

What Can Causal Process Tracing Offer to Policy Studies? A Review of the Literature

Adrian Kay and Phillip Baker

Causal process tracing (CPT) has emerged as an important method of causal inference in qualitative social science research, most notably in case study research designs. There is now a considerable literature on the aims, philosophical groundings, and methods of process tracing. This paper reviews the CPT literature to assess what new directions it may suggest for policy studies. The first part of the paper sets out the methodological advantages CPT offers in building and testing theories of policy change, most notably in supporting a theoretical pluralism to address the problem of complexity in policy studies. Building on recent scholarship across the social sciences, the second part examines step by step the recently minted “best practice” for undertaking CPT in policy studies. This part includes discussion of the possible pitfalls of CPT as a method; common errors involved in its use are set out and minimization strategies offered. In particular, while acknowledging the usefulness of Bayesian tests for causality as heuristic devices, we emphasize the limitations of applying such tests in practice. Possible correctives are suggested. The final part of the paper speculates more generally on the potential of CPT to improve our investigation of patterns of policy change over time.

KEY WORDS: causal process tracing, policy studies, case study design

Introduction

Causal process tracing (CPT) has emerged as an important method of causal inference in research that employs some form of case study design (Collier, Brady, & Seawright, 2010; George & Bennett, 2005). Reflecting a surge of interest in qualitative research design more generally, significant recent advances have been made in refining the CPT method, both as a standalone tool and a complement to other research methods (Beach & Pedersen, 2013; Bennett & Checkel, 2012; Collier, 2011; George & Bennett, 2005; Hall, 2006).

This paper contends that the CPT method offers significant potential for addressing some of the main research questions which occupy and motivate the field of policy studies. CPT can be used for between-case analysis of different causal paths in small-N case study research, including long and complicated causal chains with perhaps disproportionately large or small effects as well as the contingencies involved in different outcomes from very similar combinations of contexts and

causal drivers. However, CPT's singular advantages for policy studies are manifest in the promise of a more robust method for understanding causality from within-case accounts of policy change while allowing for comparability between single case studies.

The first part of the paper sets out the argument that CPT offers several methodological advantages in building and testing theories of policy change over time, notably in supporting a theoretical pluralism that gives answers to the problem of complexity in policy studies. Briefly, this is the view that important policy problems cannot be solved or policy responses studied in isolation; instead, they require consideration of the dense web of relationships connecting states, companies, civil society organizations, and individuals as a policymaking system as well as analysis of their mutual influences. Building on recent scholarship across the social sciences, the second part examines step by step the recently minted "best practice" for undertaking CPT at it applies to policy studies. This part includes discussion of the possible pitfalls of CPT as a method; common errors involved in its use are set out and minimization strategies offered. We review briefly the use of Bayesian tests of causality and, while acknowledging their usefulness as heuristic devices, we reflect on their practicality in policy studies. Alternative standards of good causal explanation are put forward for future development, which are compatible with but not driven by the Bayesian logic dominant in the CPT literature. The final part of the paper speculates more generally on the potential of CPT to improve our investigation of patterns of policy change over time.

CPT: Methodological Attractions and Varieties

Why Use CPT in the Study of Public Policy?

There are several methodological grounds for the expanded use of CPT in policy studies. The first and perhaps most obvious is that CPT addresses directly a core problematic: the inherent complexity in the temporal analysis of open policymaking systems (Cairney, 2013). As a method, it can be used to identify and describe policy events, and to elaborate on the single or multiple paths by which they come about (Collier, 2011). CPT holds the promise of a rich account of "how" a complex political phenomenon like public policy emerges. In particular, it enjoys advantage in analyzing "processes that are path dependent or rooted in strategic interaction" where regression analysis or statistical models can offer only limited causal claims about the relationships between individual factors (Hall, 2006, pp. 29–30).

Complexity begets an occupational hazard for all policy scholars: access to valid, reliable, and useful data. Almost invariably in policy studies, we focus on the contemporary or near past, and key variables are often hidden from view, by official decree or political imperatives, and data are fragmented and not conveniently additive. CPT offers a "tool for drawing descriptive and causal inferences from diagnostic pieces of evidence—often understood as part of a temporal sequence of events or phenomena" (Collier, 2011, p. 824). Even in a research environment of poor,

fragmented, and incommensurable data, Checkel (2006, p. 365) argues that CPT can still provide a “how-we-come-to-know nuts and bolts for mechanism-based accounts of social change [and directs] one to trace the process in a very specific, theoretically informed way,” or as George and Bennett (2005, p. 206) assert, CPT offers a means to “identify the intervening causal process—the causal chain and causal mechanism—between an independent variable (or variables) and the outcome of the dependent variable.”

Another strong methodological feature of CPT is that it offers a strong complement to commonly used methods in policy studies including comparative case study and large-N designs (Gerring, 2007). So-called “hypothetico-deductive” study designs or those that rely for example on correlation, counterfactual, and experimental study designs can offer useful information on the inputs and outputs of a given causal relationship, but the black box of causality usually remains opaque and is labeled “intervening variable.” Thus, these designs are useful for determining what potentially causes policy change. We must turn to causal mechanisms to make the mechanistic properties of the black box visible—to answer questions concerning *how* causal variables cause policy change. Employed as a form of analysis of the “how” question to complement research designs for the “why,” concerns about nonexperimental and incommensurable data in CPT are less pronounced. The workhorse case study research designs of policy studies—within-case and small-N comparison—are given a particular support by CPT to adumbrate complex causal links that are hidden in large-N statistical designs that identify correlations between variables. Importantly, CPT can be used for idiographic research: At a methodological level, it prompts researchers to disaggregate the policy process into constituent elements rather than seek a general or grand theory of the policy process. In doing so, attention is concentrated on the unique, or at least highly specific, constellations of mechanisms that give rise to particular policy events at particular times in particular places. However, this attentiveness to context does not preclude causal explanation. CPT offers that the ambition to generalize mechanisms across different spatial and temporal contexts can be combined with acknowledging the uniqueness of particular policy processes. Mechanisms, even with multiple theoretical antecedents, are portable within and across cases and can be the basis for systematic theorizing about policy processes, particularly if some kind of typological theorizing is employed to work out how mechanisms may interact in any particular situation.

An additional strength of CPT is that it can be well suited to the theoretical pluralism common in frameworks employed in policy studies research. Schlager (2007) uses the Ostrom notion of a framework as providing for multiple theories with a general class of variables and mechanisms necessary for explaining a particular phenomenon under investigation. The theoretical pluralism of a framework approach to policy studies allows the investigator to hypothesize multiple independent variables, causal mechanisms, and dependent variables to elaborate on the relationships between them and to provide an understanding as to how they might be operationalized as empirical observations when present in a particular case (Collier, 2011). Frameworks may be revised as theories improve in their explanatory

range and power. Schlager (2007) goes on to set out theories that are tested and revised through the development of models, where a model represents a particular situation, more precise in its assumptions and narrow in scope than the theory from which it derives.

A central claim of this paper is that CPT is consistent with, as well as providing justification for, theoretical pluralism in policy studies. Sabatier (2007) exhorts policy scholars to apply several different theoretical perspectives because this contributes to a clarification of assumptions across frameworks and support the accumulation of evidence over multiple studies. The application of multiple theoretical perspectives should gradually lead to clarifying conditions under which one perspective is more useful than another. Far from being seen as a weakness of the field, theoretical pluralism can offer more powerful and “theoretically triangulated” causal inference. This is one clear answer to the dilemma posed by Cairney (2013) in the *Policy Studies Journal*: How do we combine the insights of multiple theories in public policy studies? From a CPT perspective, the objective should be to not only identify whether a hypothesized variable is present or absent but to specify from among a set of potential variables those that best explain causality at each link in a causal sequence. This also allows the analyst to test theories comparatively (Rohlfing, 2014). Conversely, the use of a single theory can mean ignoring alternative and potentially equally valid explanations, and therefore comprise an important form of inferential error.

Theoretical pluralism allows the analyst to address the inherent complexity of political phenomena. Understanding change in terms of only one or two causal variables will, in many cases, likely underestimate causal complexity. There may not be just a few but many variables, and the relationships between them may be independent–dependent but also potentially interdependent as well as temporally dynamic. As Bennett and Elman (2006a, p. 251) elaborate, multiple phenomena have complex causes “including tipping points, high-order interaction effects, strategic interaction, two-directional causality or feedback loops, [as well as] equifinality . . . and multifinality.” Indeed, policy as a variable traced over time often exhibits these characteristics. For example, Head’s (2008) categorization of policy problems is one of many available to policy scholars to different types of policy problems that may vary in their degree of complexity.

The theoretical pluralism that CPT supports can help to address such complexity and open up two particularly promising lines of future inquiry to policy studies. First, CPT can be combined with typological theory (Bennett & Elman, 2006b, p. 466; Collier, LaPorte, & Seawright, 2012; George & Bennett, 2005; Steinberg, 2007) to extend analysis beyond the variables and causal mechanisms predicted by an individual theory to a consideration of how such mechanisms interact, their “higher order interaction effects.” As George and Bennett (2005, p. 147) put it, “typological theories provide a way to model complex interactions or causal mechanisms by including recurrent combinations of hypothesized mechanisms as distinct types of configurations.” They also identify “how and under what conditions [such configurations] produce effects on specified dependent variables” (George & Bennett, 2005, p. 235).

Another future avenue for theoretically guided CPT is counterfactual reasoning (Collier, 2011; Fearon, 1991; Levy, 2008). Counterfactuals posit alternative possible causal processes that could have occurred, that are counter to the actual causal process established factually. Specifying the counterfactual to a theorized hypothesis, or in other words the “null hypothesis,” allows for more powerful theorization and therefore CPT. As Levy (2008, p. 631) highlights, “A theory that specifies the consequences of both x and not x tells us more about the empirical world than a theory that specifies only the consequences of x .”

Counterfactuals are implicit to any idiographic analysis which has the ambition to establish claims of equifinality and multifinality. Hypothesized causal variables can be conceptualized ideographically in this regard as binary—as present (1, observation) or absent (0, counter observation) at each point in a causal sequence. Hypothesized causal mechanisms that convey the causal force from the causal to the outcome variable determine the direction in which the tree branches at each juncture in a causal sequence. Cumulatively, this accounts for all possible causal sequences and outcomes as a causal process tree.

CPT in Case Study Design

CPT may be applied to within-case analysis, in small- N case-comparative designs, or when N equals 1. As is established in policy studies, the former design avoids the case selection bias that can result from selecting on the dependent variable and from the “intuitive regression” of small- N comparisons (Bennett & Elman, 2006b, p. 461; George & Bennett, 2005). Of course, a major limitation of CPT applied to a within-case research design is that any claims of cause–effect relationships will be limited to that case only—that a causal mechanism or combination of mechanisms was either absent or present and that it functioned as hypothesized.

In policy studies, sometimes we know about the population under study (e.g., policy sector) from which the case is drawn, in which case comparative case study or regressions say something useful; but sometimes, we simply do not know anything. These are research problems in policy studies where regressions or comparative cases are not useful. This is where a single, within-case CPT study has benefits by addressing the external validity by reference to the importance of a common theoretical framework. Single case studies can be valid because they are comparable with other studies using the common theoretical framework. As we argue further below, generating alternative explanations using CPT for a particular case outcome *does not* require explicitly comparative work.

Of course, small- N designs which employ some form of matching and contrasting of cases have produced much celebrated policy studies research, but their “contribution to causal inference urgently needs to be supplemented by within-case analysis” (Collier, 2011, p. 824). As Bennett and Elman (2006b) point out, with CPT, “causation is not established through small- n comparison alone . . . but through uncovering traces of a hypothesized causal mechanism within the context of a historical case or cases.” It can help the analyst to determine whether a correlation

between two variables is causal and not a spurious correlation resulting from the presence of an antecedent variable (Mahoney, 2000, p. 412).

If it is not possible to make comparative case comparisons or when comparative cases are imperfectly matched so controlling is not feasible—the hardy and perennial dilemma of comparative public policy scholarship—CPT offers a way forward by supporting causal inference from within-case research designs by supporting robust theoretically plural explanations of policy change.

Varieties in CPT

Although we emphasize theoretically guided CPT in this paper, it is important to acknowledge the three variants of CPT identified in the literature¹ (Beach & Pedersen, 2013; George & Bennett, 2005; Hall, 2006). First is case-centric CPT for “explaining outcomes in a particular case,” where the key purpose is to answer the question “what mechanistic explanation accounts for [the] outcome?” This variant is used by researchers who assume a case to be very complex, multifactorial, and context specific, not easily generalized beyond the case itself (Beach & Pedersen, 2013).

Second are theory-centric CPT variants, comprised of theory building and theory testing. Theory testing is deductive—the purpose is to test existing theory and the causal mechanisms that they hypothesize by asking is the “causal mechanism present and does it function as hypothesized?” It is employed in two situations: first, when there are existing deductions concerning a causal mechanism linking X and Y in a particular type of case, or second, when deductions from existing theorization concerning a mechanism can be made easily (Beach & Pedersen, 2013).

In contrast, theory building is inductive: The aim is to develop new theory by asking “what is [the] causal mechanism between X and Y” based on the empirical evidence of a particular type of case (Beach & Pedersen, 2013, p. 13). It is employed for two purposes: first, when there is evidence of an existing correlation between X and Y but there is uncertainty about what causal mechanism links them; second, when an outcome (Y) is known but there is uncertainty about what causes it (X) (Beach & Pedersen, 2013, p. 14). All of these variants can be used for making causal inferences regarding policy change. These three variants, their synonyms, and uses are described in Table 1.

Mechanistic Building Blocks for Theories of the Process

CPT is an important method for building and testing theories of the policy process using mechanisms; it can be used to triangulate (theoretically as well as empirically) hypothesized causal mechanisms linking causes and their effects. CPT ideally provides an explanation of each significant step in a sequence of policy development by reference to a theory. Because the method disaggregates policy

Table 1. Causal Process Tracing Variants, Descriptions, and When to Use Them

Variant (Synonyms)	Description/Key Questions	When to Use
Case-centric Detailed narrative Historically specific	A detailed narrative that explains how a particular outcome or set of events came about What mechanistic explanation accounts for the outcome?	For explaining outcomes in a particular case To develop a data set for making tests for causal inference
Theory testing Theory orientated Analytical causal explanation	To test deductively derived theories and the causal mechanisms that they hypothesize Is the causal mechanism present or absent? If present, does it function as hypothesized?	When a theory is deduced from existing literature and the intention is to test whether a causal mechanism hypothesized by that theory is present or absent in a given case
Theory building Multivariate Hypothesizing and generalizing	To inductively generate new theory by identifying casual mechanisms from the empirical evidence of a particular case What is the causal mechanism between X and Y?	When the intention is to use empirical evidence to develop a theoretical explanation and to generalize to causal mechanisms of a particular case

Sources: Beach and Pedersen (2013), George and Bennett (2005), and Hall (2006).⁸

processes, it is particularly sensitive to the possibility of multiple alternative pathways to the same outcome (equifinality), as well as multiple outcomes resulting from the same causal mechanisms and conditions under which they operate (multifinality).

There is significant variation in the definition of a “causal mechanism” (e.g., Mahoney, 2012, pp. 579–80; Mayntz, 2004; Hedström & Ylikoski, 2010, p. 51). George and Bennett’s widely used and basic definition is helpful; the mechanisms are as follows: ultimately unobservable physical, social, or psychological processes through which agents with causal capacities operate, but only in specific contexts or conditions, to transfer energy, information, or matter to other entities. In doing so, the causal agent changes the affected entities’ characteristics, capacities, or propensities in ways that persist until subsequent causal mechanisms act upon them (George & Bennett, 2005, p. 137).

In CPT, causal mechanisms can be studied empirically through detailed historical analysis, and because history does not repeat itself, the outputs resulting from the mechanism are invariably contingent. CPT thereby takes the form: *X* caused *Y* through a mechanistic process of *A*, *B*, *C* in case *Z*. That is not to say that mechanisms are not portable—the same mechanism can operate the same way under different contexts. However, the attributes of that context will determine what type of outcome the mechanism generates. It is therefore important to recognize that “mechanisms alone cannot cause outcomes. Rather, causation resides in the interaction between the mechanism and the context within which it operates” (Falleti & Lynch, 2009). This introduces the concept of multifinality: The same mechanism can generate differential policy outcomes dependent on the context in which it operates.

Table 2. Primary Variables Commonly Referred to When Using the CPT Method

Variable (Synonyms)	Independent Variable	Intervening Variable ^a	Dependent Variable
	Causal Variable	Causal Mechanism and Causal Process	Outcome Variable
Position in causal sequence	First	Second, third, Xn-1	Xnth
Description	An initial phenomenon that triggers a causal process that generates an effect on the outcome variable	Processes through which agents with causal capacities operate to transmit a causal force between an independent and dependent variable	The outcome of interest; the final outcome in a casual process

^aThe use of the term intervening variable is problematic as we elaborate in this table.
Source: Adapted from Bennett and Checkel (2012).

As Bennett and Checkel (2012, p. 9) note, however, the Table 2 conceptualization of variables assumes that the causal mechanism wholly transmits an effect from the independent to the dependent variable in isolation. While this may be true in rare instances, given the complexity of policy phenomena, this conceptualization is overly simplistic—there will often be additional variables that generate important effects and variables interact; hypothesized causal variables are not necessarily mutually exclusive. To refine the use of causal mechanisms, it is useful to think in terms of antecedent, exogenous, complementary, and unobserved variables (Table 3). Together, these constitute important parameters that shape, enable, or constrain primary variables. In policy studies, these are often described implicitly as the “context” or “background” to a particular case in which the causal variable, mechanism, and outcome variable are situated. We return to the importance of background variables later in the discussion on hypothesis testing.

What to Trace? A Methodological Characterization of the Policy Variable

The nature of the dependent variable problem in policy studies is widely acknowledged (Howlett & Cashore, 2009), and a resolutely intractable conceptual discussion of the nature of public policy, both as an independent or dependent variable, continues to limit the quest for both a valid and reliable measure of policy change. Policy as a variable displays various spatial, temporal, and complexity characteristics that make it difficult to trace causally.

In the spatial dimension, policy change can occur across a nested hierarchy of layers, levels, or orders of abstraction. Mechanisms underpinning policy change can thereby operate at the micro (individual behavior), meso (the actions of policy communities or networks), and macro (institutional or social systems that structure political interaction) levels. All three levels can be important in determining or constituting a given policy process. This presents manifold methodological dilemmas: How micro does the causal process tracer go? Does CPT require a commitment

Table 3. Other Variables

Variable (Synonyms)	Antecedent Variables	Exogenous Variables Omitted Variables	Complementary Variables	Unobserved Variables Confounding Variables
Position in causal sequence Description	Prior to Variables that diminish or enhance the effects of the causal variable, thus preceding the causal sequence under analysis	First: Xn-1 Known to the analyst but excluded due to theorized casual insignificance or significant complexity. They can be included (made endogenous) at later stages if necessary.	First: Xn-1 Diminish or enhance the effects of the main variables but do not interact directly with them	Unknown Unknown to the analyst but interact with primary variables

Note: Adapted from Bennett and Checkel (2012).

to methodological individualism? What about internal psychological mechanisms such as beliefs and intentions?

Without disputing that these are central and difficult questions for policy scholarship, we argue that they do not directly affect the methodology of inference in CPT from the data to explanatory claims. Although all policy processes necessitate individual agency, policy change can be studied at meso or macro levels without reduction to decision making by individuals or further still to internal, psychic mechanisms. We agree with Falleti and Lynch (2009, p. 7) who note that “micro-level mechanisms are no more fundamental than macro-level ones.” They further elaborate on how theories of policy change invoke mechanisms across all three levels of abstraction. Furthermore, the level(s) at which the analyst traces policy change processes will depend on their research objectives or theoretical orientation (Checkel, 2006). These are scholarly judgments rather than something that is naturally or logically determined. For example, rational choice theorists specify individuals as the causal agent, network theorists specify policy communities, and institutional theorists specify broader rules, norms, and other social structures that govern the interaction of political actors. Some, like Sabatier’s Advocacy Coalition Framework, include all three layers of abstraction in order to generate inferences about meso-level policy processes (Sabatier, 2007; Sabatier & Jenkins-Smith, 1993).

However, as CPT is generally applied in practice, the causal mechanisms underpinning policy change will most often derive from middle-range theories useful once the policy process has been disaggregated into constituent elements. The purpose of CPT is not to seek a general or grand theory of policymaking but rather concentrate on the contingent conjunctions of mechanisms that may vary across time and space (or at least occur in highly specific constellations) that give rise to particular events that may or may not be identical or comparable. The application of CPT should, therefore, account for micro and macro orders of change but emphasize change at the meso level.

CPT also invokes, necessarily, time as an independent variable (Bennett & Elman, 2006a, 2006b). The layers of policy change will differ in terms of their rate of change (tempo), for how long they endure (duration), whether they are accelerating or decelerating (acceleration), and when they are important (timing) (Grzymala-Busse, 2011). For example, in the established orthodoxy in the field—punctuated equilibrium theory of policy dynamics—policy change patterns consist of extended periods of incrementalism punctuated by rapid bursts of change (Howlett, 2009).

How to Do Robust CPT Analysis

This section sets out the step-by-step best practice of how to “do” CPT as a method for making causal inferences concerning policy. Drawing on leading work in CPT from the social sciences more broadly, we identify potential sources of error and outline strategies for error minimization at each step for CPT’s application in policy studies.

Step One: Theorizing Variables and Empirical Proxies

When theory testing with CPT, the analyst identifies a set of preexisting theories and the hypothesized causal mechanisms that they predict, or when employing theory-building CPT, the analyst makes deductions from the evidence particular to a case in order to develop hypotheses that explain it. These hypotheses can generate additional tests within the same case or additional cases. However, theory-building CPT is almost invariably guided by preexisting theory, reoccurring empirical regularities particular to a type of case, the analysts' previous work, or knowledge concerning the phenomena of interest. In both theory-building and theory-testing variants therefore, theorization in one form or another precedes hypothesizing variables and the empirical proxies we would expect to observe if these are present in a particular case.

Theory-centric CPT has several steps and associated potential sources of error (Beach & Pedersen, 2013; Bennett & Checkel, 2012; Checkel, 2006; Collier, 2011; Hall, 2006). First, the analyst uses existing theory to hypothesize a set of causal mechanisms that link in a causal chain, an independent variable with the dependent (outcome) variable. As Collier (2011, p. 825) iterates, "Careful, analytically informed specification of hypotheses is essential both in selecting and interpreting pieces of evidence, and in weighing them against one another." It is also used to elaborate on the processes by which hypothesized mechanisms operate (how the mechanism conveys a causal force between phenomena *X* and associated outcome *Y*) and the assumptions that underlie them (Beach & Pedersen, 2013; Hall, 2006). As George and Bennett (2005) elaborate, "In using theories to develop explanations of cases through process tracing, all the intervening steps in a case must be predicted by a hypothesis . . . , or else that hypothesis must be amended—perhaps trivially or perhaps fundamentally—to explain the case." As noted earlier, a theory is more powerful when it states the counterfactual—what is hypothesized to happen if the theory is null. Careful specification of the null hypothesis alongside each hypothesis is another step policy scholars using CPT should consider.

Importantly, errors can emerge when using theory-centric and in particular theoretically plural variants of CPT. This includes the "first-mover advantage" problem where an analyst may favor one particular theory and discount others simply out of familiarity or personal preference. This increases the risk of making the data conform to hypothesized causal mechanisms that fit a favored theory rather than testing the data against multiple theoretical lenses (Checkel, 2006; Hall, 2006; Sabatier, 2007). Because data collection or the process of generating a narrative data set (see step two) is theoretically guided, the observations made can favor some theories over others. As Hall, following Kuhn, points out, "the 'facts' against which a theory is tested are always generated, to some extent, by the theory itself" (Hall, 2006, p. 27). To avoid these forms of error, policy scholarship should specify explicitly the set of theories within the framework that they are employing and give equal weighting to each theory in study design and evidence collection.

Theories employed by the analyst are at this stage used prognostically: to operationalize the variables the theories hypothesize as empirical proxies. Proxies

are in other words the observations the analyst would expect to observe if a hypothesized variable were present/absent or causal/noncausal in a particular case (Beach & Pedersen, 2013; Checkel, 2006). Proxies should be generated² for the independent variable(s), intervening mechanism(s), and dependent variable(s) hypothesized from the set of theories employed (Collier, 2011). Importantly, as Hall notes, "Special attention should be devoted to deriving predictions that are consistent with one theory but inconsistent with its principal rivals so as to be able to discern which among a set of competing theories is more likely to be valid." Further, "theories should be formulated so as to yield predictions that can be shown to be false by available data and that are distinguishable from the predictions of rival theories" (Hall, 2006, p. 27). Theorizing and generating proxies to causal mechanisms are therefore important steps that inform the process of collecting diagnostic evidence.

Step Two: Collecting Diagnostic Evidence

Once hypotheses and proxies have been generated; a body of evidence (or empirical observations specific to the case) is assembled. There is a comprehensive literature on building rigor into qualitative data collection (Ezzy, 2002; George & Bennett, 2005; Miles & Huberman, 1994; Robson, 2002; Yin, 1993, 1994). For this paper, the purposes of data collection in CPT are considered.

What constitutes evidence when using CPT (Collier, 2011, p. 824)? As outlined earlier, both forms of theory-centric CPT are theoretically guided, relying on prior knowledge. The variables hypothesized by that knowledge and the empirical proxies those hypotheses predict are used to guide evidence collection. However, as Hall notes, "This is not simply a search for 'intervening' variables [i.e., proxies to causal mechanisms] . . . [but] to see if the multiple actions and statements of the actors at each stage of the causal process are consistent with the image of the world implied by the theory" (Hall, 2006, p. 28).

Collier and Brady define a single unit of diagnostic evidence as a causal process observation³ (CPO) or "an insight or piece of data that provides information about context, process, or mechanism, and that contributes distinctive leverage to causal inference" (Collier, Brady, & Seawright, 2004, p. 277). Mahoney (2010, p. 127) usefully builds on this definition, classifying CPOs specific to theory-testing CPT. He identifies an independent variable CPO as "data concerning the existence of a posited cause," a mechanism CPO as "data concerning posited intervening events and processes," and auxiliary outcome CPO as "data concerning posited auxiliary outcomes." These correspond to the independent, intervening, and dependent variables proposed by a theory respectively. An important point is that CPOs are not necessarily the actual manifestation of a variable but rather they are observable indicators or markers of its existence, what have been referred to as "fingerprints" (Beach & Pedersen, 2013, Chapter 2).

Types of Evidence and Assigning Diagnostic Weightings. However, CPOs pertaining to the same variable are not necessarily equal: Some types of CPOs will have more bearing on a hypothesis than others. As George and Bennett (2005, p. 99) highlight,

it is necessary to critically evaluate evidentiary sources and weight the value of data collected. A corollary question presents: how should policy scholars assign a “diagnostic weighting” to any given CPO?

A number of factors are important for policy studies. First is the frequency of corroborating instances of a CPO. Triangulating empirical observations is a common technique for minimizing (and measuring through variability of tests on three or more technical replicates) measurement error in the natural sciences. This holds true for making causal inferences using CPT, where the analyst should collect as large and as variable a data set as possible (Bennett & Checkel, 2012; Bennett & Elman, 2006a; Hall, 2006; King, Keohane, & Verba, 2001). If three CPOs are observed (e.g., statements made by the same type of key informant) that are relevant to the same variable, and two of those CPOs are consistent whereas the third is not, we would give more weighting to the corroborating CPOs. Bennett and Checkel (2012, p. 31) articulate this clearly: “with triangulation, a researcher cross-checks the causal inferences derived from his/her process tracing by drawing upon distinct data streams.”

Second is to consider the likelihood of a CPO occurring “if the alternative explanations are true” (as quoted in Mahoney, 2010, p. 128). As Bennett and Elman (2006b) put it, “our confidence in the suggested explanation will be increased if process tracing finds evidence of observable implications that are inconsistent with alternative explanations.” This is of course the process of abduction: We can infer that a given explanation is more likely to be true, from the presence of the CPO that it predicts, but other explanations do not. The specificity of a CPO is important here. As Bennett and Elman (2006b) note, “Some evidence will be probative for many alternative explanations, and some evidence will be germane for only one explanation. A process tracing account will be stronger if key non-substitutable links in the hypothesized process are supported by the evidence.”

The third factor concerns the spuriousness of evidential sources—the degree of bias inherent to the source (Bennett & Checkel, 2012; George & Bennett, 2005). The diagnostic weighting assigned to a piece of evidence should be determined against the nature of its source. These can be differentiated as primary or secondary. Primary evidence (primary CPOs) is those created during the time period under analysis and has not been subjected to interpretation in a manner specific to the research. Documents such as media articles, speeches, manuscripts, policy documents, journal articles, gray literature, and declassified archival materials are examples. The recording unit used to store primary evidence can include summaries of entire documents, thematic paragraphs, or sentences (Robson, 2002).

In policy studies, primary CPOs are subject to the instrumentality of the actors who produce them and the context in which they are produced. Policy scholars are well used to reflexive enquiry: Who, for whom, where, when, and why was the evidence created and how do these factors insinuate bias? The interests of the author of a primary CPO are more likely to be strategic rather than neutral, portraying explanations that align with their interests modified by those of their intended audience. Statements made to some public interest groups will differ to those made to private interest groups. Statements made in private may be less

guarded than those made in public. Previously classified archival material will have greater weighting than public accounts of the time, and so on (Bennett & Checkel, 2012, pp. 28–29, 32).

Secondary evidence (secondary CPOs) is interpretations of primary evidence, created after the time period under analysis (or after a particular time period within it). These can include accounts by historians, key informant interviews, magazine articles, historical manuscripts, criticisms, and commentaries. Unlike primary CPOs, secondary CPOs provide an interpretation of the case under analysis within a particular framework and theory. It is important to consider potential biases in secondary sources. History is subject to the biases of the historian, what diagnostic evidence they used to validate their arguments, and through what theoretical lenses (if any) they interpreted events. Historical scholars are at risk of their own political and analytical instrumentality by selecting evidence that validates their arguments, political views, or preferred theoretical framework. For these reasons, the policy scholars should look to include a diversity of secondary CPOs, in particular contending accounts of the same case. By triangulating secondary CPOs, the analyst can also conceptualize what gaps in the causal narrative may be spurious or missing, identify inconsistencies in contending accounts, and attempt to address these inconsistencies through the collection of relevant primary evidence. Together with reviewing primary evidence, this is what is commonly referred to as “soaking and poking,” providing the analyst with an assessment of the evidentiary value of historical materials⁴ (George & Bennett, 2005, pp. 99–105).

We add a third form of evidence for policy studies: counterfactual (counterfactual CPOs). Counterfactual CPOs are nonobservations. That is, they are hypothetical or fictional constructs that run counter to the CPOs established in primary and secondary evidence. They take the form “what *could have* happened but did not *in fact* happen?” Thus, counterfactual CPOs can be conceptualized as nonobservations. They can add greater weighting to primary and secondary forms of evidence through the following logic: If established empirically that *x* is a cause of *y*, would *y* have occurred in the absence of *x*?

It is however impossible to collect evidence consisting of “nonobservations.” Instead, counterfactuals can be posited from theory, and counterfactual evidence can be established through key informant interviewing. If we accept the proposition that counterfactual reasoning is innate to human decision making, then we diverge from the standard “billiard ball” perspective of causation. It is usually assumed that causality proceeds in a unidirectional manner from past to future. However, in policy studies, we must account for the agency of policy actors and how the anticipation of future events (as causal factor) shape decisions previously situated in the present.

Step Three: Hypotheses Testing

With a data set collected, the next step for the theory testing variety of CPT is to specify the means by which to verify the presence or absence of hypothesized causal

Table 4. CPT Tests Developed from Bayesian Logic

		Sufficient	
		No	Yes
Necessary	No	Straw in the wind (weakest) Passing: affirms relevance of hypothesis but does not confirm it Failing: Hypothesis is not eliminated but is slightly weakened Implications for rival hypotheses: Passing slightly weakens them Failing slightly strengthens them	Smoking gun (stronger) Passing: confirms hypothesis Failing: Hypothesis is not eliminated but is slightly weakened Implications for rival hypotheses: Passing substantially weakens them Failing somewhat strengthens them
	Yes	Hoop (stronger) Passing: affirms relevance of hypothesis but does not confirm it Failing: eliminates hypothesis Implications for rival hypotheses: Passing somewhat weakens them Failing somewhat strengthens them	Doubly decisive (strongest) Passing: confirms hypothesis and eliminates others Failing: eliminates hypothesis Implications for rival hypotheses: Passing: eliminates them Failing: substantially strengthens them

Note: Adapted from Collier (2011, p. 825) who adapted this typology from Bennett (2010, p. 210).

variables and to seek to establish empirically that they were causal. As just noted, the diagnostic weighting of a CPO on a given hypothesis will vary, and the same CPO may function as proof for multiple hypotheses (Bennett & Checkel, 2012). However, the approach of making causal inferences by way of Bayesian logic is a useful heuristic device in policy studies; it does not often provide a useful practical guide for scholars.

The literature has settled on four tests⁵ based on the degree of belief in a hypothesis or a Bayesian logic: A hypothesis is inductively confirmed if the probability of it being true is higher after the diagnostic evidence is known than its probability of being true prior to collecting the evidence. These are the straw-in-the-wind, hoop, smoking gun, and doubly decisive tests⁶ (see Table 4). These tests can be used to determine (i) whether *X* and *Y* actually took place (descriptive inference) and (ii) whether *X* was a cause of *Y* (causal inference) (Mahoney, 2012). Van Evera (1997, p. 31) initiated this line of inquiry by differentiating between these four tests based on the certainty and frequency of the evidence a given theory predicts; in other words, the more unexpected a given piece of evidence is, the greater its rarity in observation so “the greater its corroborative power” (George & Bennett, 2005, p. 219). Bennett developed these concepts into a two-dimensional typology based on whether a given piece of evidence pertaining to a theory is necessary or whether it is sufficient for passing a test (Bennett, 2010; Bennett & Checkel, 2012; Collier, 2011). A necessary cause is one that is required for an outcome but does not produce that outcome in isolation. A sufficient cause is one that when present will generate the cause, although the outcome can come about through other causes and is therefore not necessary (Mahoney, Kimball, & Koivu, 2009).

Table 4 reveals how these tests vary in terms of their inferential leverage. Strong tests are preferred to weak ones (Van Evera, 1997) as a means of investigating rival

hypotheses. Passing a doubly decisive test confirms a hypothesis and eliminates alternative hypotheses, and because it requires a CPO to be both necessary and sufficient, it therefore constitutes the strongest test. Rohlfing (2014) argues the test is even stronger than currently acknowledged in the literature due to a failure to distinguish between the theoretical and empirical uniqueness of a CPO; the former refers to situations where theories do not overlap in their predictions and the latter to situations where we can only test one theory's predictions on the basis of a specific case.

Second, a smoking gun is a strong sufficiency test (Zaks, 2011). Where evidence is sufficient but not necessary to support the presence of a causal factor or the operation of its mechanism, it can confirm a hypothesis, but the absence of such evidence does not eliminate that hypothesis. When there are multiple causes necessary for the same outcome, the rarest of those causes is most important empirically. The more empirically important causes are also those that are closer to the threshold of being a sufficient cause in isolation (Mahoney et al., 2009). Third, a hoop test is a strong necessity test (Zaks, 2011). The absence of necessary evidence can eliminate the hypothesis, but the presence of sufficient evidence can establish the importance of but does not confirm it. Finally, a straw-in-the-wind test is the weakest. It is applied when evidence may have some probative value to the hypothesis but is neither necessary nor sufficient to affirm or reject (Bennett, 2010; Collier, 2011; Mahoney, 2012). Zaks (2011) described this probative value as "leverage for" or "leverage against" the hypothesis.

Taking the strongest (doubly decisive) test, the absence of both necessary and sufficient evidence to support a hypothesis can eliminate it. This emphasizes the importance of the inclusion of the counterfactual (examined earlier in this paper) when specifying hypotheses and evidential proxies. Counterfactual reasoning (abduction) can also be applied to provide evidence for a given hypothesis. Although only a weak test affirms it, a hypothesis is more likely to be true if strong tests eliminate all alternative hypotheses. This resonates with the words of Sherlock Holmes dictum that when all other factors are eliminated, whatever remains—however improbable—is the truth (Doyle, 1930, p. 93, as quoted in Bennett & Elman, 2006a).

While in principle such tests can be applied with exact certitude, in practice, it may not be possible to conduct robust tests because of the possible spuriousness of evidence and the complexity of policy phenomena. A doubly decisive test may therefore become a smoking gun or hoop test at best, and these in turn in practice can become a straw-in-the-wind test if such doubt exists (Mahoney, 2012). This presents a concern for any application of CPT in policy studies: If one can never completely eliminate nor affirm a hypothesis, then how useful are these tests? Given that causal inference is the central feature of CPT, this problematic step raises questions about the practical value of CPT to policy scholarship.

Two answers to this important challenge are sketched. The first follows the claim by Rohlfing (2014) that competitive hypothesis testing is usually not possible for case studies; instead, we should test each argument or conjecture *comparatively* in terms of their relative explanatory power. The process of comparative hypothesis testing

helps with the problem of theories with overlapping predictions. In line with theoretical pluralism, it is useful to include multiple explanatory factors in a research design on the grounds that independent variables marking separate theories will often correlate to some extent, particularly when many of the variables in different theories of the policy process are nested within the same framework. If this is the case, then controlling for the presence of other variables is important to minimize the possibility of omitted variable bias.

Comparative hypothesis testing recognizes that policy processes are complex political phenomena in which a causal factor and mechanism are context dependent, and multiple causal variables and their mechanistic parts can interact with one another to generate important interaction and nonlinear effects. Additionally, a confluence of multiple “weak” factors might simultaneously affect the likelihood of policy change not captured by a single theory. Recognizing temporally distinct policy paths, the conjunctions of causes need to be elaborated (George & Bennett, 2005, p. 26).⁷

There are concepts available to understand such conjunctions. First, an INUS cause is “An *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result” (Mackie, 1965 p. 246, as quoted in Mahoney et al., 2009, p. 125). As Mahoney et al. (2009, p. 124) describe it, “The individual causal factors are neither necessary nor sufficient; rather, they are part of an overall combination that is sufficient for the outcome.” Second, a SUIN cause is a constitutive causal component of a necessary cause. Mahoney et al. (2009, p. 126) offer a more precise definition, “a *sufficient* but *unnecessary* part of a factor that is *insufficient* but *necessary* for an outcome.” These causes align with the idea of causality as transmitted from the causal to the outcome variable via the synergistic effects of the parts of a mechanism rather than through a single intervening variable.

The second answer to challenge for CPT to contribute usefully to policy studies is to question the ascendancy of Bayesianism. From a policy scholarship point of view, Bayesianism is only one—admittedly important—approach to causal inference: There are other standards of a good causal explanation apart from it being the most likely explanation. The most likely explanation may well tend to be very thin and as an evaluative standard push policy studies in a direction away from the rich, thick descriptions of historical context and the nuanced, multidimensional causal concepts involved in existing, established accounts of policy change.

These are standards of a good causal explanation for CPT to aim to achieve as well as the most likely one. It is beyond the scope of this paper to elaborate thickness and richness as standards of good causal explanation (see, for example, Gerring, 2012; Lipton, 2004; Scheffer & Niewoehner, 2010). For our purposes here, we note that there are such standards and that the problems that Bayesian tests of causality present for CPT in policy studies are not insurmountable nor should they deter policy scholars from using CPT. Of course, the question of causal selectivity—what is included in good thick and/or rich explanations of policy change over time—is a significant and challenging one for the policy studies field to reach a consensus answer. The theoretical pluralism of CPT investigated earlier assists with the search for contrasts and alternatives not taken that are relevant to a good causal account in

policy studies. The “why this and not that” standard is widely practiced in assessments of high-quality policy scholarship, especially when dealing with the products of case study research designs. Yet this standard is not part of Bayesian standards of causal assessment. The way forward for CPT in policy studies is to acknowledge the problematic of explaining complex, contingent conjunctions which are acutely time sensitive in an open system in a single case and develop standards of richness and thickness for assessing good explanations.

Conclusions

This paper has argued that CPT offers fertile territory for supporting research into the main research questions which occupy and motivate scholarship on policy change. While CPT can be used for between-case analysis of different causal paths in small-N case study research, we also see its distinctive advantages for policy studies in contributing to within-case analysis of causal processes. Here, it holds the promise of richer accounts of causation than standard regression analysis and more explicit theory-driven accounts of policy change than narrative explanations commonly used in single-case studies of policy change over time.

If CPT is to be adopted more widely by policy scholars, commitment should be made to its application in a robust manner that minimizes sources of inferential error. It is perhaps the causal part of CPT that is in the need of the greatest work for the method to be used more widely and successfully in policy studies. Establishing causation is time consuming, difficult, and demanding, and policy studies can still produce useful and plausible results without causal inferences (particularly its fast-growing postpositivist strands), but it remains the lodestone of much qualitative case study research into policy change.

In the emerging CPT literature, a Bayesian approach has been established as the main basis for causal inferences that appeal to mechanisms. Here, the probability estimate for a hypothesis about the operation of a mechanism being true is updated as additional evidence is learned. The philosophical question of whether this actually establishes causation is not something, in one sense, the field of policy studies needs to resolve. It is useful to know the probability of *A* given *B*, and this is part of our justification for why one particular study of a policy process is a better account than another. However, in another sense, it is still important to acknowledge that the Bayesian approach, while it may be consistent with, does not drive the richly detailed, historically grounded causal explanations of policy change that already exist in policy scholarship.

Causal models have exogenous factors, endogenous factors, and background variables. The importance of background variables to the CPT method underpins the requirement for the rich, detailed, and fine-grained contextual analysis set out in steps one to three above. Distinguishing necessary and sufficient conditions is essential to establishing background causation, but because CPT is both causal and idiographic, this requires claims of the sort: *Y* would have been *y* if and only if *X* had been *x*. This is why the paper has given prominence to the need for greater use of

counterfactual reasoning within policy studies. While this practice is easy to dismiss as mere speculation, it can be useful to supplement what, how, and why questions with “what if.” These can help increase the justification of causal claims, particularly those parts that include reference to background variables. Counterfactuals are also useful when we lack evidence about particular mechanisms in the causal chain. The old adage that the absence of evidence is not the same as evidence of absence applies here, and counterfactual reasoning can help researchers to increase confidence of claims.

What standards of “what if” reasoning might be advanced to help the CPT method burnish background causation claims or plug evidence gaps? We do not have a fully developed answer to this. At a minimum, counterfactual reasoning must be useful for identifying some mechanism for which we do not possess primary or secondary data but which we suspect is germane to our model of causal relations in the particular policy process being traced. While this is imperfect, at least we might provide a known unknown and render explicit on what basis our causal claims in policy studies rest.

Adrian Kay is a professor of government in the Crawford School of Public Policy at the Australian National University. Professor Kay’s research interests are at the intersection of international and comparative public policy, with an empirical focus on agriculture, food, and health.

Phillip Baker is a postdoctoral research fellow in the Regulatory Institutions Network (RegNet) at the Australian National University. Dr. Baker’s research interests are in health policy, determinants of population health equity, and public nutrition intervention.

Notes

This paper is based upon research supported by the Australian Research Council Discovery Program under grant DP120103676, “The making and unmaking of Australian public policy: understanding the path from Medibank to Medicare.”

1. George and Bennett (2005, p. 10) also identify a fourth. The development of a more general explanation rather than tracing a detailed causal process, used when a “higher level of generality” is needed, that “does not require a detailed tracing of causal sequence.”
2. For an excellent example, see Collier (2011, p. 826).
3. Distinct from data matrices or “data-set observations” used in quantitative analysis. See Collier et al. (2004). Some of the literature on which this article draws (e.g., Brady & Collier, 2010; Mahoney, 2010) formulates arguments in terms of CPOs, rather than in terms of process tracing *per se*.
4. See George and Bennett (2005, pp. 99–105), for insights on assessing the evidentiary value of archival materials in particular.
5. For an excellent and concise description of these tests and how they can be applied in practice, see Collier (2011). Rohlfing (2014) has recently proposed an expanded typology by reinterpreting the doubly decisive test to distinguish between theoretical and empirical uniqueness.
6. Zaks (2011) has recently argued that the first three tests should logically be reframed as leverage, necessity, and sufficiency tests, respectively.
7. Mahoney has provided a set theoretical logic for differentiating between these five causes. He also introduced the concept of tightness and the method of sequence elaboration for assessing the

importance of causes. Applying this logic, he provides an inventory of causes as a useful heuristic. See Mahoney et al. (2009).

8. For further differentiations between these variants and the specific steps involved in their application, see Beach and Pedersen (2013).

References

- Beach, Derek, and Rasmus Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. Ann Arbor, MI: University of Michigan Press.
- Bennett, Andrew. 2010. "Process-Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed., eds. Henry E. Brady, and David Collier. Lanham, MD: Rowman & Littlefield, 207–21.
- Bennett, Andrew, and J. T. Checkel. 2012. *Process Tracing: From Philosophical Roots to Best Practices*. Simons Papers in Security and Development, No. 21/2012. Vancouver: School for International Studies, Simon Fraser University.
- Bennett, A., and Colin Elman. 2006a. "Complex Causal Relations and Case Study Methods: The Example of Path Dependence." *Political Analysis* 14 (3): 250–67.
- . 2006b. "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science* 9: 455–76.
- Brady, Henry, and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.
- Cairney, Paul. 2013. "Standing on the Shoulders of Giants: How Do We Combine the Insights of Multiple Theories in Public Policy Studies." *Policy Studies Journal* 41 (4): 1–21.
- Checkel, Jeffrey T. 2006. "Tracing Causal Mechanisms." *International Studies Review* 8 (2): 362–70.
- Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44 (4): 823–30.
- Collier, David, Henry E. Brady, and Jason Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, eds. Henry E. Brady and David Collier. Lanham, MD: Rowman and Littlefield, 229–71.
- . 2010. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed., eds. Henry E. Brady, and David Collier. Lanham, MD: Rowman & Littlefield, 161–200.
- Collier, David, Jody LaPorte, and Jason Seawright. 2012. "Putting Typologies to Work: Concept Formation, Measurement, and Analytic Rigor." *Political Research Quarterly* 65 (1): 217–32.
- Ezzy, Douglas. 2002. *Qualitative Analysis Practice and Innovation*. Crows Nest, NSW, Australia: Allen & Unwin.
- Falleti, Tulia G., and Julia F. Lynch. 2009. "Context and Causal Mechanisms in Political Analysis." *Comparative Political Studies* 42 (9): 1143–66.
- Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43 (2): 169–95.
- George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gerring, John. 2007. *Case Study Research*. Cambridge: Cambridge University Press.
- . 2012. *Social Science Methodology: A Unified Framework*. Cambridge: Cambridge University Press.
- Grzymala-Busse, Anna. 2011. "Time Will Tell? Temporality and the Analysis of Causal Mechanisms and Processes." *Comparative Political Studies* 44 (9): 1267–97.
- Hall, Peter A. 2006. "Systematic Process Analysis: When and How to Use It." *European Management Review* 3 (1): 24–31.
- Head, Brian. 2008. "Wicked Problems in Public Policy." *Public Policy* 3 (2): 110–18.

- Hedström, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49–67.
- Howlett, Michael. 2009. "Process Sequencing Policy Dynamics: Beyond Homeostasis and Path Dependency." *Journal of Public Policy* 29 (3): 241–62.
- Howlett, Michael, and Ben Cashore. 2009. "The Dependent Variable Problem in the Study of Policy Change: Understanding Policy Change as a Methodological Problem." *Journal of Comparative Policy Analysis* 11 (1): 33–46.
- King, Gary, Robert O. Keohane, and Sidney Verba. 2001. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Levy, Jack. 2008. "Counterfactuals and Case Studies." In *The Oxford Handbook Of Political Methodology*, eds. J. M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press, 627–44.
- Lipton, Peter. 2004. *Inference to the Best Explanation*, 2nd ed. Oxford: Routledge.
- Mackie, J. L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2 (4): 245–64.
- Mahoney, James. 2000. "Strategies of Causal Inference in Small-N Analysis." *Sociological Methods & Research* 28 (4): 387–424.
- . 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62 (1): 120–47.
- . 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41 (4): 570–97.
- Mahoney, James, Erin Kimball, and Kendra Koivu. 2009. "The Logic of Historical Explanation in the Social Sciences." *Comparative Political Studies* 42 (1): 114–46.
- Mayntz, Renate. 2004. "Mechanisms in the Analysis of Social Macro-Phenomena." *Philosophy of Social Sciences* 34 (2): 237–59.
- Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. London: Sage Publications.
- Robson, Colin. 2002. *Real World Research: A Resource For Social Scientists and Practitioner-Researchers*. Oxford, UK; Malden, MA: Blackwell Publishers.
- Rohlfing, Ingo. 2014. "Comparative Hypothesis Testing Via Process Tracing." *Sociological Methods & Research* 43 (4): 606–42.
- Sabatier, Paul A. 2007. "The Need For Better Theories." In *Theories of the Policy Process*, ed. Paul A. Sabatier. Boulder, CO: Westview, 3–21.
- Sabatier, Paul A., and Hank C. Jenkins-Smith. 1993. *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder, CO: Westview Press.
- Scheffer, Thomas, and Joerg Niewoehner, eds. 2010. *Thick Comparisons*. Leiden: BRILL.
- Schlager, Edella. 2007. "A Comparison of Frameworks, Theories, and Models of Policy Processes." In *Theories of the Policy Process*, ed. Paul A. Sabatier, Boulder, CO: Westview, 293–321.
- Steinberg, Paul F. 2007. "Causal Assessment in Small-N Policy Studies." *Policy Studies Journal* 35 (2): 181–204.
- Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- Yin, Robert. 1993. *Applications of Case Study Research*. London: Sage.
- . 1994. *Case Study Research: Design and Methods*, 2nd ed. Thousand Oaks, CA: Sage Publications.
- Zaks, Sherry. 2011. "Relationships among Rivals: Analyzing Contending Hypotheses with a New Logic of Process Tracing." Annual Meeting of the American Political Science Association, Seattle, Washington.