Process Tracing: Defining the Undefinable?

Christopher Clarke*

Forthcoming in *The Oxford Handbook of the Philosophy of Political Science*

Abstract

A good definition of process tracing should highlight what is distinctive about process tracing as a methodology of causal inference. I look at eight criteria that are used to define process tracing in the methodological literature, and I dismiss all eight criteria as unhelpful (some because they are too restrictive, and others because they are vacuous). In place of these criteria, I propose four alternative criteria, and I draw a distinction between process tracing for the ultimate aim of testing a start–end hypothesis versus process tracing as an ultimate end in itself. Although it is clear enough how the former method works, there is still much methodological work to be done in understanding the latter method as a distinctive method of causal inference, I argue.

1 Introduction

Process tracing is a tool that political scientists use to hunt the causes of political phenomena. Indeed "process tracing is perhaps the tool of causal inference that first comes to mind when one thinks of qualitative methodology in political science" (Mahoney 2010, 123). It "represents the empirical core of many, if not most, case studies" (Rohlfing 2012, 150). Exemplars include classic works such as Skocpol (1979) on the French, Russian and Chinese revolutions, Tannenwald (1999) on the taboo against using nuclear weapons during the Cold War, and

*This is a draft of a chapter that has been accepted for publication by Oxford University Press in the forthcoming book *The Oxford Handbook of Philosophy of Political Science* edited by Jeroen van Bouwel and Harold Kincaid due for publication in 2021. Thank you Jeroen van Bouwel, Harold Kincaid, and one anonymous referees for your constructive comments on the manuscript. And thanks to an audience at the University of Kent for discussion of these ideas. This work has received funding from the European Research Council under the European Union Horizon 2020 Research and Innovation Programme, under grant agreement no 715530.

Wood (2003) on the civil war in El Salvador, amongst many other lauded studies of elections, government policy-making, international relations, and the like. What's more, it's claimed that process tracing can measure actual causal effects, in contrast to quantitative methods that supposedly can only measure probabilistic "averages" of these causal effects (Mahoney and Goertz 2006; Crasnow 2012). So it is no surprise that political methodologists have started to pay process tracing a lot of attention.

This chapter seeks a precise definition of process tracing. The main reason to seek such a definition goes back to King, Keohane, and Verba (1994), the manifesto which sparked one of the central methodological debates within political science: what are the aims, limits and logic of the methods of political science? in particular, how do quantitative methods differ from qualitative methods in this regard? As methodologists have come to see process tracing as one of the central methods in the qualitative toolkit, much of this "post-KKV" debate has begun to focus on process tracing.

As I see it, a methodologically useful definition of process tracing should start with some exemplars of process tracing in the political science literature, and it should work out what most of these exemplars have in common, epistemically speaking. What is the common rationale that warrants the conclusions that each process tracing study draws? What epistemic problems and epistemic benefits does this rationale give rise to? How is this rationale distinct from the rationale that lies behind other studies that draw causal conclusions, quantitative studies in particular? If you define process tracing too broadly, thus grouping together studies that have different epistemic problems and epistemic benefits, then you risk mis-identifying those problems and benefits. If you define process tracing too narrowly, then you miss an opportunity to understand the epistemic problems and benefits of those studies that you've excluded from your definition.

Finding a useful definition will also help alleviate the concern that process tracing is an "ad hoc" method (Gerring 2006, 178), one that is "carried out informally and without a high level of transparency" and that "lacks systematization of technique and explicitness in execution" (Mahoney 2015, 201). As Beach and Pedersen (2016, 302) put it: "process-tracing has become a buzzword in recent years, but until recently there has been little agreement about what we actually are tracing or how to conduct this tracing properly".[1]

Here's how I will proceed. My proposed definition of process tracing contains four criteria. As I explain in Section 2, everyone agrees that

> (Criterion One) For a study to count as process tracing, it must identify the presence or absence of intermediate causal links between some factor and an outcome of interest.

[1] See Bennett and Checkel (2015) and Checkel (2013) for agreement.

In this sense, process tracing is the search for a chain of causes that led to an outcome of interest. Section 2 then explains how some methodologists place a further limit on what counts as process tracing:

> (Criterion Two) For a study to count as process tracing, it must describe each of these intermediate links in terms of entities engaging in regular, well-understood activities.

Section 3 then examines the idea that, although process tracing identifies intermediate links in a causal chain, this is not the ultimate aim of a process tracing study; instead it is merely a means of pursuing the following ultimate aim:

> (Criterion Three — Version A) For a study to count as process tracing, its ultimate aim must be just to test the hypothesis that the starting factor in the chain was a cause of the final outcome. Indeed, it must pursue this ultimate aim via a particular argumentative structure, one that I try to describe in Section 3.

Section 4 contrasts this view of the ultimate aim of process tracing with an alternative:

> (Criterion Three — Version B) For a study to count as process tracing, its ultimate aim must be the identification / description of the intermediate links in a causal chain.

That is to say, the description of these intermediate links is not a means to some further end, such as testing the hypothesis that first factor in the chain was a cause of the final factor. In Section 4, I show that whether a process tracing study is a success or as a failure can depend on whether version *A* or version *B* of Criterion Three is the correct view of the ultimate aim of the process tracing study in question.

Section 5 then examines the suggestion that what distinguishes process tracing from quantitative methods is in part the fact that quantitative methods use "rectangular" data-sets. I argue that the use of rectangular data-sets is a superficial symptom of a more fundamental difference between process tracing and quantitative methods. Namely:

> (Criterion Four) A study does not count as a process tracing study if it requires (Version *A*) any of the start–end causal relationships it measures, or (Version *B*) any of the intermediate causal relationships it measures, to be "unit homogeneous" across a largish population of cases.

In addition to criteria such as these, some methodologists explicitly offer (or implicitly hint at) the following limits on what counts as process tracing:

> (F1) A study does not count as a process tracing study if it relies on numerical data;
>
> (F2) A study does not count as a process tracing study if it relies on evidence from cases other than the single case under examination;
>
> (F3) A study does not count as a process tracing study if it needs the probability distribution of some variable to take a particular form;
>
> (F4) A study does not count as a process tracing study if it uses statistical methods;
>
> (F5) A study does not count as a process tracing study if its results can be generalised.

Section 6 will argue that these suggested criteria *F1–F5* are incorrect: the limits they place on what counts as process tracing are unhelpfully restrictive. Section 7 then examines another three putative criteria:

> (E1) For a study to count as process tracing, it must only use causal process observations, that is, diagnostic evidence to uncover the mechanism under examination;
>
> (E2) For a study to count as process tracing, it must only use pattern matching;
>
> (E3) Process tracing can use a very wide variety of evidence.

I argue that criteria *E1–E3* are empty: they do not place any substantive limits on what counts as process tracing. So we are left with Criterion One Two Three and Four as the primary ways of defining process tracing.

This raises a problem for process tracing version *B*—process tracing as an end in itself. This is because standard techniques of causal inference, such as regression, require unit homogeneity (King, Keohane, and Verba 1994, 91–94). So Criterion Four Version *B* rules out using standard causal inference techniques to study these intermediate causal relationships. So it is currently unclear how process tracing version *B* proposes to establish these intermediate causal links. To repeat, criteria *E1–E3* are empty: they do not describe how process tracing establishes intermediate causal links. (In contrast, process tracing for the ultimate aim of testing start–end hypotheses can appeal to any techniques it likes to establish intermediate causal links, including standard quantitative techniques.) Therefore I offer my four criteria as a helpful starting point from which one might build an account of process tracing version *B*, process tracing as an end it itself. If I am right, the methodological literature has only begun to answer the question: how do process tracers trace these intermediate links in causal chains?

2 Process Tracing as Identifying Intermediate Links

The first defining feature of process tracing, everyone agrees, is that process tracers search for (and try to make explicit) the mechanism that helped produce a given outcome in a given case or cases. What, for example, was the mechanism that prevented the Soviet Union from sending the Red Army to crush the pro-democracy movements in Poland and East Germany in 1989? What was the mechanism that caused El Salvadorian elites to compromise with the insurgents at the end of the civil war in 1992? What was the mechanism whereby new European Union states were integrated such that they came to share the norms of the other member states?

Of course, methodologists disagree about what it means to describe the mechanism that caused a given outcome (Hedstrom and Ylikoski 2010). Among methodologists interested in process tracing, however, there is at least agreement on one point. Namely, to describe a mechanism one must at very least describe a chain of causes that ends in the outcome in question: *A* caused *B*, *B* caused *C*, *C* caused *D*, and *D* caused *E*.[2] For example, Wood describes the following causal chain: (A) insurgents in El Salvador took over plantations in the 1980s. This caused (B) the El Salvadorian economy to rely less on primary commodities. This caused (C) the El Salvadorian elite's preferences to become more democratic. This caused (D) the elite and the insurgents to begin to bargain. And this caused (E) a transition towards democracy. Thus process tracing searches for the chain of causes that led, in the case under examination, to the outcome one is interested in:

> (Criterion One) For a study to count as process tracing, it must identify the presence or absence of intermediate causal links between some factor *A* and an outcome of interest *E*.

To this requirement, some methodologists would add that to describe a mechanism, one must also describe this causal chain in such a way that each link in the chain is described as an "entity" engaging in a regular and well-understood "activity". For example, the most profitable way for capital owners to produce primary commodities is to repress their workers, one might think; but to produce other commodities and services repression is less effective. If true, this fact helps one understand the causal link between *B* (the shift away from primary commodities) and *C* (elite preferences for democracy). This link arises, one might argue, because this new economic structure *B* constrained the opportunities available to capital owners; they are no longer able to make a profit by repressing their workers. And the idea of an economic structure constraining or expanding an

[2] See Gerring (2006) and Checkel (2008) and indeed most of the citations in his chapter.

individual's opportunities is a well-understood activity. What's more, all or most individuals adhere to the norms of belief–desire psychology, one might think, and so this activity that operates across individuals in a more or less regular fashion. Therefore, to describe the link between B and C in these terms is to describe it as an entity engaged in a regular and well-understood activity. And this is required if one wants to truly be describing the mechanisms that links A with E, say many methodologists.[3] Thus some methodologists would add to our definition of process tracing:

> (Criterion Two) For a study to count as process tracing, it must describe each of these intermediate links in terms of entities engaging in regular, well-understood activities.

Some theorists would elaborate on this point and say that, in political science, the best understood and most regular activities are those in which goal-orientated agents make decisions in light of the opportunities available to them.[4] For example, an individual member of the elite decides to bargain with the insurgents because that is the best way for this individual to maximize the return on the capital she owns. Or at least, one might say, the best understood (macro) activities are those activities for which one knows how the activity in question arises out of goal-orientated agents interacting with each other. Take for example a drought making it more difficult for farmers to grow coffee and thus causing a rise in its price. This is an entity of sorts (a drought) engaging in an activity (driving up the price of coffee). Standard micro-economic theory makes some assumptions about how consumers and firms interact with each other in market contexts, and about consumers' preferences for coffee and tea and other goods; and from these assumptions it deductively proves that the price of coffee will rise. And, insofar as this story is more or less true, then, this activity is well-understood, one might think. (A comparison: a bunsen burner increases the pressure of a gas in a sealed container. One might think that this activity is only well-understood once one knows how this activity arises from the fact that a gas is just billions of individual particles moving in accordance with Newton's laws of motion.)

So we have one uncontroversial criterion (Criterion One) and one somewhat more controversial criterion (Criterion Two) for what counts as process tracing. At very least, process-tracing looks for the presence or absence of intermediate causal links between some factor A and an outcome of interest E.

[3] Citations to e.g. Beach and Pedersen (2016) who agree, and those who disagree.

[4] This suggest seems to be strongly implied in Evangelista (2015) Waldner (2012) Bennett (2008) and George and Mckeown (1985).

3 The Logic of Testing Start–End Hypotheses

Methodologists use lots of suggestive metaphors to describe what process tracers do. Process tracers "break down" or "dis-aggregate" causes (Gerring 2006). They trace causes "forward" and "backward" in time (Bennett 2010). Their method "focuses" on the "intervening steps" in a causal chain (Bennett 2010). But it is surprisingly rare for methodologists to make precise the specific logic behind process tracing. And this is notwithstanding the fact that process tracers often point to Bayesianism as the general epistemological framework within which process tracing is to be understood (more on this point in due course).

However, at times, some methodologists seem to suggest the following as the specific logic behind process tracing:

> (Criterion Three — Preliminary Version) For a study to count as process tracing, it must ultimately aim to support the hypothesis that the starting factor in the chain *A* was a cause of the final outcome *E*. And the means by which it pursues this ultimate aim must be the following argumentative structure.
>
> Premises: *A* caused *B*. *B* caused *C*. *C* caused *D*. *D* caused *E*.
>
> Conclusion: *A* caused *E*

This criterion views process tracing as the study of intermediate causal links in order to *support a start–end hypothesis*, in this case the hypothesis that *A* caused *E*. Waldner (2012) is the most explicit in identifying the logical structure of this reasoning, which he calls the logic of "concatenation".

> To claim the existence of a causal chain is to claim that given entities and mechanisms, one event constrained future events such that a subsequent event was bound to happen, or at least that an earlier event substantially shifted the probability distribution governing the subsequent event, making sequelae far more probable. … Process tracing accounts, for this reason, tend to make deterministic causal claims that, given a set of initial and scope conditions, an outcome was bound to occur. (Waldner 2012, 69)

Some other methodologists hint at this logical structure too:[5]

> For example, when analysing World War I, it seems plausible to believe that Austria's decision to initiate war with Serbia was necessary for the general

[5] See for example Mahoney (1999), Mahoney (2004), Rohlfing (2012), Bennett (2010), Mahoney (2015), Collier (2011b), and perhaps also Humphreys and Jacobs (2015), and Gerring (2006).

> war. In turn, Austria's decision may have required the assassination of Ferdinand, such that this assassination may also have been necessary for the general war. (Mahoney 2015, 213)
>
> If one can theorise a full sequence of steps, one should also test for the presence of each step because otherwise parts of the argument remain untested. In set-relational terms, this means that each step is a necessary component and they are jointly sufficient for inferring that the purported cause indeed is a cause. (Rohlfing 2012, 152)
>
> Process-tracing study provides stronger evidence of the existence of a causal process(es) linking *C* and *O* because evidence is provided for each part of the causal process. (Beach and Pedersen 2016, 176)
>
> In process-tracing, if we can collect confirmatory evidence of each part of a mechanism that we can trust, we can make a relatively strong inference that *C* is causally related to *O* through a causal mechanism. (Beach and Pedersen 2016, 273)
>
> By tracing an explicit mechanism, we would be more justified in making the claim that *C* is linked to *O* through a causal mechanism. (Beach and Pedersen 2016, 83)

It's worth noting in passing that there is disagreement amongst the methodologists quoted here as to the nature of the causes being traced here. As you can see, some treat causes as in some sense necessary for their effects; others treat causes as in some sense sufficient for their effects.

I will now show that at least one exemplary process tracing study fails criterion Criterion Three, in its preliminary form given above. This study is not a study of intermediate causal links to support a start–end hypothesis. The exemplary study I have in mind is Brady (2010)'s analysis of the 2000 presidential election in the United States. In this election, the national media reported at 8pm EST that the polls in Florida had closed. But this report was false: ten counties in Florida's western Panhandle are in the Central time zone (CST) not the Eastern time zone (EST). As a result, the polls in these ten counties remained open until 9pm EST. What's more, at 8:48pm EST the national media reported that Al Gore had won Florida, even though polls in these ten counties were open for another 12 minutes. In these two respects, the media's call was premature. Did this premature media call cause would-be voters in these ten staunchly pro-Bush counties not to vote?

Lott (2005) uses a standard quantitative technique (differences-in-differences regression) to estimate that, if the media had not prematurely called the Florida election, Bush's victory over Al Gore in Florida would have been 7,500 to 37,500 votes greater than it actually was. Against this, Brady uses process tracing to dispute Lott's conclusions. Brady calculates that (F) there were 379,000 would-be voters in the Florida panhandle. That is to say, people who intended to vote in

the election. Since the media declared a Gore victory twelve minutes before the close of the polls, Brady focuses his attention on late would-be voters: people who hadn't yet voted by the last twelve minutes of the polls being open, but who intended to do so. Drawing on evidence that in 1996 Florida voters didn't vote in disproportionately large numbers in the last hour, and assuming that 1996 and 2000 are similar in this respect, Brady infers from (F) that (G) there were at most 4,200 would-be late voters in 2000.[6] Drawing on general evidence about media reach, Brady infers from (G) that (H) at most 840 would-be late voters heard the media call. And drawing on evidence about the support for Bush in the Florida panhandle, he infers from (H) that (I) at most 560 would-be late Bush-voters heard the media call. And drawing on general evidence about voters' responses to premature media calls, he infers from (I) that (J) at most 56 would-be Bush-voters decided not to vote because of the media call. Even assuming that no would-be Gore-voters decided not to vote because of the media call, it follows that (K) if the media hadn't called the election prematurely, Bush's victory would at most have been 56 votes greater. Quite a different estimate from Lott's estimate of 7,500 to 37,500![7] This is a quick but (I think) faithful reconstruction of Brady's process tracing argument.[8]

What is the logical structure of Brady's process tracing argument? On a first glance, one might be tempted to say the following: F caused G, and G caused H, and H caused I, and I caused $\mathcal{J}$. So Brady's argument is clearly a study of intermediate causal links to support a start–end hypothesis, one in which the causal chain runs F to G to H to I to $\mathcal{J}$. But a second look shows that this interpretation is incorrect.[9] To see this, note that a inference to support a start–end hypothesis using a causal chain running from F to $\mathcal{J}$ is an argument for the conclusion that F caused $\mathcal{J}$. Applied to this case, the conclusion would be: the fact that there were 379,000 would-be voters in the panhandle caused the fact that at most 56 would-be Bush-voters decided not to vote because of the media call. But this is

[6] Brady provides a second argument for the figure of 4,200, an argument that is more dubious in my view. This argument simply assumes that would-be voters did not intend to vote in disproportionately high numbers in the last hour. But the warrant for this assumption is highly questionable.

[7] However, Brady is clearly focusing only on the second premature call (the 8:48pm report that Gore had won). But Lott seems to focus on the two premature calls together (including the 8pm report that the polls had closed).

[8] Brady talks about 'voters' rather than 'would-be' voters as I do. But it seems to me that for Brady's argument to work he needs to be talking about would-be voters. So to make Brady's argument sound, one needs to tinker with the figures that Brady gives. For example, I suspect that Brady's figure of 379,000 is a report of the number of actual votes cast rather than an estimate of the number of would-be voters.

[9] One might also dispute that some of these links in the chain are truly causal—the link between H and I to take just one example. But I will not explore this point further.

not Brady's conclusion. Brady's conclusion is instead $\mathcal{J}$ itself: at most 56 would-be Bush-voters decided not to vote because of the media call.[10] And so the structure of Brady's argument—despite appearances—is not a study of intermediate causal links to support a start–end hypothesis on the causal chain running from F to $\mathcal{J}$.

Instead, the logic of Brady's argument is different. Here's how I see it. For any would-be voter we can ask whether the following conditions hold: (1) did the early media call A cause (B) this would-be voter to hear the media call? and (2) did hearing the media call B cause (C) this would-be voter not to vote, rather than voting for Bush? The evidence that Brady provides suggests that of the 379,000 would-be voters, at most $1/5$ heard the media call, and so $1/5$ satisfy condition one. It also suggests that at most $1/90$ intended to vote late (rather than vote early or vote via absentee ballot) of which $2/3$ intended to vote for Bush, and of which $1/10$ would be put-off by an early call, so at most $1/1350$ satisfy condition two. So at most $1/6750$ would-be voters satisfy both condition one and condition two. That is to say, at most 56 people. However, one might claim that A can only cause C via a causal chain running through B. The only way in which the media call could influence a given voter's behavior is through that voter actually hearing the media call. It follows that the number of people for which A (the media call) caused C (that individual to refrain from voting for Bush) is at most 56 people. That's the logic of Brady's argument as I see it. (Interestingly, on my reconstruction, there is only one intermediate link in the chain being analysed, rather than three or four as appears on a first glance. The inferential complexity instead arises because Brady needs to piece together lots of assumptions about population frequencies—$1/90 \times 2/3 \times 1/10$—in order to infer that at most $1/1350$ would-be voters satisfy condition two.)

Note that, as I've reconstructed it, at no point does Brady's argument support a start–end hypothesis. Yes, it's true that—for any individual for whom A caused B and for whom B caused C—one might infer that A caused C. But this extra inference is entirely dispensable to Brady's overall argument. Instead the key move in Brady's argument has an altogether different structure. It's a study of intermediate causal links in order to *undermine a start–end hypothesis*:

> Premise: A caused C only if A caused B, and only if B caused C.
>
> Premise: A did not cause B (or B did not cause C).
>
> Conclusion: A did not cause C.

Taking stock, I've identified two types of argumentative structure that are the basis of some process tracing studies: studying intermediate links to support a start–end hypothesis and studying intermediate links to undermine a start–end hypothesis.

[10] Or K itself.

I now want to clarify these two argumentative structures by responding to the following sort of worry. Most methodologists agree that the general epistemological framework in which process tracing should be understood is Bayesian epistemology (Humphreys and Jacobs 2015; Fairfield and Charman 2017): a causal hypothesis is tested by the observations that one expects to make, if that hypothesis is true. But my two forms of studying intermediate links don't seem to do that. They are presented as arguments in a premise–conclusion form, rather than as a probability function, which is the main currency of Bayesian epistemology. It's therefore unclear how they fit into the Bayesian framework. That's the concern.

(Readers unfamiliar with the Bayesian epistemological framework might consult Howson and Urbach ([1989] 2006) to understand the notation and underlying ideas in this section. Note that Bayesian epistemology differs from Bayesian statistics. Bayesian epistemology is a fully general epistemological framework capable of handling hypotheses of any type, causal hypotheses for example; Bayesian statistics is a more narrow and less philosophical. It's a set of tools for estimating the parameters of probability distributions.)

To address this important concern, I will now reformulate these two types of argumentative structure to bring out what is common to both, and to show how they fit within the Bayesian epistemological framework. To begin, let e denote the evidence uncovered by a given study in political science. And let h denote the start–end causal hypothesis that the study ultimately aims to test: *A* caused *C*. And consider the following two intermediate hypotheses: h_{AB} is the hypothesis that *A* caused *B*; and h_{BC} is the hypothesis that *B* caused *C*.

Now, for any given study one can ask: does the evidence e that it uncovers bear on start–end hypothesis h only by first bearing on these intermediate hypotheses? The intuitive idea here is that if one were to become certain of the truth or falsity of h_{AB}, and similarly of h_{BC}, then e would have no further bearing upon h. Suppose, for example, one were to learn of a given individual that h_{AB} the early media call caused her to hear the early media call, and h_{BC} hearing the early media call did indeed cause her not to vote, rather than voting for Bush. In this case, the evidence e' that two thirds of the voters in the panhandle are pro-Bush becomes irrelevant for evaluating h, the hypothesis that the media call caused this voter not to vote (rather than voting for Bush). And the same goes for all the other evidence e collected in Brady's study. Put in terms of probabilities, the currency of the Bayesian framework: $P(h|h_{AB}h_{BC}) = P(h|h_{AB}h_{BC}e)$ and $P(h|\neg h_{AB}h_{BC}) = P(h|\neg h_{AB}h_{BC}e)$ and $P(h|h_{AB}\neg h_{BC}) = P(h|h_{AB}\neg h_{BC}e)$ and $P(h|\neg h_{AB}\neg h_{BC}) = P(h|\neg h_{AB}\neg h_{BC}e)$.

What's more, one can usually split the overall evidence e that a study uncovers into two or more parts. For example, one part of Brady's evidence was e_1 his general evidence about media listening rates; and a distinct part of Brady's evidence

was his evidence e_2 about the support for Bush in the panhandle. Thus for each intermediate hypothesis one can ask: is there part e' of the overall evidence e that bears on this intermediate hypothesis alone? In the Brady study, for example, evidence e_1 bore on h_{AB} but not on h_{BC}. That is to say, his general evidence about media listening rates bore on the hypothesis that the early media call caused a given individual to hear the media call; but it did not bear on the hypothesis that a given individual hearing the media call would cause her not to vote for Bush. In terms of probabilities: $P(h_{AB}|e_1) \neq P(h_{AB})$ but $P(h_{BC}|e_1) = P(h_{BC})$. Similarly, evidence e_2 bore on h_{BC} but not on h_{AB}. That is to say, Brady's evidence about the support for Bush in the panhandle bore on the hypothesis that a given individual hearing the media call would cause her not to vote for Bush; but it did not bear on the hypothesis that the early media call caused given individual to hear the media call. In terms of probabilities: $P(h_{BC}|e_2) \neq P(h_{BC})$ but $P(h_{AB}|e_2) = P(h_{AB})$.

This now puts me in a position to define the method of identifying intermediate causal links for testing start-end hypotheses:

> (Criterion Three — Version *A*) For a study to count as process tracing, it must ultimately aim to test the hypothesis that the starting factor in the chain was a cause of the final outcome. And the means by which it pursues this ultimate aim must be the following logical structure.
>
> (I) The overall evidence e bears on hypothesis h only by first bearing on the intermediate hypotheses.
>
> (II) For each intermediate hypothesis, there is a piece of evidence e' that bears on this intermediate hypothesis alone.[11][12]

In the Brady example that I've just analysed, I considered the start–end hypothesis h that *A* caused *C* and the intermediate hypothesis h_{AB} that *A* caused *B* and the intermediate hypothesis h_{BC} that *B* caused *C*. That is to say, I considered a single causal chain with only one intermediate link. But Criterion Three

[11] Condition *II* is necessary, I suggest, because many studies that test hypothesis h but don't intuitively do so "via" testing h_{CD} and h_{DE} will nevertheless satisfy condition *I*, I suggest. At any rate, for broad agreement with my suggestion here, see Beach and Pedersen (2016) and Gerring (2006).

[12] Note that Criterion Three does not require that one can partition the evidence e into evidence e_{AB} that bears on h_{AB} alone, and into evidence e_{BC} that bears on h_{BC} alone, with no evidence left over. Take for example a scenario in which one's evidence breaks down into three parts: e_A is evidence about the observed values of variable *A*, e_B is evidence about the observed values of variable *B*, and e_C is evidence about the observed values of variable *C*. Imagine further that e_A and e_B (but not e_C) bear on the intermediate hypothesis h_{AB} that *A* caused *B*; while e_B and e_C (but not e_A) bear on the intermediate hypothesis h_{BC} that *B* caused *C*. Therefore, e_A bears on h_{AB} alone, and e_C bears on h_{BC} alone, and so this case satisfies Criterion Three, even though e_B bears on both h_{AB} and on h_{BC}. (Thank you Harold Kincaid for pointing this out to me.)

encompasses more complicated examples. One might consider the start–end hypothesis h that *A* caused *E* and the intermediate hypothesis h_{AB} that *A* caused *B*, and h_{BE} that *B* caused *E*. And one might, in addition, consider the intermediate hypothesis h_{AC} that *A* caused *C*, and h_{CD} that *C* caused *D* and h_{DE} that *D* caused *E*. In this more complicated set-up, one is examining two causal chains that might run from *A* to *E*. The first has one intermediate link *B*, and the second has two intermediate links, *C* and *D*.

I conclude that many process tracing studies conform to the logic given in criterion Criterion Three (Version *A*), namely identifying intermediate causal links for the ultimate aim of testing start–end hypotheses.

4 Identifying Intermediates as an End in Itself?

In this section, I want to examine an alternative idea, the idea that process tracing does not ultimately aim to test start–end hypotheses. Instead:

> (Criterion Three — Version *B*) For a study to count as process tracing, its ultimate aim must be the identification / description of the intermediate links in a causal chain.

In other words, the previous section dealt with the identification of intermediate links for the ultimate end of testing start–end hypotheses; and this section will deal with identifying intermediate links as an ultimate aim in itself. (Of course, there are also other aims to which one might put a process tracing study in addition to these two aims. But these two aims seem to me to be primary. For example, the aim of making effective policy is parasitic on these two aims, I'd suggest: for any given policy objective, to gain the knowledge required for effective policy interventions is just to gain the causal knowledge of a start–end relationship or a intermediate relationship that the policymaker can exploit.)

Why might testing one or more intermediate hypotheses be valuable for its own sake? One answer is that, although the knowledge that *A* causes *E* provides some understanding of why *E* occurred, this understanding is limited. In contrast, if one knows that *A* caused *E* via one causal chain that runs through *B*, and also via a second causal chain that runs through *C* and *D*, one has a deeper understanding of why *E* occurred—especially if the links in this causal chain are described in terms of entities engaged in well-understood activities (see section 2).

To make the distinction between these two aims more vivid, imagine a scenario in which a method other than process tracing is able to establish that *A* caused *E*, to a high degree of certainty. Imagine, for example, that instrumental variables regression suggests that (A) the existence of institutions that guarantee

private property rights causes (E) long run economic growth (Acemoglu, Johnson, and Robinson 2005). Let's also imagine that the assumptions needed for this technique to be reliable are known to be true. In this case, this instrumental variables regression puts us in a position to know the start-end hypothesis that private property rights *A* cause growth *E*. Is it still worth tracing the process between *A* and *E*? If you answer "no" then your ultimate aim is just to test the start–end hypothesis that *A* caused *E*. But if you answer "yes" then you take the identification of the intermediate causal links between *A* and *E* to be an ultimate end in itself.

Of course, there is no reason why one cannot adopt both of these two aims together as the ultimate aim of a single process tracing study. Nevertheless, it is important to analyse these aims separately, I will now argue. The reason I think one needs to analyze these two aims separately is that testing a start–end hypotheses (via identifying intermediate links) is often more difficult than testing one or more intermediate hypotheses themselves. I will give three examples of this.

Example one. Consider the hypothesis that *A* the rise of Nazism in the 1930s was a cause of *E* the formation of the state of Israel in 1948. Suppose that one tests the following intermediate hypotheses: (i) the rise of Nazism *A* caused *B* the genocide of Jews in Europe in the 1940s as well as their persecution in the 1930s; (ii) *B* was a cause of *C* Jewish refugees fleeing to Mandatory Palestine from 1933–47; and (iii) *C* was a cause of *E* the formation of Israel. Let's also assume the difference-making view of causation: to say that *C* was a cause of *E* is to say that *C* made a positive difference to *E*, which is to say that if *C* had not occurred, then *E* would not have occurred either. Suppose that one then discovers evidence that Israel would have been formed at any rate, even without the Jewish refugees who fled to Palestine 1933–47. On this basis one concludes that *iii* is false: *C* was not a cause of *E*. In this case, I think, this discovery has deepened one's understanding of why the state of Israel was formed, in that one has ruled out one of the plausible causes of this outcome.[13] However, this discovery has not done much to test the start–end hypothesis that *A* was a cause of *E*, I suggest. After all, there are lots of other plausible causal chains that link *A* to *E*. For example: the rise of Nazism *A* caused *B** the war between the British Empire and Nazi Germany, which caused *C** damage to Britain's economy, which caused *D** Britain's inability to govern Palestine in the 1940s, which was a cause of *E* the formation of Israel in 1948. In sum, the study of intermediate links succeeds in rejecting the intermediate hypothesis that *C* caused *E*—and it thereby increases our understanding of the outcome *E*—but it fails to test the start–end hypothesis that *A* caused *E*. Judged by one ul-

[13] Philosophers often talk about understanding-why as causal knowledge. But I see no reason why this knowledge need be knowledge of the presence of a cause, rather than knowledge of the absence of a cause.

timate aim it's a somewhat successful analysis; judged by the other ultimate aim it's mostly unsuccessful.

Example two. Recall the World War I example from Section 3. Suppose one fails to find evidence for or against the hypothesis that *A* the assassination of Franz Ferdinand caused *B* Austria to start a war with Serbia. But suppose one finds strong evidence that *B* Austria's starting a war with Serbia caused *E* the general war. In this case, one has almost completely failed to test the start–end hypothesis that *A* the assassination of Franz Ferdinand caused *E* the general war. Nevertheless, one has increased one's understanding of why the general war occurred: it's because Austria started a war with Serbia. In this case, the study of intermediate links succeeds in establishing the intermediate hypothesis that *B* caused *E*—and it thereby increases our understanding of the outcome *E*—but again it fails to test the start–end hypothesis that *A* caused *E*. Judged by one ultimate aim it's a somewhat successful analysis; judged by the other ultimate aim it's mostly unsuccessful.

Example three. Repurposing an example from Hitchcock (2001), suppose that one discovers that *A* high investment in the railroads in the 1890s in the United States caused *B* low investment in canals and roads in the 1890s. And suppose that one also discovers that *B* low investment in canals and roads caused *E* transport costs to be high. This analysis succeeds in the aim of testing these intermediate hypotheses. (And it seems to me that, by that same token, it deepens one's understanding of why US transport costs are high.) However, one might think that this analysis of intermediate links, as it stands, fails to test the start–end hypothesis that *A* caused *E*. For example, one might think that causation is just difference making; and one might think that railroad investment *A* made a difference to transport costs *E* via a positive route and a negative route. The positive route is this: railroad investment *A* made a positive difference to *D* railroad efficiency, which itself made a positive difference to *E* transport costs. The negative route is this: railroad investment *A* made a negative difference to *B* canal and road investment, which itself made a positive difference to *E* transport costs. Therefore, one might think, whether *A* railroad investment made a positive or negative or no difference to *E* transport costs will be determined by the relative strength of these positive and negative routes. And—let's imagine—this relative strength is something that we have yet to examine. So the analysis of intermediate links succeeds in establishing several intermediate hypotheses—and it thereby increases our understanding of the outcome *E*—but again it doesn't go far in testing the start–end hypothesis that *A* caused *E*. Judged by one ultimate aim it's a totally successful analysis; judged by the other ultimate aim it's mostly unsuccessful. (Of course, some people reject the view of causation as difference making. And for some of these alternative views, causation is transitive: if *A* caused *B* and if *B* caused *E* then it's impossible for *A* not to have also caused *E*.)

On the basis of these three examples—this third example especially—I conclude that one can often test one or more intermediate hypotheses (and thereby deepen one's understanding of the final outcome) while also falling short of testing a start–end hypothesis. It follows from this that there is no such thing as a good or bad study of intermediate causal links. Rather there are good and bad studies relative to the ultimate aim of testing start-end hypotheses, and relative to the ultimate aim of testing intermediate hypotheses alone. It follows from this that we need two sets of methodological guidelines for analyzing intermediate links in a causal chain: one for those political scientists who aim only to test a start–end hypothesis, and another for those who aim only to test some intermediate hypotheses. (Political scientists who pursue both aims in a single study will need to follow both sets of guidelines. They will need to trade-off these aims against each other if and when the guidelines conflict with each other.) For this reason, it's important for methodologists to draw a distinction between these two ultimate aims for which one might study intermediate links.

Unhappily, the literature does not yet draw a sharp distinction between these two ultimate aims.[14] For example, Beach and Pedersen (2016) distinguish between "explaining outcome" process tracing on the one hand and "theory building" and "theory testing" process tracing on the other. But this distinction—it is crucial to note—is nothing like my distinction between studying intermediate links as an ultimate end in itself versus as a means that serves the ultimate aim of testing start–end hypotheses. As far as I understand Beach and Pedersen, "explaining outcome" process tracing means process tracing that reconstructs a very long and detailed causal chain. As a result, this precise causal chain will be idiosyncratic to the case in question and will not be found in other cases—even though any individual link in the chain may be found is some other cases. So Beach and Pedersen's distinction has to do with whether a causal chain exhibits uniformity across cases. It has nothing to do with the distinction between ultimately aiming to test a start–end hypothesis versus ultimately aiming to test one or more intermediate hypotheses.

Even though the literature does not sharply draw this distinction, there are some hints as to which aim various methodologists have in mind when they talk about process tracing. On the one hand, Section 3 has already cited methodologists who seem to endorse testing start–end hypotheses as the ultimate aim of process tracing. On the other hand, Checkel (2013) thinks that one important aim of process tracing is to test the hypothesis that A caused B caused E against the rival hypothesis that A caused D caused E. This is clearly testing intermediate

[14] Waldner (2012) gestures towards this distinction but does not clarify it or study its implications. Beach (2017) and Bennett and Checkel (2015) mention these two aims, but do not draw attention to the distinction or study its implications.

hypotheses as an ultimate end itself, not for the ultimate aim of testing the start–end hypothesis that *A* caused *E*. Similarly, Hall (2003) states that an important aim of process tracing is to increase the "explanatory power" of one's theory by establishing that *A* caused *B* caused *E*, so this sounds again like testing intermediate hypotheses as an end it itself (deepening understanding); see also Bennett and Checkel (2015).

In fact, some methodologists seem to take process tracing as having both these ultimate aims, even though they don't highlight these aims as distinct, or explore the implications of this distinction.[15] Contrast the citations of Beach and Pedersen in Section 3 with:

> When embarking on a study of the mechanisms linking a cause and outcome, we often already possess cross-case knowledge about patterns of difference-making between the two factors. *The* reason we then trace mechanisms using in-depth case studies is to investigate the "how actually" question, shedding more light on whether there is evidence of a mechanism and thereby also informing us how a cause contributes to co-produce an outcome. (Beach and Pedersen 2016, 41, emphasis mine)

5 Rectangular Datasets and Comparability

This section will examine the suggestion that the type of evidence used in process tracing studies differs from that used in quantitative studies: unlike process tracing, quantitative studies uses "comparable" data arranged into a "rectangular" datasets. I will argue that the best way of understanding this suggestion is that process tracing, unlike quantitative methods, does not rely on an assumption of "unit homogeneity".

Process tracing studies are often said to draw upon qualitative evidence, which is to be contrasted with quantitative evidence. One obvious way of making this idea more precise is to offer the following criterion:

> (F1) A study does not count as a process tracing study if it relies on numerical data, data of the form 'the number of *X*s is . . .' or 'the proportion of *X*s that were *Y*s is . . .'.

A moment's thought shows that this criterion is too restrictive. Exemplary instances of process tracing often rely on numerical data (Brady 2010; Gerring 2006). For example, Wood (2003) uses as evidence the percentage of farmland

[15] See Beach (2017) and Bennett and Checkel (2015) for further examples.

in El Salvador that was put to various purposes, the number of people killed at various stages of the civil war, and the proportion of town majors that resigned during the war.

Instead of criterion *F1*, a more promising suggestion is that process tracing studies handle numerical evidence differently from standard quantitative methods:

> Because each [numerical] observation is quite different from the rest, they do not collectively constitute a sample. Each observation is sampled from a different population. This means that each [numerical] observation is qualitatively different. It is thus the non-comparability of adjacent observations, not the nature of individual observations, that differentiates the process-tracing method from standard research designs. (Gerring 2006, 179)

What does Gerring mean when he suggests that quantitative studies use datasets in which the data is comparable? A common way of interpreting this suggestion is as the suggestion that quantitative studies use rectangular data sets: each dataset is constituted by a number of cases (the rows of the dataset) and by a number of variables (the columns of the dataset) and for each case one observes the value of each of those variables (the cells of the dataset); see Waldner (2012), Collier, Brady, and Seawright (2010), and Rohlfing (2012) for discussion.

But this suggestion cannot be quite right. To see this, consider the phenomenon of missing data. Quantitative studies that use many cases and many variables are often unable to observe the values of some of the cells in the rectangular dataset, so their dataset deviates from this rectangular ideal. But, according to the current suggestion, the more that a study's dataset deviates from this rectangular ideal, the more methodologically similar that study is to process tracing studies, and the less similar it is to standard quantitative studies. But that can't be right, everyone would agree: quantitative studies with missing data are most fruitfully thought of as a sophisticated form of quantitative study, not as something that approaches a qualitative or process tracing method. This illustrates, I suggest, that the use of rectangular datasets by quantitative methods is, in fact, a superficial symptom of a deeper and more interesting epistemic distinction between quantitative studies and process tracing studies.

What might this deeper distinction be? Let's examine the logic of quantitative studies in more detail. Take for example a political scientist who is interested in the extent to which natural resource wealth R and ethnic diversity D contributed to the duration Y of the civil war in El Salvador, for example. Let r^a denote El Salvador's actual resource wealth (2 billion dollars say) and let d^a denote El Salvador's actual ethnic diversity (10 percent say) and let y^a denote the actual duration of the civil war in El Salvador (14 years).

Contrast this actual scenario A with a hypothetical scenario B in which El Salvador's resource wealth is instead r^b (1 billion dollars say) although its ethnic diversity r^b is the same as r^a (10 percent). Let y^b denote what the duration of the civil war in El Salvador would have been, if El Salvador had instead been in this hypothetical scenario B. Most quantitative methodologists agree that $y^a - y^b$ denotes the causal contribution that resource wealth being r^a (2 billion) rather than r^b (1 billion) made to the duration of the civil war in El Salvador. But how to measure $y^a - y^b$?

Consider any hypothetical scenario s you like, in which resources r^s are anything you like, and in which diversity d^s is anything you like. So y^s denotes what the duration of the civil war in El Salvador would have been, if El Salvador had instead been in scenario s. Accordingly $E(Y^s)$ denotes the propensity of the civil war in El Salvador to endure in this scenario s. Metaphorically speaking, if one were to repeat history over and over again a large number of times, each time ensuring that El Salvadorian resources were r^s and diversity were d^s, but allowing other variables to vary, $E(Y^s)$ denotes the average duration of the civil war in El Salvador, averaged across these metaphorical repeats of history.

Now, to estimate causal contribution $y^a - y^b$, quantitative political scientists proceed by thinking about a larger population of cases—for example, all civil wars since 1945. And they assume that in each of these cases the propensity of the civil war in question to endure is a function of resource wealth and ethnic diversity. For example they might assume that, for each civil war post 1945, $E(Y^s) = \beta_1 r^s + \beta_2 d^s$ is true for any hypothetical scenario s. (Here β_1 and β_2 are unknown coefficients that are assumed to be the same for each scenario s and for each civil war post 1945.)

With this key assumption in place, it follows that $E(Y^a - Y^b) = \beta_1(r^a - r^b) = \beta_1(2-1) = \beta_1$. This fact allows one to proceed as follows. First, one collects data on R resource wealth, D ethnic diversity and Y civil war duration, from many civil wars since 1945. Second, one uses standard statistical techniques to measure the values of β_1 and β_2 that "fit" this dataset best, or "explain" this dataset best. Third, since $E(Y^a - Y^b) = \beta_1$, we can now use our estimate of β_1 as an (unbiased) estimate of $y^a - y^b$, the causal contribution that resource wealth being r^a (2 billion) rather than r^b (1 million) made to the duration of the war in El Salvador. This is the econometric approach to causal inference in a nutshell.

Note that our key assumption entails that there is a single mathematical function that describes the propensity of each civil war post 1945 to endure, for any hypothetical scenario s; for example, the function $E(Y^s) = 3r^s + 2d^s$. That is to say, all the civil wars post 1945 exhibit "unit homogeneity", it is assumed (King, Keohane, and Verba 1994, 91–94). To be clear, unit homogeneity does not say that the actual causal contribution $y^a - y^b$ of an extra billion dollars is the same for

El Salvador as it is for Northern Ireland, say. This is because the above mathematical function is a description of a probabilistic average $E(Y^s)$, not of the hypothetical outcome y^s itself. Indeed, unit homogeneity itself does not even require that that the average contribution $E(Y^a - Y^b)$ is the same for El Salvador as it is for Northern Ireland. To see this, note that the mathematical function that satisfies unit homogeneity might be more complicated than the one I gave as an example. It might be $E(Y^s) = (3 + d^s)r^s$ for instance—in which case, those civil wars with greater ethnic diversity would be wars in which 1 billion extra resource wealth makes a greater causal contribution (on average) to the war's duration. To be clear, unit homogeneity is simply the assumption that there is a single mathematical function—and a single specification of the values of the coefficients in the function—that describes the propensity of each civil war post 1945 to endure.

This knowledge of unit homogeneity, across some largish population of cases, is absolutely essential to the reliability of the econometric method (King, Keohane, and Verba 1994). (Of course, there is no need to assume unit homogeneity across all cases: if the US Civil War isn't in one's dataset, for example, or in the population of civil wars that one wants to study, then there is no need to assume that the above mathematical function describes the propensity of the US Civil War to endure.)

Taking stock, these reflections suggest another essential feature of process tracing. If you take the ultimate aim of process tracing to be to identify the start–end causal relationship between factor *A* and factor *E*, then:

> (Criterion Four — Version *A*) A study does not count as process tracing if it requires unit homogeneity for the starting factor *A* and the outcome *E*, that is, the existence of a largish population of cases for which there is a single mathematical function that describes the propensities governing *E*, given variation in *A* and some other variables.

In contrast, if you take the ultimate aim of process tracing to be to identify intermediate relationships in a causal chain, the suggestion is:

> (Criterion Four — Version *B*) A study does not count as process tracing if it requires unit homogeneity for any of the intermediate links in the chain, that is, the existence of a largish population of cases for which there is a single mathematical function that describes the propensities governing factor *D*, for example, given variation in *C* and some other variables.

To understand the difference between these two versions of Criterion Four, imagine a researcher who does the following. She examines a sample of 20 civil wars post 1990, and uses econometric methods to show that *C* caused *D* for civil wars post 1990; she next uses another sample of 20 civil wars in Central America,

and uses econometric methods to show that D caused E for civil wars in Central America. On this basis, she argues that C caused E in civil wars in Central America post 1990. In this case, she has not assumed that the relationship between C to E is homogeneous across a largish population of cases, because there are only a handful of cases that are Central American civil wars post 1990. So this researcher has not violated version A of Criterion Four, although she has violated version B.

Criterion Four also distinguishes process tracing from those quantitative techniques—developed in psychology and imported into political science—that are labelled interchangeably as "mediation analysis" or "path analysis" or "structural equation modeling" (Mackinnon 2007; Tarka 2018).[16] Criterion Four also distinguishes process tracing from more recent quantitative innovations in political science that build on these older mediation-analysis techniques (Imai, Keele, and Tingley 2010; Imai et al. 2011). These techniques are all forms analyzing intermediate links in causal chains, but they do not count as process tracing, according to Criterion Four.

6 Unhelpful Restrictions on Process Tracing

Why have I not suggested the following additional criterion? one might ask. Namely: a study does not count as process tracing study if it needs any feature of the case in question to resemble any other case. This criterion seems to me unhelpfully restrictive. For example, many process tracers make assumptions that the agents they are studying behave in familiar ways. A process tracer might assume that agents will usually tend to hide information that casts them in a good light but be reluctant to volunteer information that casts them in a bad light. As a result a process tracer might credit negative confessions from her informants more strongly than she does positive boasting. But the process tracer's warrant for doing this is that she knows that the informants she is talking to is in this respect similar to most of the other people the process tracer has come to know in life. Therefore, I suggest, it is far too restrictive to debar process tracing from relying on any form of resemblance between her case and other cases.[17] (In contrast, Criterion Four debars process tracing from relying on a particular form of resemblance, namely unit homogeneity, but this leaves open the possibility that process tracing studies might rely on other forms of resemblance.)

[16] This is not to be confused with structural equation modelling within economics. SEM in economics tend to be based on systems of simultaneous equations, and tend not to model latent (unobserved) variables other than of course the error term.

[17] See Crasnow (2012) for a distinct argument to a similar conclusion.

For precisely the same reason, I reject the following criterion: (F2) a study does not count as a process tracing study if it relies on evidence from other cases. After all, my process tracer in the above example is relying on evidence about other people (people not within her case) to support her conclusions. (Sometimes criterion *F2* is endorsed more or less explicitly: process tracing is the use of "observational within-case empirical material left by the workings of a causal mechanism within an actual case to make inferences about the existence of a mechanism in a case" (Beach 2017). Other times criterion *F2* is heavily hinted at.[18])

Why have I not suggested the following additional criterion? one might ask. Namely: (F3) a study does not count as a process tracing study if it needs the probability distribution of some variable to take a particular form. Again, this criterion seems to me unhelpfully restrictive. As I've already noted, some process tracers like to use proportions as evidence—for example, the proportion of women who fought in the militias in El Salvador. But often to measure these proportions one will have to draw sample from a wider population: there are simply too many members of the militia to count them all. But the moment one seeks to use this sample to measure the proportion in a wider population, one will need to make probabilistic assumptions about one's sampling method, if only implicitly and informally.

This also highlights why I think is not strictly true that process tracing cannot use statistical methods, as criterion *F4* says. This is contrary to what Mahoney (2004) and George and Mckeown (1985) claim. For example, Wood's study would still be an exemplary process tracing study, even if she had used formal statistical methods to establish the proportion of women who fought in the militias.

Why have I not suggested the following additional criterion? Namely: (F5) a study does not count as a process tracing study if its results can be generalized. *F5* is false because it says that (having used process tracing to establish the causes of a given civil war) one is never licensed to use this as evidence that the same causes are operating in another civil war. Whether or not one is licensed to extrapolate from one civil war to a second one depends primarily on one's knowledge of how similar these wars are, not on the methods one used to study the first war. Indeed, many process tracers explicitly say that process tracing often tries to generalize or extrapolate. As Hall (2003, 395) puts it, process tracing "is an effort to elaborate and assess the validity of theories capable of explaining a broad class of events or outcomes. It seeks generalizations that are both more simple and more portable than those at which historians typically aim." "As regards the ambition to go beyond the single case, this involves attempts to identify which mechanisms are systematic and non-systematic in the specific case study" (Beach and Pedersen

[18]See Bennett and Elman (2006) and Beach and Pedersen (2016) and Mahoney (2010) and Humphreys and Jacobs (2015) for just some examples.

2016, 310).

In sum, *F1* to *F5* place unhelpful restrictions on what can count as a process tracing study.

7 The Elusive Nature of Process Tracing

Recall criterion Criterion Four Version *B*: process tracing studies do not test intermediate causal hypotheses by methods that assume unit homogeneity, for example by standard quantitative methods. This immediately raises the question: if an intermediate causal hypothesis is not to be tested via standard quantitative methods, then how else is it to be tested?

To explore this issue, this section will examine some further criteria for what counts as process tracing. I will dismiss these putative criteria as empty: they do not offer any substantive, additional constraints over and above Criteria One to Four.

The first putative criterion I have in mind is: (E1) a study only counts as a process tracing study if relies solely on "causal process observations" (Brady 2010). But what are causal process observations supposed to be, one might ask? A promising preliminary clarification is to interpret criterion *E1* as follows: (E1*) a study only counts as a process tracing study if it uses "diagnostic evidence" to uncover the mechanism under examination. This criterion is offered by Collier (2011a) and Collier (2011b) among others. But what might it mean to say of some evidence that it is "diagnostic evidence" for some hypothesis? On one reading, all evidence for a given hypothesis is diagnostic evidence. In this case *E1* just becomes Criterion One and Criterion Two, the idea that process tracing uncovers mechanisms. On another reading, evidence counts as diagnostic evidence for an hypothesis about a given case whenever the evidence itself is a fact about the very same case. For example, the fact that Sophie's skin is yellow is diagnostic evidence that she has liver failure. In contrast, the fact that 90 percent of people with yellow skin have liver failure is evidence that Sophie has liver failure; but it is not diagnostic evidence, one might say, because it is a fact about other people, not about Sophie herself. (This seems to be the view of Beach (2017).) On this reading of the meaning of "diagnostic evidence", however, *E1* just becomes the criterion *F2* that I rejected in the last section.

Another interpretation of criterion *E1* is to re-express it as follows: (E2) a study only counts as a process tracing study if it uses pattern matching. See Mahoney (1999), Mahoney (2004), Bennett and Elman (2006), and Beach (2017) and Goertz and Mahoney (2012) for agreement. Pattern matching is the process of taking an hypothesis—for example the hypothesis that Sophie has liver failure—and asking

how well it can "account for" the patterns one sees in the data? How well can it account for the fact that Sophie has yellow skin, for example?

The problem is that according to most philosophies of scientific inference *all* scientific inference is pattern matching. Hypothetico-deductivism says that all scientific inference is pattern matching, where "accounts for the patterns in the data" is read as "the hypothesis entails the pattern as a matter of deductive logic". Falsificationism is just hypothetico-deductivism with the caveat that hypotheses are never supported by the data, only falsified. Inference to the best explanation says that all scientific inference is pattern matching, where "accounts for the patterns in the data" is read as "the hypothesis, if it were true, would explain the patterns in the data" (Lipton 1991). Bayesian epistemology says that all scientific inference is pattern matching, where "accounts for the patterns in the data" is read as "the scientist in question has a high degree of confidence that these patterns will be occur, supposing that the hypothesis is true" (Howson and Urbach [1989] 2006).

So *E2* does not put any restriction on what counts as process tracing. It fails to tell us what is distinctive about process tracing as a method of causal inference. (The only philosophy of scientific inference that seems difficult to fit with the pattern matching idea is the frequentist philosophy of statistical inference. So my best guess at interpreting *E2* is as a round-about way of endorsing *F4*: a study does not count as a process tracing study if it uses statistical methods. Or perhaps it's a roundabout way of endorsing the more defensible version of *F1*, namely my Criterion Four.)

The third and final putative criterion I have in mind is the following: (E3) process tracing can use a very wide variety of evidence. For example, Bennett and Elman (2007, 183) says that process tracing uses "a wide variety of sources (often including archived documents, contemporary news accounts, secondary histories, biographies or memoirs, and interviews) with due attention to the potential motivated and informational biases of each source". Fairfield and Charman (2017, 368) adds that it often uses "information about timing and sequencing, actors' goals and intentions, and other aspects of causal mechanisms, as obtained from a wide range of sources including interviews, archives, media records, and secondary literature." All this is certainly true, but notice that it does not place any restrictions on what counts as process tracing. There is not a single study for which *E3* suggests that that study doesn't count as process tracing. Instead, *E3* is a kind of meta-criterion, which says that the criteria that we place on what counts as process tracing should not be too restrictive. In particular, the criteria we place on process tracing should allow process tracers to draw upon a wide variety of evidence. So, as a criterion for defining process tracing, *E3* is entirely non-restrictive.

Taking stock, I've argued in this section that criteria *E1*–*E3* are empty: they do not place any substantive constraints on what counts as process tracing, additional

to Criteria One to Four.

8 Conclusion

I conclude that Criteria One to Four are the primary ways of defining process tracing that are suggested by the present methodological literature. Note that I offered two versions of these criteria. Version *A* is *process tracing for the ultimate aim of testing a start–end hypothesis*. For a study to count as this type of process tracing:

> It must identify the presence or absence of intermediate causal links between some factor *A* and an outcome of interest *E*.
>
> And, perhaps, it must describe each of these intermediate links in terms of entities engaging in regular, well-understood activities.
>
> It must ultimately aim to test the hypothesis that the starting factor in the chain was a cause of the final outcome. And the means by which it pursues this ultimate aim must be the following logical structure: (I) the overall evidence e bears on hypothesis h only by first bearing on the intermediate hypotheses; (II) for each intermediate hypothesis, there is a piece of evidence e' that bears on this intermediate hypothesis alone.
>
> It must not require unit homogeneity for the starting factor *A* and the outcome *E*, that is, the existence of a largish population of cases for which there is a single mathematical function that describes the propensities governing *E*, given variation in *A* and some other variables.

Version *B* is *process tracing as an ultimate end in itself*. For a study to count as this type of process tracing:

> It must identify the presence or absence of intermediate causal links between some factor *A* and an outcome of interest *E*.
>
> And, perhaps, it must describe each of these intermediate links in terms of entities engaging in regular, well-understood activities.
>
> Its ultimate aim must be this identification / description of the intermediate links in a causal chain.
>
> It must not require unit homogeneity for any of these intermediate links in the chain, that is, the existence of a largish population of cases for which there is a single mathematical function that describes the propensities governing factor *D*, for example, given variation in *C* and some other variables.

As it stands, process tracing as an ultimate end in itself is under-theorized. Given what I argued in Section 7, the methodological literature has barely begun

to answer the question: if an intermediate causal hypothesis is not to be tested via methods that assume unit homogeneity, then how else is it to be tested? See Runhardt (2015) for a similar worry. Some methodologists seemed to answer this question by claiming that intermediate hypotheses could be directly observed to be true (Brady 2010). But I think it's fair to say that most methodologists now reject this claim (Beck 2006; Crasnow 2012).

How then to proceed? In my view, what we urgently need is a taxonomy of various different ways in which process tracers are supposedly able to test intermediate causal hypotheses without relying on unit homogeneity. This taxonomy would allow us to evaluate process tracers' claims that process tracing is good at eliminating rival causal hypotheses, that process tracing is not an inherently micro enterprise, and that process tracing really is its own method, distinct from quantitative methods.

8 References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven N. Durlauf, IA:385–472. Amsterdam: North Holland.

Beach, Derek. 2017. "Process-Tracing Methods in Social Science." https://doi.org/10.1093/acrefore/9780190228637.013.176.

Beach, Derek, and Rasmus Pedersen. 2016. *Causal Case Study Methods: Foundations and Guidelines for Comparing,matching,and Tracing*. Ann Arbor MI: University of Michigan Press. https://doi.org/10.3998/mpub.6576809.

Beck, Nathaniel. 2006. "Is Causal-Process Observation an Oxymoron?" *Political Analysis* 14: 347–52. https://doi.org/10.1093/pan/mpj015.

Bennett, Andrew. 2008. "The Mother of All "Isms": Organizing Political Science Around Causalmechanisms." In *Revitalizing Causality: Realism About Causality in Philosophy Andsocialscience*, edited by Ruth Groff, 205–19. London: Routledge.

———. 2010. "Process Tracing and Casual Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 207–20. Lanham MD: Rowman; Littlefield.

Bennett, Andrew, and Jeffrey T. Checkel. 2015. "Process Tracing: From Philosophical Roots to Best Practices." In *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel, 3–38. Cambridge: Cambridge University Press.

Bennett, Andrew, and Colin Elman. 2006. "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science* 9: 455–76. https://doi.org/10.1146/annurev.polisci.8.082103.104918.

———. 2007. "Case Study Methods in the International Relations Subfield." *Comparative Political Studies* 40: 170–95. https://doi.org/10.1177/0010414006296346.

Brady, Henry E. 2010. "Data-Set Observations Versus Causal-Process Observations: The 2000USpresidential Election." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 237–42. Lanham MD: Rowman; Littlefield.

Checkel, Jeffrey T. 2008. "Process Tracing." In *Qualitative Methods in International Relations: A Pluralist Guide*, edited by Audie Klotz and Deepa Prakash, 114–30. New York: Palgrave Macmillan.

———. 2013. "Transnational Dynamics of Civil War." In *Transnational Dynamics of Civil War*, edited by Jeffrey T. Checkel, 3–30. Cambridge: Cambridge University Press.

Collier, David. 2011a. "Teaching Process Tracing: Examples and Exercises." http://ssrn.com/abstract=1944646.

———. 2011b. "Understanding Process Tracing." *Ps: Political Science and Politics* 44: 823–30. https://doi.org/10.1017/s1049096511001429.

Collier, David, Henry E. Brady, and Jason Seawright. 2010. "Sources of Leverage in Casual Interference: Toward an Alternativeviewof Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 161–200. Lanham MD: Rowman; Littlefield.

Crasnow, Sharon. 2012. "The Role of Case Study Research in Political Science: Evidence Forcausalclaims." *Philosophy of Science* 79: 655–66. https://doi.org/10.1086/667869.

Evangelista, Matthew. 2015. "Explaining the Cold War's End: Process Tracing All the Way down?" In *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel, 153–85. Cambridge: Cambridge University Press.

Fairfield, Tasha, and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities,andcaveats." *Political Analysis* 25: 363–80. https://doi.org/10.1017/pan.2017.14.

George, Alexander, and Timothy J. Mckeown. 1985. "Case Studies and Theories of Organizational Decision Making." In *Advances in Information Processing in Organizations*, edited by Robert F. Coulam and Richard A. Smith, 2:21–58. Greenwich CT: JAI Press.

Gerring, John. 2006. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.

Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research Inthesocial Sciences*. Princeton NJ: Princeton University Press.

Hall, Peter A. 2003. "Aligning Ontology and Methodology in Comparative Poli-

tics." In *Comparative Historical Analysis in the Social Sciences*, edited by James Mahoney and Dietrich Rueschemeyer, 373–406. Cambridge: Cambridge University Press.

Hedstrom, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49–67. https://doi.org/10.1146/annurev.soc.012809.102632.

Hitchcock, Christopher. 2001. "A Tale of Two Effects." *Philosophical Review* 110: 361–96.

Howson, Colin, and Peter Urbach. (1989) 2006. *Scientific Reasoning: The Bayesian Approach*. 3rd ed. La Salle IL: Open Court.

Humphreys, Macartan, and Alan M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109: 653–73. https://doi.org/10.1017/s0003055415000453.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15: 309–34. https://doi.org/10.1037/a0020761.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning About Causal Mechanisms-fromexperimental and Observational Studies." *American Political Science Review* 105: 765–89. https://doi.org/10.1017/s0003055411000414.

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

Lipton, Peter. 1991. *Inference to the Best Explanation*. London: Routledge.

Lott, John R. 2005. "The Impact of Early Media Election Calls on Republican Voting RatesinFlorida's Western Panhandle Counties in 2000." *Public Choice* 123: 349–61. https://doi.org/10.1007/s11127-005-7166-1.

Mackinnon, David P. 2007. *Introduction to Statistical Mediation Analysis*. New York: Routledge.

Mahoney, James. 1999. "Nominal, Ordinal, and Narrative Appraisal in Macrocausal Analysis." *American Journal of Sociology* 104: 1154–96. https://doi.org/10.1086/210139.

———. 2004. "Comparative–Historical Methodology." *Annual Review of Sociology* 30: 81–101.

———. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62: 120–47. https://doi.org/10.1017/s0043887109990220.

———. 2015. "Process Tracing and Historical Explanation." *Security Studies* 24: 200–218. https://doi.org/10.1080/09636412.2015.1036610.

Mahoney, James, and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitativeresearch." *Political Analysis* 14: 227–49. https:

//doi.org/10.1093/pan/mpj017.
Rohlfing, Ingo. 2012. *Case Studies and Causal Inference*. London: Palgrave Macmillan. https://doi.org/10.1057/9781137271327.
Runhardt, Rosa W. 2015. "Evidence for Causal Mechanisms in Social Science: RecommendationsfromWoodward's Manipulability Theory of Causation." *Philosophy of Science* 82: 1296–1307. https://doi.org/10.1086/683679.
Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France,Russia,and China*. Cambridge: Cambridge University Press.
Tannenwald, Nina. 1999. "The Nuclear Taboo: The United States and the Normative Basisof Nuclearnon-Use." *International Organization* 53: 433–68. https://doi.org/10.1162/002081899550959.
Tarka, Piotr. 2018. "An Overview of Structural Equation Modeling: Its Beginnings, Historicaldevelopment,usefulness and Controversies in the Social Sciences." *Quality and Quantity: International Journal of Methodology* 52: 313–54. https://doi.org/10.1007/s11135-017-0469-8.
Waldner, David. 2012. "Process Tracing and Causal Mechanisms." In *The Oxford Handbook of Philosophy of Social Science*, edited by Harold Kincaid, 65–84. Oxford: Oxford University Press.
Wood, Elisabeth Jean. 2003. *Insurgent Collective Action and Civil War in El Salvador*. Cambridge: Cambridge University Press.