Toward a Pluralistic Vision of Methodology

Henry E. Brady, David Collier, and Jason Seawright
Department of Political Science, 210 Barrows Hall,
University of California Berkeley, Berkeley, CA 94720-1950
e-mail: hbrady@berkeley.edu (corresponding author)
e-mail: dcollier@berkeley.edu
e-mail: seawrigh@socrates.berkeley.edu

1 Introduction

The discussion of qualitative and quantitative research methods in political science has followed a "long arc of development," according to Andrew Bennett (this symposium). He sees our book *Rethinking Social Inquiry: Diverse Tools, Shared Standards (RSI)* as a "key turning point" in this development and as contributing to a more pluralistic vision of methodology. We would certainly be gratified if we have indeed made this contribution. Yet this is an ongoing discussion, and this symposium points to fruitful directions this discussion can take.

Our book has, in fact, produced diverse responses. There seems to be general agreement with Philip Shively's observation that "*Rethinking Social Inquiry* is science as it should be, with serious and respectful discussion of matters that often elicit a rather different tone of voice." The qualitative and "medium-N" methodologist Benoît Rihoux praises the book in strong terms, underscoring among other things its value for teaching methodology; he nonetheless sees the volume as contributing more to a debate in North America between quantitative and qualitative methodologists than to European social science, which is more oriented toward qualitative methodology. Philip Schrodt, a meticulous quantitative methodologist, joins us in underscoring the limitations of conventional quantitative research in political science—based on linear regression analysis and econometric refinements on regression. Indeed, Schrodt takes an even harsher view than we do of the mainstream quantitative methods that have been at the center of political science methodology for at least two-and-a-half decades. He thereby reinforces our rationale for exploring how quantitative and qualitative tools can complement one another—indeed, why and how each may be more powerful when used in conjunction with the other.

On the other hand, our colleague Nathaniel Beck adopts a skeptical view of our enterprise. A central goal of our book is to establish the contrast between, on the one hand, *dataset observations* (DSOs)—which are located in a "rectangular data set" of variables and cases and which are the basis for correlation and regression analysis—and, on the other hand, *causal-process observations* (CPOs)—which we define as insights or pieces of data that provide information about context, process, or mechanism and that contribute distinctive leverage in causal inference (*RSI*, 277). Beck is unconvinced that CPOs, taken outside a rectangular data set, contribute to causal inference. He argues that CPOs cannot be "adjoined" with DSOs in making inferences. In the title of his article, he suggests that the idea of a CPO may be an "oxymoron." Only DSOs provide a viable basis for causal inference.

We address these alternative reactions to *RSI*, first by underscoring some of the main arguments developed in the book, then by considering the insights offered by the contributors to this symposium, and finally by further exploring the idea of CPOs, reviewing examples from the natural sciences, epidemiology, and political science. Some political scientists might view DSOs as the sole foundation for more rigorous research. Hence, it is particularly relevant to review examples from other disciplines that are seen as more rigorous and in which CPOs—and also the practice of adjoining CPOs and DSOs—play an exemplary role.

2 Toward a Pluralistic Vision of Methodology

2.1 *Quantitative Methods, Qualitative Methods, and Statistical Theory*

A central concern of *RSI* is with the interrelations among quantitative methods, qualitative methods, and statistical theory. Of course, a complex set of distinctions lies behind the quantitative-qualitative dichotomy, as we emphasize in the first footnote in our text (*RSI*, 4, n. 1; also 244–50). Yet this dichotomy provides a heuristic contrast that usefully structures important parts of the present discussion.

On the quantitative side, we focus on what may be called *mainstream quantitative methods* in political science, based to a substantial degree on research tools involving regression analysis and econometric refinements on regression. This tradition is well institutionalized and broadly used within our discipline. Beyond being an important approach in its own terms, mainstream quantitative methods are sometimes evoked as providing a "quantitative template" that can be applied to political science research in general. Indeed, it is sometimes argued that other methodologies, including qualitative methods, should be subordinated to this template.

On the qualitative side, a well-defined set of analytic tools has emerged (*RSI*, passim), although qualitative methodology has only become institutionalized more recently.[1] Regarding the relation of qualitative methods to quantitative methods, one can identify a defensive posture that actively resists the imposition of the quantitative template. Yet some qualitative methodologists are convinced that their approach has strengths lacking in quantitative research and that this latter body of work has much to learn from the qualitative tradition.

A third perspective is that of statistical theory, which we understand as a broad tradition for reasoning about evidence and inference. Statistical theory provides the foundation for quantitative research. Yet statisticians are often concerned that social scientists who work with observational (rather than experimental) data move too quickly toward making causal inferences and that in the process they leave themselves vulnerable to many inferential mistakes. Thus, far from providing unambiguous support for the use of mainstream quantitative methods as an engine of causal inference, statistical theory may, in light of this skepticism and under some circumstances, in fact provide a statistical rationale for alternative methodologies such as qualitative research (*RSI*, 4, n. 2; 254, n. 29).

Given these diverse perspectives, we have the foundation for a productive dialogue between the quantitative and qualitative approaches, informed by the perspective of statistical theory. Central to our view of this dialogue is that both qualitative and quantitative

---

[1] This institutionalization has been reflected, since the year 2000, in the formation of the Qualitative Methods Section of the American Political Science Association and the rapid expansion of its membership, in the success of the Institute for Qualitative Research Methods held at Arizona State University every January, and in the expanding number of political science graduate programs that offer courses in qualitative methods.

analyses are hard to do well. This contention is more than a rhetorical lament about the difficulty of carrying out good research. Too many qualitative and quantitative studies are simply bad work. We believe that both sides in the qualitative-quantitative debate would be more credible if they began by acknowledging how hard it is to do good social science. Modesty would become us all.

The assertion that *qualitative* analysis is difficult to carry out in a rigorous manner may seem obvious to a great many *quantitative* researchers. Many scholars in the latter group would acknowledge that qualitative studies unearth valuable knowledge of cases and context. Yet these quantitative scholars routinely see the qualitative tradition as lacking standardized, transparent, and replicable criteria for data collection and measurement and also for causal inference. Both descriptive and causal inferences are also seen as suffering from the many problems that can derive from working with a small N. Many challenges must be overcome.

However, mainstream quantitative analysis also faces major pitfalls—notwithstanding the valuable standardization of research procedures, the replicability of these procedures, and the crucial contributions of formal statistical tests to description and to causal inference. Some of the pitfalls of quantitative analysis are spelled out in Schrodt's contribution to this symposium, and others are identified in Bartels's chapter in *RSI*, entitled "Some Unfulfilled Promises of Quantitative Imperialism" (chap. 4). The difficulties noted by Schrodt, Bartels, and others are not primarily a consequence of the failure to use cutting-edge quantitative techniques; rather, they are inherent to the task of making cross-case causal inferences on the basis of observational data. Central to these difficulties is the fact that, as Leamer put it, with observational data "there is no formal way to know what inferential monsters lurk beyond our immediate field of vision" (quoted in *RSI*, 230).

2.2   *Causal-Process Observations*

In light of these divergent strengths and weaknesses of quantitative and qualitative methods, a major theme of our book is to explore one of several approaches to bridging their respective strengths—i.e., by combining the valuable standardization of procedures of the former, with the insight into cases of the latter. We develop this idea through the juxtaposition, already introduced, of data-set observations and causal-process observations.

DSOs are the basis for the standard rectangular data set of the quantitative researcher, with rows corresponding to cases and columns corresponding to variables. This data set is the foundation for correlation and regression analysis. In relation to this rectangular data set, the term "observation" has a very specific meaning. It is not the ordinary language meaning, in the sense that one "observes" phenomena in the real world. Rather, an observation is specifically an entire row in the rectangular data set. It is all the scores for a given case.

A CPO, by contrast, is an insight or piece of data that provides information about context, process, or mechanism and that contributes distinctive leverage to causal inference. It is not part of a rectangular data set; it provides a separate type of inferential leverage. Our goal in selecting this label is to incorporate the term "observation," which as just noted has a special status in relation to causal inference in quantitative research, and to juxtapose it with the idea of causal process (*RSI*, 252–5).

The label causal-process observation could raise the concern that researchers are understood as directly "observing" causation.[2] However, this would not make sense, given

---

[2] This issue is raised by Beck and indirectly by Bennett.

that causation is an abstract concept about which analysts simply make inferences. It is not observed directly, either with CPOs or on the basis of the regression coefficients that are built on DSOs.[3] Given this potential misunderstanding as to whether researchers directly observe causal processes—if some readers find it more helpful to think of this as "causal-process information," that is a useful alternative label. However, to reiterate, we deliberately called these pieces of data CPOs to emphasize that this kind of evidence merits the same level of analytic and methodological attention as DSOs in quantitative research (*RSI*, 253).

Some quantitative methodologists and statisticians agree that what we call CPOs are a valuable component of credible causal inference. For example, Goldthorpe argues that information about mechanism and causal process "must be *added* to any statistical criteria before an argument about causation can convincingly be made."[4] Further, Goldthorpe cites statisticians who, with some variations in emphasis, present this same argument (*RSI*, 54, n. 1).

CPOs can thus be used in conjunction with quantitative analysis, as well as separately. CPOs can be the primary basis for inference in a given study, with DSOs serving initially to frame the analysis or subsequently to place the conclusions in a wider comparative context. When CPOs are used in a study that is primarily quantitative, they may be invaluable at the beginning, when the researcher needs clues about which variables to include. CPOs can play a role in an intermediate phase of a study, for example, to build on insights already developed using DSOs and to point to fruitful directions for further analysis. Finally, at the end of a quantitative analysis, CPOs can serve as a check on whether the causal inferences derived from DSOs are plausible. The balance and sequencing[5] between the two types of observations may vary greatly, and overall, neither necessarily plays a more decisive role.

3 Model Specification, Bayesian Analysis, and Strategies for Combining Methodologies

The commentaries in this symposium build constructively on the arguments of our book. We address three main topics that emerge in this discussion.

3.1 *Model Specification*

Schrodt emphasizes the serious inferential difficulties that can arise from reliance on linear model specifications in quantitative research. Researchers who employ these models must recognize the highly problematic instability of results under alternative specifications, the challenges of collinearity, and the often inappropriate use of significance tests in conjunction with the models. Schrodt emphasizes the usefulness of nonlinear statistical models, such as Charles Ragin's (1987) Qualitative Comparative Analysis (QCA), as an alternative to the difficulties of linear modeling.

For those who might think that Schrodt takes these concerns too far, similar concerns are in fact expressed in Achen's (2002) review of political methodology in which he pointedly proposes "A Rule of Three (ART): A statistical specification with more than three explanatory variables is meaningless." He proposes this rule because "often the

---

[3] David Hume made the classic and decisive case against the direct observability of causation.
[4] Goldthorpe (2001, 8). Italics in the original.
[5] See Tarrow, *RSI*, chap. 10.

causal patterns are dramatically different across the cases" and "with more than three independent variables, no one can do the careful data analysis to ensure that the model specification is accurate and that the assumptions fit as well as the researcher claims" (2002, 446–7).

We read Achen's solution not so much as a recommendation that precisely three explanatory variables should be employed but rather as a strong admonition that the statistical model should be parsimonious and carefully specified.[6] He is right about the nature of the problem. Schrodt, Ragin, and Achen are all concerned that causal relationships are most likely nonlinear, interactive, and logically complex.[7] It is telling that Ragin, a qualitative methodologist, has been a leader in pointing out this problem and that other qualitative researchers often refer to the possibility of "conjunctural" causation (*RSI*, 278).

It is also telling that Achen recommends looking at subsets of cases because "a study that gets the unique causal patterns of black Protestants approximately right and throws everyone else out of the sample is better than an analysis that tosses every group into the statistical soup and gets them all wrong" (2002, 447). Of course, quantitative researchers are right when they say that regression models can accommodate subgroup differences through higher order interaction terms, but as Achen points out, in practice quantitative researchers seldom take seriously the need to look for distinct causal patterns across different cases in their data sets.

Shively's contribution to this symposium focuses centrally on another issue related to model specification: selection bias.[8] He argues, contrary to the argument advanced in *RSI* (chap. 6), that selection bias is just as problematic for within-case as for cross-case analysis. He suggests that, for cases with extreme scores on the dependent variable, within-case analysis will be thrown off by the disproportionate presence of extreme scores for unusual reasons—which will lead to idiosyncratic, although not incorrect, causal inferences. Shively makes a useful point, but he is not saying that within-case analysis is threatened by selection bias, rather he is saying that the causal relationships found through within-case analysis in a case selected based on the value of the dependent variable might not be generalizable to other cases. Researchers, therefore, should be careful not to generalize from what they find in the Netherlands or Ireland to other cases. That seems like a standard warning, but Shively goes on to amend it in an interesting way by noting that one of the major insights of *RSI* is to emphasize that the contributions of a case study are to test theories and to develop them further. Hence, he argues "it may indeed be useful to select cases deliberately with regard to how they fit into a particular theory." This conclusion agrees nicely with the argument in *RSI* that selection bias is too often cited as a bogeyman when the real issues lie elsewhere.

---

[6]The problem of specification bias, as with inadequate linear specifications in a multiple regression, cannot be solved by omitting all but three variables from the regression (thus introducing another form of specification bias) and getting the specification of the remaining three "just right." Achen proposes that three explanatory variables would be enough variables if we had strong theory, but even that seems either insufficient or impossible for most situations in which nonexperimental data are being analyzed.

[7]It is worth reflecting on the fact that if a binary outcome can be produced through two alternative sets of sufficient conditions, each consisting of the interaction of two dichotomous necessary conditions (which technically speaking are INUS conditions—Mackie 1974), then the correct regression specification involves two interaction terms and one fourth-order interaction of all four conditions. No terms will simply be linear combinations of the necessary conditions.

[8]Although selection bias is primarily an issue involving the analytic consequences of case selection, the literature on econometric solutions to selection bias has emphasized that it can be reconceptualized as a missing-variables problem.

3.2 *Bayesian Analysis*

Schrodt not only criticizes the overuse of linear models but also notes that "the ubiquity of exploratory statistical research has rendered traditional frequentist significance tests all but meaningless." He suggests that a Bayesian perspective might provide a more coherent theory of inference. Bennett's remarks extend the domain of Bayesian reasoning into discussions of qualitative research design, arguing that qualitative case selection, process tracing, and generalization on the basis of case studies involve a Bayesian as opposed to a frequentist statistical logic. From his discussion, it would appear that Bennett is particularly concerned with two themes that he sees as distinctively connected with Bayesian statistics: the insight that some kinds of cases may be more valuable for a particular inferential problem than others and the imperative of closely linking theory with causal inference.

McKeown's (*RSI*, 158–62) idea of a "folk Bayesian" approach may be most helpful here, rather than a technically complex form of Bayesian statistics. Without entering into debate over these alternatives, we certainly agree with Schrodt and Bennett. Some categories of cases are more valuable than others for specific research tasks. As Munck emphasizes (*RSI*, 120), qualitative researchers have developed a range of case selection strategies for developing and/or refining theoretical explanations, and debates about case selection for causal inference are lively and ongoing. Furthermore, it is clear that causal inference without links to theory contributes relatively little to the development of the social sciences.[9] Hence, a Bayesian perspective, at least in the informal sense noted above, is invaluable.

3.3 *Geographic Location of the Debate and Strategies for Combining Methods*

Benoît Rihoux observes that the methodology debate addressed by *RSI* is located in the United States. In Europe, by contrast, political science is more strongly centered in qualitative research, and this debate is less relevant and more unfamiliar. Some U.S. political scientists might suggest that Rihoux's argument simply demonstrates that Europe is outside of the disciplinary "conversation." Yet a word of caution is needed here. In Hix's (2004) ranking of political science departments, almost half of the 200 departments evaluated are located outside of the U.S., as are eight of the top-ranked 26 departments. We may well be experiencing a globalization of political science, and Rihoux's concern with the U.S.-centric character of this methodology debate merits serious attention.

Rihoux argues further that for a non-U.S. audience, more attention could have been given in *RSI* to strategies for combining methodologies. He refers in particular to Charles Ragin's (1987) QCA methods and to recent work that advocates the joint use of qualitative and quantitative methods in a single research project. Rihoux expresses concern that *RSI* misses an opportunity to discuss more thoroughly these multimethod research strategies.

In fact, the book strongly endorses these strategies. First of all, combining DSOs and CPOs, as we have advocated, is one mode of bridging quantitative and qualitative methods. Notable examples, discussed in *RSI*, include the use of CPOs in Stokes' study of policy switches by Latin American presidents (*RSI*, 256–7; Stokes 2001) and in Brady's analysis of voter turnout in Florida during the U.S. 2000 presidential election. Tarrow's discussion of triangulation (*RSI*, chap. 10) spells out alternative multimethod strategies,

---

[9]Even so, we would second Schrodt's valuable caution: appeals to use theory in order to resolve methodological dilemmas create circular arguments; in fact, the major reason we need empirical evidence in the first place is to adjudicate between competing, and often equally plausible, theoretical arguments.

and we see strong convergences between Tarrow's approach and Coppedge's (2005) idea of nested inference, Lieberman's (2005) idea of nested analysis, and Laitin's (2002) tripartite method that combines theory, comparative-quantitative analysis, and narrative case evidence.

Thus, we agree with Rihoux's emphasis on searching for ways of combining qualitative and quantitative methods in research on a given topic. However, we would add a cautionary note: both the qualitative and the quantitative components in a multimethod design must stand up to scrutiny in its own right. Both components must separately meet the standards for persuasive inference. Hence, although combining methods can have major advantages, it also imposes a corresponding methodological burden on the researcher.

4 Causal Process Observations: An Oxymoron?

We now return to the idea of CPOs and the interesting critique advanced by Nathaniel Beck that this idea may be an oxymoron. Beck views DSOs as the primary tool for causal inference, and he has strong misgivings about CPOs. He argues that CPOs do not solve problems of research design in a novel way and in particular fail to resolve problems of defective research design. Good research design is about ruling out rival explanations, and it is unclear to him what CPOs contribute to meeting this challenge. He allows that what we call CPOs may potentially yield descriptive information about cases that can improve causal inference based on quantitative analysis, but he believes that they do not do so in their own right.

Beck's two most crucial, overall arguments are the following. First, although CPOs may play a role in evaluating explanations of singular events, they do not contribute to causal inference vis-a-vis general phenomena. Yet this latter form of inference is, in his view, the central task of social science. This shortcoming leads Beck to suggest that the idea of a CPO may be an oxymoron. Second, he asserts that *RSI* does not make it clear what it means to "adjoin" CPOs and DSOs, in the sense that they can effectively be used together. He sees this problem as undermining a number of key points in *RSI*.

Clearly, we disagree with these arguments. We believe that CPOs do contribute to assessing explanations of general outcomes and they are routinely "adjoined" with DSOs. After a further introduction to the logic of combining the two types of observations, we support our argument with a series of examples. We show that in the natural sciences— astronomy and paleontology—CPOs are used for making causal inferences regarding both singular and general outcomes. These examples demonstrate that CPOs can definitely contribute to rigorous research.

Next we discuss two examples from epidemiology, which—like political science— routinely has available DSOs with a large N, involving observational data, which can readily be analyzed with the tools of regression analysis. Nonetheless, important advances in epidemiology have depended on adjoining these DSOs with CPOs. We conclude with a review of four examples from the social sciences—three previously discussed in *RSI* and by Beck and a further example of a nested design that illustrates key ideas about linking DSOs and CPOs. Our overall point is that CPOs play an important role in rigorous inference across many domains of research.

5 Why Causal Process Observations Matter

Why, then, are CPOs inherently important? Extending our discussion above, we argue that a major part of the answer is found in the empirical and theoretical limitations of

DSOs. DSOs are especially useful for detecting probabilistic relationships when there are many observations of comparable units and when the relevant causal pathways and interactions are well understood or controlled through randomized experiments. Yet it is often hard, if not impossible, to increase the number of analytically relevant DSOs, and we seldom have adequate understanding or control of these causal pathways or interactions.

In many situations, only a few similar units or events can be studied. For example, we find fewer than two dozen truly advanced industrial countries, relatively few cases (at least until recently) where nuclear weapons might have been used, and only about a dozen social revolutions. In these situations, adding cases may be impossible (or at least foolhardy) because new cases will differ fundamentally from the original universe of concern. Furthermore, in some situations, social scientists—and also natural scientists—wish to explain a singular event, such as why World War I occurred or why the dinosaurs became extinct. In these situations, large-N studies may not be relevant.

The second limitation is even more serious. Increasing the number of DSOs can provide additional leverage when the major inferential problem is lack of statistical power stemming from a weak probabilistic relationship between a putative cause and effect. However, it can merely add cost without increasing leverage if the fundamental difficulty is a research design that diverges from a true experiment. In that case, the most vexing problem is not lack of data. Rather it is the lack of an appropriate design and the failure of a proper statistical specification that can allow for valid inferences. Proper statistical specification depends on strong theory that can guide the analysis.

Thus, the value of the regression-based model of inference, which is central to mainstream quantitative methods, depends upon a large number of comparable DSOs, which are hard to obtain, and on strong statistical specification assumptions, which are difficult to satisfy. We argue that a different strategy involving CPOs is sometimes more productive and less costly. This approach is hardly a panacea because it still demands careful data collection and strong assumptions in order to interpret CPOs. Yet it does provide an alternative when DSOs and regression analysis appear imperfect or defective.

At the simplest level, CPOs are diagnostic pieces of information that provide key insights in assessing explanations. A standard metaphor employed in discussing CPOs involves the parallel to criminal detective work. Detectives make their diagnoses on the basis of dogs that do not bark (as in Sherlock Holmes' famous "Silver Blaze" story), missing suicide notes, other clues that are or are not found at crime scenes, and stories that "just don't add up." As will be shown in the examples below, this type of analysis goes beyond a simple model of "cause and effect" and recognizes that a causal process typically involves complex mechanisms, mediators, and markers that can provide alternative ways to test theories and to develop explanations. Paying attention to these mechanisms, mediators, and markers can reveal causal processes, and they are the foundation for CPOs.

5.1 *Examples from Natural Science*

5.1.1 Astronomy: From Brahe to Galileo and Innovation in the 20th Century

Both CPOs and DSOs figure prominently in the history of astronomy. Two especially interesting periods are 1540–1620, when the heliocentric (sun-centered) model replaced the earth-centered Ptolemaic model, and the mid-20th century, when an unexpected CPO lent support to the "big bang" theory.

In the earlier period, the thinking of the great observational astronomer Tycho Brahe (1546–1601) was shaped by three simple observations, one DSO and two CPOs, that changed how he understood the cosmos. A fourth observation—a CPO made by Galileo—provided the key piece of information for substantiating the theories developed through the efforts of Copernicus, Brahe, Kepler, and Galileo.

In 1563, Brahe observed a conjunction of Saturn and Jupiter on August 17—occurring a month before it was predicted by the Alfonsine Tables based on the refinements of the Ptolemaic theory and several days before it was predicted by the newer Copernican Tables. We view this as a DSO, given that its status as an anomaly depended on its location in the data sets that were the basis for the predictions. This DSO indicated that standard theories were failing to predict the locations of the planets, and this failure set Brahe on a course of astronomical research to develop better data and a better theory of the heavens.

Nine years later, Brahe first saw a great Nova—a star so bright that it shone in the daytime. This, along with his observations of the comet of 1577, constituted CPOs, suggesting that there were causal forces operating in the heavens that made them just as changeable as events on the earth—thus contradicting the prevailing Aristotelian and Christian doctrines that the heavens were immutable and not subject to terrestrial laws.

Finally, in 1610, Galileo (1564–1642) discovered the moons of Jupiter.[10] This CPO provided a model for a counterargument against those who maintained that common sense showed that all the heavenly bodies revolved around the earth (Kuhn 1957, 222).

The Nova, the comet, and the moons of Jupiter were CPOs that disconfirmed the older theories and provided evidence for the newer ones. They contributed to causal inference with respect to general outcomes, and they were adjoined to DSOs in useful ways.

In the 20th century, the discovery of cosmological microwave background radiation (CMBR) provides an example of a CPO that changed astronomy (Lachieze-Rey and Gunzig 1999). In the early 1960s, two astrophysicists at Bell Laboratories, who were interested in the spectra of radio sources, inherited a powerful radio telescope previously employed for satellite communication. Using this telescope, they continually found their data corrupted by an annoying background radiation. After considering many possible sources of this radiation (including a pigeon's nest in the antenna) without eliminating it, they contemplated giving up some of their research goals. A chance encounter with a colleague linked them to a theoretical group at Princeton that had predicted such background radiation as a marker of the effects of the big bang. Hence, the CMBR was a "smoking gun" that one would expect to find if the universe was not infinitely old and in a steady state but rather was created through a big bang at some time in the past. These results were a turning point in the history of 20th century astronomy.

These examples of CPOs in astronomy have the same form, given that careful thinking about causal processes, based on revealing and sometime puzzling observations, makes it possible to rule out some explanations and accept others. Furthermore, in most of these illustrations, it is hard to even imagine what a relevant set of DSOs would look like for coming to the same conclusions.[11]

---

[10] See Koestler (1959, 284–90, 351).

[11] We could, perhaps, imagine a sample of various universes in which the dependent variable is CMBR (one if present, zero if absent) and a randomized assignment of the big bang (one if present, zero if absent). Then a statistically significant coefficient from this sample would mean that the big bang caused CMBR. But the logic here does not, and simply could not, look at all like that. For a discussion of the complexities of this causal logic, see Lachieze-Rey and Gunzig (1999, 27–9). It is also interesting to reflect upon whether CMBR constitutes a case of singular or general causation.

5.1.2 Dinosaur Extinction

Research seeking to explain the extinction of dinosaurs makes repeated use of CPOs. This example is especially interesting because it starts out as an attempt to explain a singular event and ends up as part of a general literature on major extinction events and declines in species diversity. Consequently, it shows how CPOs can be used for understanding both singular and general outcomes.

The extinction of the dinosaurs appeared to have occurred somewhat abruptly about 65 million years ago, in the late Cretaceous (K) period, just before the Tertiary (T) period at the "KT boundary." The abruptness of the event presented something of a theoretical problem because ever since Charles Lyell's *Principles of Geology* (1830–33), geologists and paleontologists have been gradualists who rejected catastrophic "Noah's flood" theories. However, the difficulty of dating extinctions exactly left plenty of "wiggle room" for "noncatastrophists," in that it left open the possibility of an extinction event that took place over as much as a million years.

Research on this topic, especially the dating problem, has made extensive use of DSOs from diverse locations. A typical study has looked at species diversity over time in various strata of rocks, which are dated using other geological information, including radiometric dating procedures. Technical statistical problems such as the Signor-Lipps effect—a truncation problem arising because the most recent occurrence of a fossil is unlikely to represent the last individual of the species—have been addressed using sophisticated quantitative methods (Hallam and Wignall 1997, 16–7).

But the research has also made extensive use of CPOs. Perhaps the most famous is the discovery of a thin layer of clay—bereft of fossils—with a high concentration of iridium at the KT boundary 65 million years ago. The clay layer was discovered by the Berkeley geologist Walter Alvarez at Gubbio in Italy in the 1970s, and it led him to some striking speculations.

> It took me a while to realize that the thin bed of clay at the KT boundary at Gubbio not only raised the question of what had caused the mass extinction but also seemed to contradict the gradualistic mind-set of geologists. The near extinction of forams (microscopic sea animals) at Gubbio looked very abrupt. (Alvarez 1997, 59)

When the clay was studied using neutron activation analysis of the iridium, the results were startling. The concentration of iridium was 90 times greater than in the earth's crust. Alvarez had started with a CPO of a thin clay layer at the KT boundary, and now he had an additional CPO that revealed that the clay had a superabundance of iridium. Was this a local anomaly? Were there similar clay layers in other places? Where had the iridium come from?

Alvarez searched for another case with a clay layer and large amounts of iridium at the KT boundary, and he found it in Denmark. At this point, he started to think about what might have deposited the iridium. Having begun with one CPO, Alvarez sought another CPO that would be "diagnostic" in identifying the larger causal process of which the original CPO was a part. For example, was the iridium deposited by a supernova? That would have left a trace of plutonium-244, but no such CPO was found. What about an impact event—a comet or an asteroid containing high concentrations of iridium? But if it was an impact, where was the crater? After many false starts and the consideration of many DSOs and potential CPOs, attention now focused on the large Chicxulub crater (a large CPO!) in the Yucatan peninsula of Mexico. Moreover, Alvarez and his colleagues offered still another CPO, the occurrence of shocked quartz in the clay layer, that was consistent with the hypothesis of an impact event.

Alvarez's findings have not gone unchallenged, and there is no consensus in the scientific community about his explanation for the extinction or even about whether there was a catastrophic extinction.[12] Crucially for present purposes, both DSOs and CPOs have been volleyed back and forth in the debate. Moreover, theoretical efforts have turned toward considering all five major extinction events in the paleontological record and looking for general theories of extinctions. One evolving possibility is that although each event has common features, each also involves a somewhat different set of causal factors, suggesting that developing singular and general explanations may not be distinct scientific endeavors.

This case study further demonstrates the importance of CPOs in natural science. Most strikingly, one CPO, the clay layer, started out as an interesting observation begging to be interpreted as one step in a larger causal chain. Other CPOs involve markers (iridium), mechanisms (evidence of impact events), and mediating variables (the presence of large craters). The case also shows how efforts to explain a singular event can become part of a larger effort to explain multiple events.

5.2 *Examples from Epidemiology*

Research in epidemiology faces inferential challenges parallel to those in political science. Large-scale data sets consisting of DSOs are common, but these are observational data, with all the limitations that entails. As in many areas of political science, the quality of the data sets is unclear, yet with the large N it is quite easy—perhaps all too easy—to apply sophisticated estimation procedures. Specification of a compelling model is essential for such estimation to succeed; yet it is often difficult to arrive at such a model.

In the face of these challenges, the field of epidemiology has a well-established tradition of working with CPOs, employing what is variously called "medical detective work" and the use of "shoe leather" (Reingold 2001, 34). The researcher may focus on unexpected singular outcomes or may trace the victims of a disease, one by one, potentially thereby expending a great deal of shoe leather. These CPOs are commonly analyzed jointly with DSOs, hence combining these two sources of inferential leverage. The resulting findings may explain a particular outbreak of the disease being studied, but they also contribute to the broader explanatory and theoretical understanding of the disease.

5.2.1 Cholera

John Snow's research on the explanations of cholera in the 1840s and 1850s in London illustrates the creative use of CPOs and their juxtaposition with DSOs.[13] CPOs helped Snow make an initial evaluation of alternative mechanisms of transmission, such as the hypothesis that it was carried by "miasma" or bad air. Snow documented sequences of infections in specific individuals, who successively had close personal contact with each

---

[12]For those wanting to pursue the debate, Alvarez (1997) and Officer and Page (1996) offer two contrasting perspectives and Hallam and Wignall (1997, 210–22) and Palmer (1999) offer more dispassionate perspectives. Hallam and Wignall are more gradualist (Hallam has coauthored with Officer), whereas Palmer is more catastrophist. Nevertheless, Hallam and Wignall (213) do say that "The vulcanists failed, however, to account satisfactorily for the occurrence of shocked quartz, and by the end of the decade, ... the debate seemed to be substantially resolved in favour of the impact supporters. It would be a mistake, however, to consider that these are the only serious contenders in accounting for the end-Cretaceous extinctions, because several other Earthbound causes must also be discussed."

[13]In addition to Snow's (1936) own book, see Cameron and Jones (1983) and Paneth (2004); also Freedman (1991 and 2005, 6–9).

other (Snow 1936, 3–10). For example, one individual died from cholera after staying in the room in a boarding house previously occupied by a sailor, newly arrived in London, who had cholera. Snow used this and similar information to infer that cholera could be transmitted from person to person. Such CPOs helped Snow to discard environmental explanations, such as miasma, and to focus instead on other vectors by which the disease might travel.

Snow drew on details regarding the typical course of the disease, which almost always began with severe disruptions of the digestive system, producing what he concluded was infected human excrement and suggesting that cholera was transmitted by contact with a toxic substance contained in the excrement (Snow 1936, 10–6). To narrow the field of possible sources for that toxic substance, Snow considered the spatial clustering of cholera, showing that victims typically shared a water source contaminated with sewage. The most famous of these localized sources was the Broad Street pump in London, which appeared to have played a central role in the epidemic of 1854. The incidence of cholera was high among individuals who lived close to the pump, but Snow also observed the case of a woman who liked the taste of water from this pump so much that she had it brought to her home in a different area of London. The epidemic came to a close shortly after Snow convinced the authorities to remove the pump handle, thereby preventing people from using the contaminated water (Snow 1936, 19–56). This sequence of specific, localized CPOs, most dramatically the consequences of removing the pump handle, were diagnostic markers that linked the cause (contaminated water) to the effect (cholera).

These diverse CPOs strongly suggested that consumption of water polluted with excrement from people who already had cholera was a major vector for the transmission of the disease. To test this hypothesis, Snow employed a simple quantitative analysis, based on DSOs, in which he compared cholera death rates during the 1853 outbreak in two parts of London receiving water from two different companies: one company provided water from the Thames near London that was heavily mixed with city sewage and the other company provided relatively clean water from upstream. In areas using the polluted water from the first company, many more people died of cholera; where the second company supplied the water, few died of cholera. In further analysis that generated DSOs and provided the basis for statistical analysis, Snow did surveys of the water source and the incidence of cholera deaths, for a large number of individual households, yielding an even more precise assessment that further confirmed the hypothesis about the water-borne transmission of cholera.

5.2.2 Childbed Fever

Puerperal fever, also called childbed fever, is a form of blood poisoning that typically occurred in maternity wards. For centuries childbed fever was the most important cause of early death among women. The analysis that uncovered the source of this disease again mixed DSOs and a dramatic CPO, involving a single, unexpected observation of its occurrence.

In 1846, Ignaz Semmelweis was hired as an administrator at the Vienna General Hospital, specifically in the 19th-century equivalent of a maternity ward—one of two such clinics in the hospital. The clinic where Semmelweis worked had a serious problem that was readily seen in an analysis of simple DSOs with an N of 2: from 1839 until 1847, the maternal mortality rate in his clinic was more than two and a half times higher than in the second clinic. The excess mortality in the first clinic was due largely to a high incidence of childbed fever.

Two DSOs focused attention on the potential connection between childbed fever and the presence of doctors versus midwives in the clinics. Starting in 1839, the first clinic in this teaching hospital had been used for training aspiring doctors, whereas the other clinic—which had a far lower incidence of childbed fever—was used for training midwives. Further data likewise focused attention on this connection. During the prior period from 1833 to 1839, when medical students and midwives were commingled in both of the clinics, the clinics' maternal death rates were nearly indistinguishable.

What, then, was causing the differential death rate? Semmelweis obtained a valuable clue, involving a CPO, when a colleague Jakob Kolletschka died after an accident during an autopsy. As Semmelweis tells the story, this colleague

> often conducted autopsies for legal purposes in the company of students. During one such exercise, his finger was pricked by a student with the same knife that was being used in the autopsy ... the disease from which Kolletschka died was identical to that from which so many maternity patients died (Semmelweis [1860] 1983, 65).

From this isolated but crucial piece of evidence, the proper causal interpretation of the quantitative data on differences in maternal mortality rates between the two clinics became apparent. Medical students, but not midwives, routinely participated in autopsies. The students washed their hands with soap—but not with a disinfectant—and then proceeded to conduct obstetric examinations on expectant mothers. Semmelweis hypothesized that childbed fever was a result of the contamination of pregnant women with "decomposing animal organic matter" (Loudon 2000, 96), a concept that was a precursor to the germ theory of disease.

To prevent this contamination, Semmelweis in May of 1847 introduced a policy requiring all medical students and doctors to disinfect their hands using chloride of lime before examining pregnant women. After this new policy was introduced, the maternal mortality rates in the first clinic fell to the same levels as in the second clinic (Semmelweis [1860] 1983, 131).

In this instance, a small-N quantitative analysis of DSOs preceded a critical CPO, involving the accidental death of just one person. After a change in hospital policy that was motivated by this CPO, the subsequent analysis of DSOs suggested that the inference based on the CPO was correct. CPOs and DSOs were thus successfully adjoined, providing the basis for a successful inference about the causes of this dangerous disease.

5.3 *Political Science Examples*

In expressing his skepticism about CPOs, Beck discusses the three political science examples presented in *RSI* (256–8; Appendix): Brady's analysis of the 2000 presidential election in Florida ($N = 1$), Tannenwald's examination of the use/nonuse of nuclear weapons by the United States in four historical episodes from World War II to the Gulf War ($N = 4$), and Stokes's assessment of dramatic shifts in economic policy in Latin America ($N = 38$). In the following discussion, we briefly review these studies and then turn to a fourth example: Evan Lieberman's study of tax collection.

5.3.1 Brady, Stokes, and Tannenwald

We view the studies by Brady (*RSI*, Appendix), Stokes (2001), and Tannenwald (1999) as successfully adjoining DSOs and CPOs. For example, Brady seeks to make a causal inference about a singular outcome—i.e., whether the early media call of the election yielded a substantial vote loss by Bush in the Florida panhandle. Brady argues against

the large-N, regression-based study carried out by Lott, which played a critical role in the national debate on the election, given Lott's conclusion that Bush did indeed suffer a major vote loss due to the early media call. Brady disputes Lott's findings, basing his analysis on CPOs rather than DSOs, as in Lott's analysis, utilizing a form of quasi-detective work focused on a sequence of steps that would have had to occur for the vote loss to be plausible. Far from offering an informal analysis that impressionistically suggests the linkages between the putative cause (the early media call) and the supposed effect (lower turnout), Brady presents a carefully structured evaluation focused on a sequence of necessary conditions. These necessary conditions are not met, and the causal claim presented by the large-N study is effectively ruled out. Contra Beck, this is an instance of successfully adjoining CPOs and DSOs to address the same problem of inference.

Stokes is concerned with whether political leaders switch to neoliberal policies because they are convinced that this switch serves the wider economic needs of the country or whether the choice is driven by narrower rent seeking. In addition to exploring questions of this type with her N of 38, she closely examines three national cases, using CPOs to increase the plausibility of her wider finding. Do these CPOs definitively support her conclusion? Definitely not. Do the coefficients in her quantitative analysis definitively support her conclusion? Again, definitely not. Both make a contribution to her overall inference.

Tannenwald adopts a parallel strategy. She hypothesizes the existence of a "nuclear taboo" that helps to account for the nonuse of nuclear weapons by the United States in three international crises following WWII. Within her framework, the nuclear taboo did not yet exist in WWII itself (the fourth DSO), and the Unites States did use nuclear weapons. In the subsequent three crises, the taboo did exist, and nuclear weapons were not used. Support for the hypothesis is thus provided by a small group of DSOs, limited to an N of 4, which is obviously a weak basis for inference. In this context, direct information about decision processes—involving CPOs—is a reasonable alternative source of insight. Of course, decisions may have been made on the basis of criteria other than those reflected in the documents Tannenwald examined, and the criteria reflected in those documents might have involved dissembling on the part of decision makers. These CPOs certainly do not definitively rule out rival explanations—indeed, no available method will definitively rule them out. The crucial point is that in this study, both CPOs and the DSOs do serve to increase the plausibility of Tannenwald's line of explanation.

5.3.2 The Politics of Tax Collection

Lieberman's (2003) study of tax collection is of special interest here because it moves from a large-N comparison based on DSOs to an analysis of South Africa and Brazil that makes central use of CPOs and then back to a cross-national comparison that uses DSOs. Using the approach of "nested analysis" (Lieberman 2005), his book provides an extended and well-organized illustration of how these two forms of evidence can be adjoined.

In his initial, large-N analysis (71 cases) that frames the study, Lieberman shows that gross domestic product (GDP) per capita is a reasonably good predictor of state capacity to tax the income of its citizens. Interestingly, the two countries of particular concern in his study are predicted to have similar capacities to tax income. Yet South Africa's capacity for taxation is substantially higher than its GDP would predict and Brazil's substantially lower. A central challenge is to account for these discrepancies, thereby potentially generating explanations that might provide further insight into the larger data set.

Lieberman's analysis of the two countries is in part a paired comparison, and this N of 2 does create DSOs. Yet with such a small N, the paired comparison in many ways makes its greatest contribution through helping to conceptualize and frame the analysis, rather than providing tests of hypotheses.

Instead, a crucial inferential tool for South Africa and Brazil is to trace out a series of causal steps that link a hypothesized cause in the study—alternative conceptions of the national political community—to tax policy and tax collection. One of the key contrasts in these conceptions is between a more exclusionary definition of the community, as in South Africa, such that elites are more willing to submit to taxation because there is a narrower scope of beneficiaries and because a more fiscally sound state helps to enforce the exclusion, as opposed to a more inclusionary definition, as in Brazil, where elites resist taxation to a greater degree because the set of beneficiaries is more diffuse. The links between this explanation and the tax outcomes are complex, and Lieberman employs numerous CPOs to evaluate these steps and assess their place in his larger explanation.

In the penultimate chapter, Lieberman returns to a larger cross-national analysis, focusing on a subset (21 cases) of the initial large-N comparison and seeking to test his new arguments against a wider range of cases. This cross-national subset involves countries that Louis Hartz called "fragment societies," a concept appropriate to Lieberman's analysis because of its central relevance to alternative conceptions of the national political community. This return to DSO-based analysis provides further support for Lieberman's argument.

Thus, Lieberman initially frames his research question with DSOs, gains insights into how cause reaches effect through the analysis of CPOs in Brazil and South Africa, and then uses DSOs to show that these same insights have explanatory power in a much larger set of cases. This study adjoins DSOs and CPOs—indeed, in a highly disciplined manner.

6 Conclusion

Both data-set observations and causal-process observations are important for causal inference. DSOs make their contribution through a quantitative logic of comparison, routinely using regression analysis. DSOs are deservedly a major research tool in political science. Yet as we argue above, and as suggested by Schrodt (this symposium), when DSOs are based on observational data, they must be analyzed with great care. Many problems—very centrally that of confounders—are simply hard to address. DSOs are valuable, but they must be analyzed with an eye to these limitations.

CPOs provide a different form of analytic leverage but leverage that can definitely be rigorous. Our examples from natural science and epidemiology show that CPOs do play a crucial role in such domains, which of course are more strongly identified with analytic rigor than are the social sciences. The rigor associated with CPOs depends on a careful specification of the inferences to be drawn from the presence or absence of particular CPOs as diagnostic markers. Of course, information that could be treated as the basis for CPOs may simply be presented in a narrative form, with a different analytic framing than we have in mind here. Such narratives can provide valuable insight, and they may play a key role in research. But our concern is with carefully constructed arguments as to why particular CPOs—either arrived at serendipitously or deliberately sought—are especially diagnostic. When such arguments are lacking, CPOs may not be very useful for causal inference. When they are present, CPOs can be a powerful tool.

Bringing together DSOs and CPOs in a single study can be especially powerful. It could, of course, become a cliche to assert this is a virtuous research strategy. But as our

examples have shown, major studies have indeed adjoined these alternative kinds of data, thereby achieving strong inferential leverage.

References

Achen, Christopher H. 2002. Toward a new political methodology: Microfoundations and ART. *Annual Review of Political Science* 5:423–50.

Alvarez, Walter. 1997. *T. rex and the crater of doom.* Princeton, NJ: Princeton University Press.

Cameron, D., and I. G. Jones. 1983. John Snow, the broad street pump and modern epidemiology. *International Journal of Epidemiology* 12:393–6.

Coppedge, Michael. 2005. Explaining democratic deterioration in Venezuela through nested inference. In *The third wave of democratization in Latin America*, ed. Frances Hagopian and Scott Mainwaring, 289–316. New York: Cambridge University Press.

Freedman, David A. 1991. Statistical models and shoe leather. *Sociological Methodology* 21:291–313.

———. 2005. *Statistical models: Theory and practice.* Cambridge: Cambridge University Press.

Goldthorpe, John H. 2001. Causation, statistics, and sociology. *European Sociological Review* 17:1–20.

Hallam, A., and P. B. Wignall. 1997. *Mass extinctions and their aftermath.* Oxford: Oxford University Press.

Hix, Simon. 2004. A global ranking of political science departments. *Political Studies Review* 2:293–313.

Koestler, Arthur. 1963 [1959]. *The sleepwalkers: A history of man's changing vision of the universe.* Universal Library ed. New York: Grosset & Dunlap.

Kuhn, Thomas. 1957. *The Copernican revolution.* Cambridge, MA: Harvard University Press.

Lachieze-Rey, Marc, and Edgard Gunzig. 1999. *The cosmological background radiation.* Cambridge: Cambridge University Press.

Laitin, David D. 2002. Comparative politics: The state of the subdiscipline. In *Political science: The state of the discipline*, ed. Ira Katznelson and Helen V. Milner, 630–59. New York: Norton.

Lieberman, Evan S. 2003. *Race and regionalism in the politics of taxation in Brazil and South Africa.* New York: Cambridge University Press.

———. 2005. Nested analysis as mixed-method strategy for comparative research. *American Political Science Review* 99:435–52.

Loudon, Irvine. 2000. *The tragedy of childbed fever.* Oxford: Oxford University Press.

Mackie, J. L. 1974. *The cement of the universe, a study of causation.* Oxford: Clarendon Press.

Officer, Charles, and Jake Page. 1996. *The great dinosaur extinction controversy.* Reading, MA: Addison-Wesley Publishing Company.

Palmer, Trevor. 1999. *Controversy: Catastrophism and evolution—The ongoing debate.* New York: Kluwer Academic.

Paneth, N. 2004. Assessing the contributions of John Snow to epidemiology: 150 years after removal of the broad street pump handle. *Epidemiology* 15:514–6.

Ragin, Charles C. 1987. *The comparative method: Moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.

Reingold, Arthur L. 2001. Field epidemiology training in Europe. *Euro surveillance: European communicable disease bulletin* 6:33–4.

Semmelweis, Ignaz. [1860] 1983. *The etiology, concept, and prophylaxis of childbed fever.* Translated by K. Codell Carter. Madison, WI: University of Wisconsin Press.

Snow, John. 1936. *Snow on cholera.* London: Humphrey Milford, Oxford University Press.

Stokes, Susan Carol. 2001. *Mandates and democracy: Neoliberalism by surprise in Latin America.* New York: Cambridge University Press.

Tannenwald, Nina. 1999. The nuclear taboo: The United States and the normative basis for nuclear nonuse. *International Organization* 53:433–68.