Process Tracing

From Metaphor to Analytic Tool

Edited by

Andrew Bennett
Georgetown University

Jeffrey T. Checkel
Simon Fraser University

Contents

8

Improving process tracing

The case of multi-method research

Thad Dunning

Introduction

Social scientists increasingly champion multi-method research – in particular, the use of both quantitative and qualitative tools for causal inference.[1] Yet, what role does process tracing play in such research? I turn in this chapter to natural experiments, where process tracing can make especially useful and well-defined contributions. As I discuss, however, several lessons are relevant to other kinds of multi-method research.

With natural experiments, quantitative tools are often critical for assessing causation.[2] Random or "as-if" random assignment to comparison groups – the definitional criterion for a natural experiment – can obviate standard concerns about confounding variables, because only the putative cause varies across the groups. Other factors are balanced by randomization, up to chance error. Simple comparisons, such as differences of means or percentages, may then validly estimate the average effect of the cause, that is, the average difference due to its presence or absence. Controlling for confounding variables is not required, and can even be harmful.[3]

However, much more than data analysis is needed to make such research compelling. In the first place, researchers must ask the right research questions and formulate the right hypotheses; and they must create or discover research designs and gather data to test those hypotheses. Successful quantitative analysis also depends on the validity of causal models, in terms of which hypotheses are defined. The formulation of questions, discovery of strong designs, and validation of models require auxiliary information,

[1] See *inter alia* Brady and Collier 2010; Bennett 2007; Dunning 2008b, 2010, 2012.
[2] Natural experiments are observational studies – those lacking an experimental manipulation – in which causal variables are assigned at random or *as-if* at random. See Freedman 1999, 2009; Dunning 2008a, 2012; or Angrist and Pischke (2008).
[3] Freedman 2008a, 2008b, 2009; also Dunning 2012; Sekhon 2009; or Gerber and Green (2012).

which typically does not come from analysis of large data sets (Freedman 2010). This depends instead on disparate, qualitative fragments of evidence about context, process, or mechanism. What has come to be called process tracing – that is, the analysis of evidence on processes, sequences, and conjunctures of events within a case for the purposes of either developing or testing hypotheses about causal mechanisms that might causally explain the case (Bennett and Checkel, this volume, X) – is the major means of uncovering such pieces of diagnostic evidence, which Collier *et al.* (2010) describe as "causal-process observations" (CPOs).

For instance, researchers using natural experiments face the challenge of validating the definitional claim of as-if random. To do so, they require evidence on the *information, incentives,* and *capacities* of key actors with control over *processes* of treatment assignment (Dunning 2012, chapter 7). This helps them to assess whether actors had the desire and ability to undercut random assignment.[4] To appraise central assumptions of standard causal and statistical models (e.g., "no interference"), researchers may use qualitative information on the mode and possible effects of interactions between units in the treatment and control groups. Finally, deep engagement with research contexts – even "soaking and poking" (Bennett and Checkel, this volume, XX) – can generate the substantive knowledge required to discover the opportunity for a natural experiment, as well as to interpret effects. Qualitative evidence may thus be a requisite part of successful quantitative analysis, and it can also make vital contributions to causal inference on its own (Brady and Collier 2010).

Despite these virtues, scholars have encountered challenges in developing process tracing tools for multi-method research. In the first place, recognition of the general utility of evidence on context or process does not provide researchers with a ready guide to practice. It is one thing to say that causal-process observations play a critical role in causal inference; it is quite another to say which *particular* CPOs are most persuasive or credible (Bennett and Checkel, this volume, XX–XX; see also Jacobs, this volume; Waldner, this volume). Indeed, qualitative evidence can also lead researchers astray, as useful examples from the biological sciences surveyed in this chapter's third section suggest. Reflecting such concerns, methodologists have focused more centrally on the challenge of distinguishing more and less valid CPOs. The elaboration of standards to evaluate process tracing is an important part of improving research practice – in terms of

[4] When "as-if random" fails, treatment assignment is not independent of potential outcomes – that is, the hypothetical outcomes each unit *would* experience if exposed to treatment or control.

this volume's title, of turning process tracing from metaphor to analytic tool (see also Collier 2011).

There are several difficulties with implementing such standards, however. Bennett and Checkel's recommendation that process tracers "cast the net widely for alternative explanations" seems critical, as is their suggestion to "be equally tough on the alternative(s)" (Bennett and Checkel, this volume, XX–XX). Yet it is not easy to demonstrate adherence to such advice. One challenge is that absence of evidence does not constitute evidence of absence (Bennett and Checkel, this volume, XX). With natural experiments, failure to find information that disproves, say, the assumption of as-if random does not constitute positive proof of its validity. Thus, and to invoke Van Evera's (1997) framework, "hoop" or "straw in the wind" tests of the assumption of "as-if random" assignment appear common: passing such tests increases the plausibility of as-if random, and may be necessary for as-if random to hold, but it is not alone sufficient.

In contrast, "smoking gun" or "doubly-decisive" evidence in favor of as-if random is rare – though smoking gun evidence that assignment was not random may be sufficient to cast serious *doubt* on the assertion of as-if random assignment.[5] In other words it is much easier to *disprove* the hypothesis of "as if random" than it is to provide sufficient evidence in *favor* of this hypothesis. Thus, researchers may triangulate between quantitative tests – e.g., evaluating whether assignment to categories of a treatment variable is *consistent* with a coin flip[6] – and qualitative information on the assignment process. Yet, without true randomization, there is always the possibility of lurking qualitative or quantitative information that would undermine the plausibility of random assignment, if it were only uncovered. Casting the net widely for alternative evidence on the process that assigns units to treatment or control groups is crucial; yet it can be difficult for researchers to know when they have cast the net widely enough.

Implementing standards for process tracing raises other epistemological and practical obstacles, related in part to the esoteric information that is often required. As with the discovery of natural experiments, process tracing typically requires deep substantive

---

[5] Recall that pieces of evidence are judged to have passed a "straw in the wind test" if they merely increase the plausibility that a hypothesis is true; a "hoop test" if passing does not confirm the hypothesis but failing to pass disconfirms it; a "smoking gun test" if passing confirms the hypothesis (but not passing does not disconfirm it); and a "doubly-decisive test" if passing confirms, and not passing disconfirms, the hypothesis. See Bennett and Checkel (this volume, XX–XX) and Bennett in the Appendix (this volume).

[6] For example, they may demonstrate that the treatment and control groups are *balanced* on measured, pre-treatment covariates, just as they would be (in expectation) in a randomized experiment.

engagement with disparate research contexts; it involves sifting through both confirmatory and potentially falsifying pieces of evidence. Our expectations about the accessibility of evidence on the information, incentives, or capacities of key actors may also affect our assessment of the probative value of any qualitative fact (Bennett and Checkel, this volume, XX; see also Jacobs, this volume). Yet precisely in consequence of the deep engagement and specialized knowledge required, the number of scholars who possess the requisite knowledge to evaluate the quality of process tracing may be small.[7]

This creates challenges relating to the ways in which observations on causal process are reported and evaluated by communities of scholars. Contrary evidence that would invalidate a given research design or causal model may indeed exist in the historical record or at a given field site. Unless it is elicited and offered by scholars, however, readers cannot use it to evaluate the persuasiveness of the process tracing. How are we in the community of scholars to know whether individual researchers have indeed sufficiently canvassed the available evidence – both supportive and potentially disconfirming? And how can researchers successfully demonstrate that they have done so, thereby bolstering the transparency and credibility of their findings (see also Waldner, this volume)?

In this chapter, I discuss these challenges further, focusing on both the promise and the pitfalls of process tracing in multi-method research. I begin by describing two ways in which process tracing may help validate design and modeling assumptions in natural experiments: through the discovery of what I have called treatment-assignment CPOs and the testing of model-validation CPOs (Dunning 2012: chapter 7).[8] While my illustrative examples show how qualitative evidence has been used productively in studying natural experiments, the discussion is also aspirational. In many studies, observations on causal process could be used more effectively to assess design and modeling assumptions.

I then turn to the challenge of appraising the quality of process tracing, describing in more detail the epistemological and practical difficulties mentioned above. I argue that while the formulation of *best practices* for what constitutes good process tracing is appealing – and the criteria suggested by Bennett and Checkel in Chapter 1 are

[7] The number of scholars who possess the interest or expertise to evaluate critically both the quantitative and qualitative analysis may be even smaller.

[8] As I make clear below, qualitative evidence has many important roles to play besides bolstering the validity of quantitative analysis; however, this plays an especially critical role in multi-method research.

excellent—they may also be quite difficult to apply and enforce. In other words, it may be challenging to develop general criteria with which to evaluate the persuasiveness and evidentiary standing of given pieces of qualitative evidence, or specific instances of process tracing. For individual researchers, it may also be difficult to demonstrate that they have adhered to those criteria – though specific attention in published work to Bennett and Checkel's best practices would undoubtedly help.

Instead, I argue that it may be productive to focus on research *procedures* that can bolster the credibility of process tracing in multi-method research. A central question is whether and to what extent such procedures can facilitate open scholarly contestation about the probative value of qualitative evidence. Thus, transparent procedures – including the cataloguing of interview transcripts or archival documents – should assist scholars with relevant subject-matter expertise in debating the evidentiary weight to accord to specific observations on causal process (Moravcsik, 2010).

The idea that scholarly contestation can improve the quality of process tracing has some parallels in the theory of "legal adversarialism," in which competing advocates offer evidence in support of different theories. While this seems unlikely to produce perfect validation of qualitative evidence – for reasons that reflect not just basic epistemological difficulties but also the sociological organization of scholarly production – I pinpoint a few major challenges and suggest some modest proposals for improving process-tracing practice. These research procedures can complement the process tracing best practices outlined by Bennett and Checkel and bolster our confidence that researchers have indeed adhered to several of those criteria.

Process tracing in multi-method research

How does process tracing contribute to causal inference in multi-method research? I use process tracing, as do Bennett and Checkel (this volume, XX), to denote a procedure for developing knowledge of context, sequence, or process – essentially, for generating causal-process observations (CPOs). Collier *et al.* describe a causal-process observation as "an insight or piece of data that provides information about context, process, or mechanism" (2010: 184).[9] At times, CPOs function like clues in detective stories (Collier 2011), playing the role of "smoking guns" (Collier *et al.* 2010: 185); at other times, they are simply pieces of contextual information upon which researchers can draw to evaluate

[9] These are contrasted with quantitative "data-set observations" (DSOs) i.e., the collection of values on the dependent and independent variables for a single case, i.e., a row of a "rectangular data set."

particular assumptions or hypotheses. My usage of process tracing also follows Mahoney (2010: 124), who notes "process tracing contributes to causal inference primarily through the discovery of CPOs."

In multi-method as in other forms of research, process tracing can in principle help confront two key challenges in making causal inferences:

> 1. the challenge of understanding the selection process that assigns units to categories of causal/treatment variables (i.e. levels of an independent variable);
>
> 2. the challenge of model validation, for instance, validation of assumptions about causal process that are embedded in quantitative models.

On the former, in observational studies, where treatment assignment is not under the control of an experimental researcher, confounding variables associated with both a putative cause and effect may play an important role. Control of confounding variables thus requires close understanding of selection processes, and ideally, the discovery of research settings in which assignment to the causal variable is independent of other variables that may influence outcomes – as in strong natural experiments.

Regarding (2), in experiments and observational studies alike, quantitative analysis proceeds according to maintained hypotheses about causal process. Yet, if these hypotheses are wrong, the results cannot be trusted (Freedman 2009). Finding ways to probe the credibility of modeling assumptions is thus a critical part of successful quantitative inference.

For both of these challenges, qualitative evidence can and should play a crucial role. Here, I would emphasize both *can* and *should*; in practice, qualitative evidence is not always (or even usually) explicitly deployed in this fashion, and some qualitative evidence may not always contribute decisively or productively to causal inference. Thus, a major theme for consideration is how qualitative evidence could be deployed more effectively in such settings to bolster causal inference (see also Schimmelfennig, this volume).

To develop these ideas further, it is useful to introduce two running examples: Snow's famous study of cholera (Freedman 1999, 2009; see also Dunning 2008, 2012) and the Argentina land-titling study of Galiani and Schargrodsky (2004, 2010). Snow used a natural experiment to study the causes of cholera transmission, and to test hypotheses engendered by a series of causal-process observations. His study was occasioned by the move of the intake pipe of the Lambeth Water Company to a purer water source, higher up-river on the Thames, prior to a cholera outbreak in 1853–54; a competitor, Southwark & Vauxhall, left its own pipe in place, lower on the Thames and downstream from more sewage outlets. According to Snow, the move of the Lambeth Company's water pipe meant that more than three hundred thousand people were:

> divided into two groups without their choice, and, in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and, amongst it, whatever might have come from the cholera patients, the other group having water quite free from such impurity. (Snow 1855: 75)

The contrast in death rates from cholera was dramatic: the household death rate among Lambeth customers was 37 per 10,000, compared to 315 per 10,000 among customers of Southwark & Vauxhall (Freedman 2009). Why this study design provided a compelling natural experiment is discussed in the next subsection, but Snow touted it thus: "It is obvious that no experiment could have been devised which would more thoroughly test the effect of water supply on the progress of cholera than this" (Snow 1855: 74–75).

The Argentina land-titling study provides another example, with a design quite similar to Snow's. In 1981, squatters organized by the Catholic Church occupied an urban wasteland on the outskirts of metropolitan Buenos Aires, dividing the land into similar-sized parcels that were allocated to individual families. After the return to democracy in 1983, a 1984 law expropriated this land, with the intention of transferring title to the squatters. However, some of the original owners challenged the expropriation in a series of court cases, leading to delays of many years in the transfer of titles to the plots owned by those owners. Other titles were transferred to squatters immediately. The legal action therefore created a "treatment" group – squatters to whom titles were ceded immediately – and a "control" group – squatters to whom titles were not ceded. As in Snow's study, nearby households found themselves exposed in an apparently haphazard way to different treatment conditions. Galiani and Schargrodsky (2004, 2010) find significant differences across the groups in subsequent housing investment, household structure, and educational attainment of children, though not in access to credit markets (thus contradicting De Soto's (2000) theory that the poor will use *de jure* property rights to collateralize debt).

In both of these studies, qualitative information about context and process plays a number of critical roles. In the first place, such information is crucial for recognizing the existence of natural experiments. Indeed, substantive knowledge and "shoe leather" work is typically a *sine qua non* for discovering the opportunity for such research designs

(Freedman 1991). Yet, process tracing and causal-process observations can also play an especially important role with respect to the two challenges (1) and (2) noted above, as I now describe.

Treatment-assignment CPOs

Regarding challenge (1), understanding the process of selection into categories of an independent variable is vital for evaluating threats to valid causal inference from confounding variables. Here, treatment-assignment CPOs (Dunning 2012) – pieces or nuggets of information about the *process* by which units were assigned to treatment and control conditions – are critical.

For example, qualitative evidence on the process of treatment assignment plays a central role in Snow's study. Information on the move of Lambeth's water pipe and, especially, on the nature of water markets helped to substantiate the claim that assignment of households to source of water supply was *as-if* random – the definitional criterion for a natural experiment. The decision of Lambeth Waterworks to move its intake pipe upstream on the Thames was taken before the cholera outbreak of 1853–54, and contemporary scientific knowledge did not clearly link water source to cholera risk.[10]

Yet, there were some important subtleties. The Metropolis Water Act of 1852, which was enacted in order to "make provision for securing the supply to the Metropolis of pure and wholesome water," made it unlawful for any water company to supply houses with water from the tidal reaches of the Thames after August 31, 1855. While Lambeth's move was completed in 1852, Southwark & Vauxhall did not move its pipe until 1855.[11] In other words, Lambeth chose to move its pipe upstream earlier than it was legally required to do, while Southwark & Vauxhall opted to keep its pipe in place; for the companies, assignment to water supply source was self-selected. In principle, then, there could have been confounding variables associated with choice of water supply – for instance, if healthier, more adept customers noticed Lambeth's move of its intake pipe and switched water companies.

[10] The directors of the Lambeth Company had apparently decided to move the intake for their reservoirs in 1847, but facilities at Seething Wells were only completed in 1852. See *Lambeth Waterwork History*, UCLA Department of Epidemiology. URL.

[11] To comply with the legislation, the Southwark & Vauxhall Company built new waterworks in Hampton above Molesey Lock in 1855. Ibid.

Here, qualitative knowledge on the nature of water markets becomes crucial. Snow emphasizes that many residents in the areas of London that he analyzed were renters; also, absentee landlords had often taken decisions about water supply source years prior to the move of the Lambeth intake pipe. Moreover, the way in which the water supply reached households – with heavy interlocking fixed pipes making their way through the city and serving customers in side-by-side houses – also implied a limited potential for customer mobility, since landlords had either signed up for one company or another (presumably when the pipes were being constructed). As Snow put it,

> A few houses are supplied by one Company and a few by the other, *according to the decision of the owner or occupier at that time when the Water Companies were in active competition.* (Snow 1855: 74–75, italics added)

This qualitative information thus suggests that residents did not largely self-select into their source of water supply – and especially not in ways that would be plausibly related to death risk from cholera. Even if the *companies* chose whether to move their intake pipes upstream or not, as Snow emphasizes, *households* were assigned sources of water supply without their choice, and often without their knowledge. Thus, qualitative knowledge on water markets is critical to buttressing the claim that assignment to water supply source was as good as random for households – in particular, that it was not linked to confounding variables that might explain the dramatic difference in death rates across households served by either company.

In the Argentina land-titling study, qualitative evidence on the process by which squatting took place, and plots and titles were obtained, also plays a central role. Recall that squatters organized by Catholic Church activists invaded the land in 1981, prior to the return to democracy in 1983. According to Galiani and Schargrodsky (2004), both church organizers and the squatters themselves believed that the abandoned land was owned by the state, not by private owners; and neither squatters nor Catholic Church organizers could have successfully predicted which *particular* parcels would eventually have their titles transferred in 1984 and which would not. Thus, industrious or determined squatters who were particularly eager to receive titles would not have had reason to occupy one plot over another – which helps to rule out alternative explanations for the findings whereby, for instance, organizers allocated parcels to certain squatters, anticipating that these squatters would one day receive title to their property. Nor did the quality of the plots or attributes of the squatters explain the decisions of some owners and not others to challenge expropriation in court. On the basis of their interviews and other qualitative fieldwork, the authors argue that idiosyncratic factors explain these decisions. In sum, evidence on the *process* of treatment assignment suggests that potentially confounding characteristics of squatters that might otherwise explain differences in housing investment or household structure – such as family background, motivation, or determination – should not be associated with whether they received title to their plots.[12]

For both the cholera and land-titling studies, such evidence does *not* come in the form of systematic values of variables for each squatter – that is, as data-set observations (DSOs). Instead, it comes in the form of disparate contextual information that helps validate the claim the treatment assignment is as good as random – in other words, causal-process observations (CPOs). To be sure, Galiani and Schargrodsky (2004) also use quantitative tests of their design assumptions, for instance, assessing whether balance on pre-treatment covariates across the treatment and control groups is consistent with a coin flip.[13]

Yet, qualitative evidence on the process of treatment assignment is just as critical: fine-grained knowledge about context and process is crucial for bolstering the case for as-if random assignment. In Snow's study, causal-process observations are also central to supporting the claim of as-good-as-random assignment – and causal-process observations would likely be needed to challenge Snow's account as well.[14] In many other natural experiments, qualitative evidence is also critical for validating the assertion of as-if random.[15]

It is useful to note here that understanding the process of assignment to treatment and control conditions is also critical in other kinds of research – including conventional observational studies (i.e., those that lack plausible random assignment). For instance, researchers may use multivariate regression (or analogues such as matching) to compare units with similar values of covariates (age, sex, and so on) but different exposure to treatment conditions. There, analysts typically assume that within groups defined by the covariates, treatment assignment is as good as random (i.e., that "conditional independence" holds). Yet, why would this be? Along with *a priori* arguments, qualitative

---

[12] Thus, potential outcomes – the outcomes each squatter *would* experience under assignment to a title or assignment to the control group – should be independent of actual assignment.

[13] For instance, characteristics of both squatters and parcels are similar across the treatment and control groups; see Galiani and Schargrodsky (2004, 2010).

[14] For instance, evidence that customers did switch companies after Lambeth's move of its pipe might undercut the claim of *as-if* random. This evidence might come in the form of DSOs or CPOs.

[15] Dunning (2012, chapter 7) provides further examples.

evidence on the process of treatment assignment – that is, process tracing – is critical for making this assertion credible, and thus for heightening the plausibility of causal inferences drawn from the analysis. Explicitly addressing this process element may not be typical, but it is no less important in conventional observational studies than in natural experiments – even if, in many cases, as-if random assignment is unlikely to hold even within matched groups.

Model-validation CPOs

Just as important in multi-method research as understanding selection into treatment is the specification of the causal model – that is, the *response schedule* that says how units respond to hypothetical manipulations of a treatment variable (Freedman 2009). Before a causal hypothesis is formulated and tested quantitatively, a causal model must be defined, and the link from observable variables to the parameters of that model must be posited. Thus, the credibility and validity of the underlying causal model is always at issue.

Hence stems the importance of (2) Model-Validation CPOs, that is, nuggets of information about causal process that support or invalidate core assumptions of causal models. As one example, both the Neyman causal model (also known as the Neyman-Rubin-Holland or potential outcomes model) and standard regression models posit that potential outcomes for each unit are invariant to the treatment assignment of other units (this is the so-called "no-interference" assumption).[16] Yet, how plausible is this assumption? Close examination of patterns of interaction between units – for instance, the information they possess about the treatment-assignment status of others – can heighten or mitigate concerns about such modeling assumptions.

Consider, for example, the Argentina land-titling study. A key hypothesis tested in this study is that land titling influenced household structure – in particular, fertility decisions by teenagers. The study indeed provides some evidence that titled households had fewer teenage pregnancies. Yet, does the difference between titled and untitled households provide a good estimator for the causal effect of interest—namely, the difference between average pregnancy rates if all households were assigned titles and average pregnancy rates if no households were assigned titles? It does not, if fertility decisions of people in untitled households are influenced by the assignment of titles to

[16] Following Rubin (1978), this is sometimes called the Stable Unit Treatment Value Assumption (SUTVA).

their neighbors in the treatment group. Indeed, if titling also influences neighbors in the control group to have fewer children, then comparing pregnancy rates in titled and untitled households does not provide a reliable guide to the causal effect of interest.

The key point is that the plausibility of the assumption that no such "interference" between treatment and control groups exists could in principle be investigated using a range of methods – including process tracing. For example, detailed knowledge of interactions between neighbors – insights into how fertility decisions of households are linked to those of other squatters – may be quite helpful for assessing the extent to which interference poses obstacles for successful inferences about average causal effects.[17] In Snow's study, too, non-interference would be important to establish: are cholera death rates in households served by Southwark & Vauxhall plausibly influenced by the assignment of neighbors to water from Lambeth? Information on causal process can also help researchers appraise other modeling assumptions, such as whether or not there is clustered assignment to treatment conditions.[18]

In sum, process tracing can contribute to buttressing or undermining the validity of such modeling assumptions – and can therefore play a critical role in experiments and natural experiments, as in conventional observational studies. Of course, modeling assumptions are just that – assumptions – and they are therefore only partially subject to verification. Researchers would do well to heed Bennett and Checkel's advice to "be tough on alternative explanations" and "consider the potential biases of evidentiary sources" (this volume, XX–XX). This will include imagining the ways in which their modeling assumptions might go off the rails in a given substantive context – say, by considering how interference between treatment and control groups could arise. This is not easy to do. Yet, the examples in this section illustrate the important contribution that qualitative evidence obtained via process tracing can make to quantitative analysis – thus suggesting how multi-method work may, in principle, lead to more valid causal inferences.

The challenges of validation

[17] Of course, quantitative measures of interactions between neighbors (DSOs) – for instance, survey self-reports in which respondents are asked systematically about interactions with their neighbors – may be useful as well, a point further developed by Bennett and Checkel in their concluding chapter (Chapter 10, this volume).

[18] For further examples of treatment-assignment and model-validation CPOs, see Dunning 2012, especially chapters 7–9.

Despite the merits of process tracing, the examples in the previous section also suggest important difficulties that confront the use of causal-process observations. Process tracing can certainly "generate a line of scientific inquiry, or markedly shift the direction of the inquiry by overturning prior hypotheses, or provide striking evidence to confirm hypotheses" (Freedman 2010: 338). Yet, it can also lead researchers down the wrong path. The medical sciences provide useful illustrations. Snow, for instance, did not stop with cholera. In fact, Snow also believed:

> by analogy with cholera [that] plague, yellow fever, dysentery, typhoid fever, and malaria … were infectious waterborne diseases. His supporting arguments were thin. As it turns out, these diseases are infectious; however, only dysentery and typhoid fever are waterborne. (Freedman 2010: 353)

Another example comes from James Lind, who carried out an experiment of sorts in 1747 to show that the absence of citrus fruits is a cause of scurvy. Lind assigned twelve sailors suffering from scurvy to ingest different nutritional supplements, with two sailors assigned to each of six treatment regimes: (1) a daily quart of cider; (2) twenty-five gutts of elixir vitriol, three times a day; (3) two spoonfuls of vinegar, three times a day; (4) a course of sea water; (5) nutmeg, three times a day; or (6) two oranges and one lemon each day. At the end of a fortnight, "the most sudden and visible good effects were perceived from the use of the oranges and lemons" (Lind, cited in De Vreese 2008: 16).[19]

According to De Vreese (2008), Lind rejected the evidence from his own experiment because he could not imagine mechanisms linking nutritional deficiencies to scurvy. Rather, his explanatory framework, inherited from the eighteenth-century theory of disease, focused on how moisture blocked perspiration, thought to be vital for inhibiting disease. Lind thought that lemons and oranges counteracted this property of moisture. Instead, he focused on humidity as the ultimate cause of scurvy, due to a series of observations apparently consistent with his theory (moisture constricting skin pores, leading to corrupted fluids in the body). Lind's focus on a wrongly identified mechanism – apparently supported by causal-process observations – thus led him astray.

Such discouraging examples raise important questions, not only about how to validate hypotheses generated by causal-process observations, but also how to distinguish useful from misleading process tracing.[20] As Freedman puts it, "If guesses cannot be

[19] Note that treatment assignment was not randomized; and with only 12 sailors, chance variation would have been pronounced. I am grateful to David Waldner for suggesting the De Vreese reference.

[20] Waldner (this volume) also explicitly addresses this issue, in his case, by advocating a "completeness standard" for process tracing. Also, Bennett, in the Appendix (this volume) notes that in some

verified, progress may be illusory" (2010: 353). Success stories demonstrate the importance of qualitative evidence. Yet, few recent writings on qualitative or mixed-method research describe misleading qualitative observations that lead scholars in the wrong direction.[21]

There seem to be two major challenges. First, how can we appraise the value of any discrete piece of evidence offered in support of a design assumption or a substantive conclusion – without knowledge of other relevant diagnostic pieces of evidence, or additional confirmatory testing? In fact, as I argue below, the probative value of a causal-process observation often depends on the existence or non-existence of certain *other* pieces of evidence, as well as analysis of data sets from strong designs. Yet, this leads to a second issue, because the full set of potential diagnostic evidence may or may not be elicited and reported by individual researchers. We thus face important challenges in terms of how we as individual researchers – and as a research community – can best operationalize Bennett and Checkel's injunction to "cast the net widely for alternative explanations" (this volume, XX–XX).

Consider, first, the probative value of particular causal-process observations, taking Snow's compelling examples as illustrations. In one cholera epidemic, Snow found that the second person known to die from cholera had taken a room in a boarding house previously occupied by a deceased boarder, who was the epidemic's first recorded case – plausibly suggesting that cholera might have spread from the first to the second boarder through infected waste. In his famous study of the Broad Street pump, Snow probed several anomalous cases. Households located near the pump where no one died from cholera turned out to take water from another source, while some households that experienced cholera deaths but lived further away turned out, for disparate reasons, to have taken water from the Broad Street pump. This heightened the plausibility of Snow's inference that infected water from the pump was spreading the disease

Finally, Snow noted that sailors who docked at cholera-affected ports did not contract the disease until they disembarked, striking a blow to the prevailing theory that cholera travels via miasma (bad air). According to this theory, sailors should have contracted cholera by breathing bad air before coming ashore. As a whole, these fragments of evidence are convincing. Combined with Snow's natural experiment, they

circumstances evidence consistent with an hypothesis can actually lower the likelihood that the hypothesis is true.

21 Freedman's (2010) account could be accused of a mild form of this selection bias: it mainly narrates success stories, in which qualitative insights led to important medical discoveries.

lead strongly to the inference that cholera spreads through infected waste or water (even if this conclusion was not fully accepted by epidemiologists for another fifty years).

The persuasiveness and probative value of each single causal-process observation is nonetheless debatable. For example, other evidence appeared consistent with the miasma theory (e.g., territorial patterns of the disease's spread); and Snow's evidence from sailors does not suggest that cholera is borne by infected waste or water. The taxonomy of process-tracing tests proposed by Van Evera (1997) and discussed and further elaborated both in Chapter 1 and the Appendix provides a helpful way of organizing our thinking about the strength of Snow's evidence. Some of his CPOs appear to provide "straw in the wind" tests, others might be "hoop" tests.[22] However, such taxonomies do not provide a ready guide for assessing whether any *particular* CPO provides strong evidence in favor of his hypothesis (see also Waldner, this volume). Researchers would be left to argue that a particular piece of evidence is indeed a "smoking gun" or is "doubly decisive;" and their readers may lack firm criteria for deciding when these claims are true.

Moreover, there is a potential circularity involved in assessing the probative value of particular CPOs in light of subsequent testing. Snow's series of causal-process observations led him to develop his natural experiment, in an approach very much in the spirit of multi-method research: both qualitative and quantitative evidence are leveraged in complementary ways at different stages of a research program. However, validation of a hypothesis through a subsequent confirmatory natural experiment does not necessarily validate a hypothesis-generating causal-process observation *qua* causal-process observation. Even with Snow, there may be a tendency for "post hoc, ergo propter hoc" thinking. That is, we may tend to see the process tracing in Snow's (1855) cholera study as powerful – and his supporting arguments for yellow fever as thin – because subsequent studies showed that he was right about cholera and wrong about yellow fever. Yet it does not follow that Snow's causal-process observations in the case of cholera were necessarily more powerful than in the case of yellow fever.

Of course, Snow's natural experiment (and subsequent studies, including later work by microbiologists) did ultimately confirm his conjecture that cholera is waterborne; other research helped to pin down the distinctive causes of transmission of plague,

[22] For example, the miasma theory may have failed a "hoop test," due to Snow's observation that sailors did not contract cholera until disembarking; yet, this is at best a "straw-in-the-wind test" for his theory that cholera is a waste- and water-borne disease.

yellow fever, and malaria.[23] The quality of the evidence that led to Snow's initial conjecture about cholera, and to his misleading hypotheses about yellow fever, may not appear to matter much in retrospect. This conclusion would be short sighted, however. In many settings, perhaps especially in the social sciences, things are not so clear-cut: the conclusions drawn from subsequent testing may not be so sharp. It therefore does matter to try to validate particular CPOs – i.e., not only to establish the truth of some general hypothesis generated through process tracing, but also to confirm the evidentiary value of a causal-process observation itself.

However, Snow's study also suggests that it may be quite tricky to evaluate the independent persuasiveness of a given CPO *without* subsequent or complementary confirmatory evidence. Indeed, as the previous section showed, knowledge of context or process often plays a critical role in validating or invalidating research designs – including the very designs expected to provide critical tests of hypotheses. In Snow's natural experiment, qualitative knowledge of the nature of water markets in nineteenth century London played a critical role in making plausible the "as-if" random assignment of households to sources of water supply. If such claims about water markets are mistaken, then the case for the natural experiment itself is substantially weakened. Thus, even in Snow's study, the quality of CPOs matters for interpreting the credibility of confirmatory tests: those pieces of evidence are used to validate the natural experiment – even as the results of the natural experiment seem to validate other CPOs.

In sum, the evidentiary value of a given causal-process observation may depend on the existence or non-existence of certain *other* pieces of evidence, as well as analysis of data sets from strong designs. As Collier (2011: 824–825) puts it, "Identifying evidence that can be interpreted as diagnostic depends centrally on prior knowledge … The decision to treat a given piece of evidence as the basis [for a process-tracing test] can depend on the researcher's prior knowledge, the assumptions that underlie the study, and the specific formulation of the hypothesis."[24] In particular, the quality of any piece of evidentiary support must be evaluated in the context of existing background knowledge and theory, and especially, in light of other diagnostic pieces of evidence. Thus, the evidentiary weight of a given CPO clearly depends not only on its own veracity but also on the non-existence of other CPOs that might provide countervailing inferences.

[23] See e.g. Freedman (2010) for references and a review of this research.

[24] Zaks (2011), for instance, assesses the relationship of process tracing evidence to alternative theories and discusses how to use process tracing to adjudicate between them.

Situating pieces of diagnostic evidence within a broader field of other causal-process observations is therefore critical for buttressing the claim that process tracing has produced genuinely dispositive evidence in favor of a particular assumption or hypothesis.

This point raises a second major challenge, however, regarding how to elicit potentially disconfirming as well as confirmatory evidence – an important challenge in natural experiments, where much hinges on detailed information about the process of treatment assignment. In the Argentina land-titling study, for example, we are told that Catholic Church organizers did not know the state would expropriate land and allocate titles to squatters; and even if they had, they could not have predicted which particular parcels would have been subject to court challenges. Thus, evidence that Church organizers *did* have reason to suspect that land would be expropriated from original owners – or that they had a basis for predicting which absent landowners would challenge expropriation in court and which would not – could undermine the plausibility that land titles are assigned as-if at random. In Snow's natural experiment, evidence that residents did in fact self-select into source of water supply – e.g., by changing suppliers after the move of Lambeth's pipe – could similarly invalidate the claim of as-if random.

I do not have reason to believe that such countervailing evidence exists in these particular examples. On the other hand, I do not have strong reason *not* to believe countervailing evidence exists: in the main, I know only what I am told by the authors of the studies. This again brings into focus an Achilles Heel of natural experiments with as-if random assignment, relative to studies in which treatment assignment is truly randomized (Dunning 2008a). "Absence of evidence" is not "evidence of absence," a point emphasized in both this book's introduction (Bennett and Checkel, this volume, XX–XX) and Appendix. To assess the evidentiary value of qualitative evidence offered by researchers in defense of a particular claim, the research community would benefit from access to a range of *other* potential causal-process observations that might have been offered but perhaps were not.

How can individual researchers, and the research community, best meet this second challenge of validation? In the first place, it seems to be a responsibility of the original researchers not only to look for evidence that supports an assumption such as "as-if" random but also evidence that would *undercut* it – again, similar to the admonition in Chapter 1 that scholars should "cast the net widely" yet "be equally tough on alternative explanations" (Bennett and Checkel, this volume, XX–XX). Moreover – and especially if researchers fail to find such evidence – they should report how and where they looked. In other words, and as emphasized in both Bennett and Checkel's opening (this volume, XX) and concluding chapters (this volume, Chapter 10), transparency in research procedures is all important. In our two running examples, for instance, researchers could report the types and number of people they interviewed and the other sources they consulted, and then state that they found no evidence that people could or did switch water companies on the basis of water quality (Snow) or chose plots in anticipation of which landowners would challenge expropriation in the courts (Argentina land-titling study).[25]

Yet, researchers may still face challenges in communicating in a credible and transparent way that they have adhered to Bennett and Checkel's best practices. With many natural experiments, the requirements in terms of substantive knowledge and mastery of the details of the process of treatment assignment are demanding; such research designs often involve intensive fieldwork. Indeed, this is a major virtue of the approach, because it brings researchers into close engagement with the research context, thereby "extracting ideas at close range" (Collier 1999). At the same time, this very level of substantive knowledge implies that many other scholars may not have first-hand knowledge – for instance, of the incentives, information, and capacities of key actors involved in assigning a given treatment – that is required for evaluating the plausibility of as-if random. Moreover, researchers themselves can only be held accountable for the evidence they do uncover; but again, absence of evidence does not always constitute evidence of absence. In sum, it is not easy to rule out completely the possibility that qualitative evidence not uncovered or offered by researchers might undermine their case for the research design.

Thus, it may often be quite tricky to assess whether Bennett and Checkel's criteria for good process tracing have been applied. The development of general best practices is surely a helpful step forward. Yet, it is just as critical for researchers to be able to communicate credibly that they have adhered to these standards – for instance, that they have "cast the net widely for alternative explanations" or have been "equally tough on the alternative explanations." At its core, the challenge is to verify that researchers have

[25] Snow implicitly does something similar when he notes that water pipes were laid down in the years when companies "were in active competition" (1855: 75). Galiani and Schargrodsky point to such interviews, though do not always specifically report to whom they spoke. For more detailed descriptions, see Snow 1855; Freedman 1999, 2009, 2010; or Dunning 2008a, 2012.

indeed successfully sought both confirming and potentially disconfirming causal-process observations – so that disconfirming evidence will appear, if it exists.

The utility of adversarialism?

One potential solution to this challenge of validation might be found in specific research *procedures*. Thus, researchers could catalogue disparate pieces of qualitative evidence, so that communities of scholars working in particular substantive areas might more readily examine them. Scholars can then subject more easily to critical scrutiny the truth of crucial claims, such as assertions of as-if random that rest on somewhat esoteric details about treatment assignment processes. One might therefore appeal to the utility of competition between scholars with different vested interests in upholding or subverting the veracity of a given claim. Such "organized skepticism" (Merton 1973; see MacCoun 1998) may provide the most feasible way of confronting the problem that absence of evidence is not evidence of absence. Indeed, competition between scholars may boost the chance that potentially disconfirming bits of qualitative evidence are fruitfully brought to light. Much as in a court of law (at least in the non-inquisitorial tradition), where competing advocates adduce evidence in favor of or against a particular interpretation or causal claim, scholars with different theoretical commitments might seek to uncover evidence that supports or undermines a particular hypothesis.[26]

This image of seeking truth through scholarly rivalry is familiar. In the case of multi-method, design-based research, however, the specific focus is novel. For instance, scholars claiming to use a natural experiment might point to the aspects of an assignment process that support as-if random. Other scholars might seek to bring forward evidence that makes this assertion less plausible.

To be sure, the analogy to courts of law is only partially appropriate for scholarly research. Legal adversarialism pre-commits dueling attorneys to providing whatever evidence is most consistent with the position they have been assigned to attack or defend, which one would hope (!) does not characterize scholars even with the very strongest theoretical commitments. And unlike, say, some civil courts – in which an ostensibly disinterested judge adjudicates between competing truth claims – in the academic realm there is no neutral third-party arbiter of justice. (The analogy to criminal trial by a jury of one's peers might be somewhat more apt). Still, the more-than-passing

[26] MacCoun (1998) contrasts the adversarial and inquisitorial traditions.

resemblance of legal adversarialism to what some scholars do at least some of the time suggests this analogy might be fruitfully explored.

Thus, one might ask: how successfully has organized skepticism interrogated the validity of research designs or modeling assumptions in multi-method research – or put the validity of supporting causal-process observations themselves under dispute? At the most general level, one can find plentiful contemporary examples of disputes about the quality of evidence. Many of these seem to focus on questions of conceptualization and especially measurement. Examples include Albouy's (2012) critique of the settler mortality data used by Acemoglu *et al.* (2001); Kreutzer's (2010) criticism of Cusack *et al.* (2007); or Rothstein's (2007) appraisal of Hoxby's (2000) coding decisions. Such assessments of data quality can certainly require qualitative knowledge of context and process. They may even sometimes involve causal-process observations – perhaps of the type Mahoney (2010) calls "independent-variable CPOs," where the main issue involves verifying the presence or absence of a cause. However, they do not typically involve the research design and causal modeling assumptions on which I have focused in this chapter.

In contrast, critiques of modeling assumptions do abound in the literature on natural experiments; yet these often take the form of assessing observable *quantitative* implications of these modeling assumptions. For example, the assertion of as-if random implies that variables not affected by the notional treatment should be about equally distributed across treatment and control groups – just as they would be, in expectation, if treatment were assigned through a coin flip.[27] Thus, Caughey and Sekhon (2011), critiquing Lee (2008), show that winners of close elections in the US House are *not* like losers of those elections on various pre-treatment covariates, especially partisanship (Democratic incumbents tend to win the close races more than Republican challengers).

Sovey and Green (2011) critique the claim of as-if random in Miguel *et al.*'s (2004) study of the effect of economic growth on the probability of civil conflict in Africa. Here, rainfall growth is used as an instrumental variable for economic growth, implying an assumption that rainfall growth is assigned as-if at random; yet, using Miguel *et al.*'s replication data, Sovey and Green suggest that "factors such as population, mountainous terrain, and lagged GDP significantly predict rainfall growth or lagged rainfall growth, although these relationships are not particularly strong and the predictors as a group tend

[27] Formal statistical tests may then be used to assess whether any observed imbalances are consistent with chance variation.

to fall short of joint significance" (2011: 197). Thus, here we see examples of efforts to assess the quantitative implications of as-if random, using statistical tests.

Such examples lean rather less heavily on causal-process observations, however – for example, on whether key actors have the information, incentives, and capacity to subvert random assignment – to validate design and modeling assumptions. To be sure, one can find good examples of the use of qualitative methods by researchers to substantiate their own claims of as-if random. Iyer (2010), in an article in the *Review of Economics and Statistics*, uses extensive documentary and archival evidence on the Doctrine of Lapse during the reign of Governor Dalhousie, a central component of her effort to find as-if random variation in the presence of princely states in colonial India (as opposed to direct colonial rule by the British). Posner's (2004) article on interethnic relations in two African countries also makes very effective use of qualitative evidence, though it does so mostly to explore mechanisms more than to support the claim of as-if random placement of a colonial border between modern-day Zambia and Malawi. In the Argentine land-titling study and other settings, qualitative evidence also clearly plays a central role.

Yet, a review of the literature on natural experiments tends to suggest the *potential* utility of causal-process observations for interrogating design and modeling assumptions in multi-method work.[28] Critiques of as-if random have tended not to draw extensively on qualitative evidence – perhaps precisely because of the extensive case knowledge and detailed information required to do so. It therefore seems there is much opportunity for greater use of qualitative methods in natural-experimental research to probe design and modeling assumptions, yet one of the difficulties concerns how best to elicit and use varied qualitative information on processes of treatment assignment.

What sorts of research procedures might promote better use of CPOs, and particularly better validation of their evidentiary value? One possibility is to promote better cataloguing of qualitative data from fieldwork interviews, archival documents, and so forth. The new Qualitative Data Repository at Syracuse University is one effort to provide a platform for public posting of qualitative evidence.[29] There, researchers will be

[28] Caughey and Sekhon (2011) also scour newspaper accounts from a random sample of close elections for qualitative evidence that could explain why Democrats win close races; however, here they are interested in evidence on mechanisms, so the qualitative evidence itself is not as central to evaluating violations of as-if random.

[29] The data repository has been established through a grant from the US National Science Foundation, with Colin Elman and Diana Kapiszewski as Principal Investigators.

able to post field notes, transcripts of interviews, archival documents, and the like; the aim, *inter alia*, is to boost transparency and perhaps replicability in qualitative research.

There may be various forms of posting and accessing qualitative information. For instance, "active citation" allows references to archival documents or other sources in a publication or a research report to be hyperlinked to a partial or full virtual copy of the document (Moravcsik 2010). Other cataloguing procedures involve posting of entire interviews or transcripts – qualitative data *qua* data. Both sorts of information could be useful to researchers, within the adversarial tradition in which one scholar contests another's CPOs. For example, the suspicion that extracts from interviews may be cherry-picked to support particular propositions or interpretations might be mollified (or exacerbated) by review of an entire interview transcript.

Better use of such transparent research procedures would likely make for better process tracing by individual researchers in several ways. It would encourage them to think through whether they have "cast the net widely for alternative explanations" – in particular, for evidence that would undermine as well as support key assumptions – before the publication of their research. And it might also help researchers as well as their critics assess whether they are being "equally tough on the alternative explanations." And it may facilitate consideration of "the potential biases of evidentiary sources" – all best practices advanced by Bennett and Checkel in Chapter 1.

Consider the following concrete advantages of posting or registering interview transcripts or other qualitative materials.

- A careful reading of interview transcripts in which key issues of information, incentives, and capacity are broached might make claims of "evidence of absence" more compelling. For instance, the assertions in the Argentina land-titling study that Catholic Church organizers did not know abandoned land was owned by the state, or that the legal owners of expropriated plots who challenged their taking in court did so for idiosyncratic reasons unrelated to the characteristics of squatters, may be made more credible by perusing transcripts of interviews with key actors.
- In contrast, the non-appearance of key interview questions, parenthetical remarks by informants, or other kinds of evidence might weaken the credibility of such claims. For example, we might expect researchers to probe informants' incentives, information, and capacities as they relate to the assertion of as-if random. The expectation of preparing qualitative materials for public posting might thus make individual researchers more self-conscious about the use of such tools.

- Individual researchers might be more prone to use various cognitive tricks to avoid confirmation bias – for instance, by assuming that they are wrong in their conclusions, or that their design assumptions such as as-if random are incorrect – and then asking how the evidence they have catalogued might support such alternative interpretations.
- Such documentation may also facilitate direct appeals to the expertise of communities of scholars. For instance, individual researchers might run key portions of their texts or even primary materials by area experts, historians, or key actors and informants who may be in a position to judge whether the scholars have misread key evidence bearing on issues such as as-if random.

To be sure, such documentation will not fully solve the problem of "missing CPOs" – that is, the problem that absence of evidence may not constitute evidence of absence. However, more complete recording of qualitative evidence – as laborious as that can be to provide – would surely improve on the current state of affairs. Researchers suspicious of an assertion such as as-if random would have a place to start looking for nuggets of information on context, process, or mechanism that would help to subvert such claims. And individual researchers adhering to such a transparent protocol could stake a more credible claim to have followed Bennett and Checkel's ten criteria for good process tracing.

Of course, the provision of more extensive documentation of qualitative evidence is probably only part of the solution to the challenges of validation. Like lawyers and judges, researchers have various incentives. Using CPOs culled from such documentation to probe the plausibility of as-if random involves substantial costs in time and effort, a point also recognized by Bennett and Checkel when they counsel scholars not "to give up" when confronting their ten best practices (this volume, XX–XX). Moreover, the intellectual reward may be uncertain and the professional returns meager – especially since, for better or worse, professional attention and credit seems likely to go to the discoverer of the design and less likely to accrue to the eager critic.

Finally, the deep and specialized substantive knowledge that is often required to identify potentially falsifying CPOs may also limit the utility of peer review. And those with the basis to know whether the full record of CPOs supports a claim of as-if random, or a particular modeling assumption like non-interference, might well have other incentives to attack or undermine another researcher's use of CPOs. This could leave the outside observer on shaky ground to determine what is true.

These caveats notwithstanding, greater provision of supporting qualitative documentation does seem likely to aid efforts to improve process tracing and thus to turn it from metaphor to analytic tool. Like the posting of quantitative datasets and replication files for published articles (which is by no means a universal practice, but certainly one backed by emerging norms), this effort can lend readers a reasonable expectation that contrary CPOs (i.e., those that contradict a main claim or hypothesis) would stand a decent chance of coming to light. Of course, there can be many practical or ethical issues that arise in posting qualitative data (such as protecting subject confidentiality), so the feasibility of providing supporting documentation may vary by project, or by type of evidence within projects.[30]

For multi-method research in the design-based tradition, this is good news. The assumptions of strong research designs often have sharper testable implications than conventional quantitative methods. For example, as noted above, as-if random assignment suggests that comparison (treatment) groups should be balanced on covariates and that policy makers or the units themselves should not have the information, incentives, and capacity to select into treatment groups in a way that may be correlated with potential outcomes. Each of these implications can be tested through a range of quantitative and qualitative evidence.

Conclusion: improving process tracing in multi-method research

The importance of multi-method work – in particular, of leveraging both qualitative and quantitative tools for causal inference – is increasingly recognized. With strong research designs, quantitative analysis can provide social scientists with powerful tools for assessing causation. Yet analysis of data sets is rarely sufficient. To develop strong designs, validate causal models, and interpret effects, analysts typically require fragments of information that give crucial insights into causal processes of interest. Process tracing is a label for a set of techniques and methods designed to generate such insights. As such, it plays an important role in social-science research.

However, it is critical to assess the quality and probative value of particular instances of process tracing. The standards put forth by Bennett and Checkel in Chapter 1 are useful in this regard. Yet, it can be difficult for researchers to demonstrate – and for the scholarly community thus to certify – that they have indeed "cast the net widely for

[30] The Qualitative Data Repository is working actively with researchers on how to address such issues.

alternative explanations," been "equally tough on alternatives," "considered the potential biases of evidentiary sources," and so forth. One major difficulty is that the value of a particular causal-process observation a researcher presents must be set in the context of other information, including possibly disconfirming evidence. Such information can be hard to find – and unless it is elicited and presented by researchers, the research community cannot assess its relative import.

Overcoming these challenges, if only partially, may involve: (1) the adoption of more transparent cataloguing practices for qualitative data, for instance, the posting of transcribed interviews and archival documents, and the use of active citations; and (2) facilitation of scholarly contestation of process tracing claims, which will in turn be aided by transparent cataloguing. Thus, scholars could use comprehensive qualitative information – including the data provided by individual researchers under (1) – to interrogate and perhaps contest specific claims about the information, incentives, and capacities of key decision-makers. Cataloguing interview transcripts and other sources would allow critics to focus on questions *not* asked, or answers *not* reported – which might allow some assessment of evidence of absence, as well as absence of evidence. To the extent that such information substantiates or undercuts researchers' design or modeling assumptions, it would be particularly useful for multi-method work, such as natural experiments. Thus, while standards for "good" process tracing may be difficult to implement, research *procedures* that boost transparency in qualitative research may help substantially to close this implementation gap. The result will be to advance this volume's central goal – i.e., improving process tracing. For individual researchers, the adoption of such procedures could facilitate credible and transparent claims about the probative value of process tracing evidence.

In this chapter, I have illustrated these points with respect to natural experiments, but similar arguments are likely applicable to other forms of multi-method research – such as those combining formal theoretical models or cross-national regressions with case studies. Of course, the specific inferential issues may vary in those contexts: with more complicated models and less plausible design assumptions, the challenges of validation may be even greater. But in principle, the major challenges – of (i) understanding *selection processes* that assign cases to alternative causal conditions or categories of a treatment variable; and (ii) validating key modeling assumptions – also apply to these forms of research. Thus, "treatment-assignment" and "model-validation" causal-process observations apply in these other settings as well.

In the end, however, there is unlikely to be a silver bullet. Social science is difficult, and causal inference is especially challenging. Discomfortingly, Green and Gerber (2012) may have it right in their discussion of experimental designs, when they argue that "the metaphor of a gold standard is correct … if we mean extracting nuggets of gold from legions of sediment."[31] In a similar way, Freedman avers, "Scientific inquiry is a long and tortuous process, with many false starts and blind alleys. Combining qualitative insights and quantitative analysis – and a healthy dose of skepticism – may provide the most secure results" (2010) For mixed-method research, good process tracing can thus play a central role, together with many other techniques, but there are likely to be many failures along with the successes.

[31] N.B. - Quotation needs double-checking.