisons *ex ante*. My own view, moreover, is that we must acknowledge that social status and disciplinary prestige explain what sorts of provocations and surprising comparisons are accorded merit and deemed insightful or revealing despite their refusal to accommodate standard templates for comparative analysis (on this and related points, see Cribb 2005). It also makes the value of comparison depend on the status quo state of knowledge of the audience in question. A comparison that provokes Europeans to think differently about state strength may be utterly conventional to Africanists, or to scholars from Africa.

The conundrum that emerges from a capacious and inclusive approach to qualitative comparative inquiry is that by opening a space for alternative logics or rationales for comparison in empirical social science research, we lose some of the communal agreement about how to evaluate that research. Simmons and Smith are aware of this challenge, and they proceed pragmatically, describing their work as providing a "vocabulary to describe their approach" (p. 11) and opening possibilities without determining the domains in which they are appropriate. The next step toward a constructive and cumulative methodological research agenda is to apply the same depth of criticism to nonmainstream qualitative comparisons, building on Simmons and Smith's vocabulary and the foundations they have built. Having learned that there are bad controlled comparisons, now let us entertain other associated and equally relevant questions: What is the bad ethnography? What are the irrelevant or misleading comparisons? And whose voice has authority to establish which surprising uncontrolled comparisons are productive, provocative, insightful, or novel? We will know more about

what makes for insightful comparisons if we are willing to contrast them with other, less insightful ones.

Concluding Thoughts

If there is a common thread that emerges from reading these important new books next to one another, it is that social science is a communal enterprise. Arguments depend not only on the facts in question, and the logic applied, but on subjective judgments and pragmatic evaluations of researchers and the audiences with whom they are communicating. This is not a new observation about the scientific enterprise, but it is helpful to be reminded that it is unavoidable—no matter what approach one takes to building a science of politics.

References

- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Clarke, Kevin A., and David M. Primo. 2012. *A Model Discipline: Political Science and the Logic of Representations*. Oxford: Oxford University Press. DOI: [10.1093/acprof:oso/9780195382198.001.0001](https://doi.org/10.1093/acprof:oso/9780195382198.001.0001).
- Cribb, Robert. 2005. "Circles of Esteem, Standard Works, and Euphoric Couplets." *Critical Asian Studies* 37(2): 289–304. DOI: [10.1080/14672710500106408](https://doi.org/10.1080/14672710500106408).
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945–60. DOI: [10.1080/01621459.1986.10478354](https://doi.org/10.1080/01621459.1986.10478354).
- Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.
- Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. Hoboken, NJ: Wiley.
- Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100(469): 322–31. DOI: [10.1198/016214504000001880](https://doi.org/10.1198/016214504000001880).