Process Tracing

Keith Dowding

For

Harold Kincaid and Jeroen Vanbouwel

Philosophy of Political Science

Oxford University Press

1. Introduction

Alexander George (1979) introduced the term 'process tracing' into political science from cognitive psychology. Methodologically it has developed in various directions (see, for example, Bennett and Checkel 2015a). In its broadest sense, process tracing is systematic qualitative analysis using various sources of information in order to narrate a sequence of events leading to some outcome of interest.[1] In a narrower sense, it involves specific recommendations for how to go about detailed case studies, within-case, and historical analysis in a systematic and theorized manner for methodologically disciplined research. Process tracing is now widely used for qualitative case-study research in political science and international relations, and has accumulated a large and fruitful literature on qualitative historical and case-study methods (George and Bennett 2005; Bennett and Checkel 2015b).

When using process tracing in empirical work, two approaches are recommended. One is inductive, and largely directed at explaining how a given outcome came about – such accounts can be theory generating (George and Bennett 2005; Bennett 2008). The other is

[1] I use the term 'outcome' or sometimes 'E' (for effect) in a broad sense to stand for some event or institution.

usually described as deductive and is thought to be theory testing. Either can also be used to generate claims or hypotheses about turning points, decisive moments, and critical junctures in history. What is expected of excellent qualitative research is not in dispute; process tracing is largely about proclaiming such qualitative research as scientific endeavour. Faced with the theory-testing claims of quantitative research, qualitative researchers have argued not only that their methods can also test theories and discover causation, but are indeed superior.

I do not examine in this chapter the various recommendations of process tracing for qualitative researchers (see Collier 2011 or Beach and Pedersen 2016 for detailed discussions), though I will briefly discuss some techniques in Section 2. Rather, I will concentrate upon some of the philosophical claims made on its behalf with regard to causal analysis and theory testing. This discussion ought to affect what qualitative researchers try to do, but will not change the practical techniques recommended by process-tracing experts.

The issues I cover in this chapter include causality and causal mechanisms, theory testing, generalizations versus mechanisms, distal and proximate causation, and the appropriate level of analysis.

2. Systematic Qualitative Methodology

Case studies provide detailed analyses of the course of events. They naturally form a narrative, and every narrative takes on a causal form in the sense that it drives towards a conclusion. In any narrative what we choose to report will be based upon what we think is important in that case study; and what we think important is bound to be what we think causes, in some sense, the outcomes that make that case study of interest. Process tracing is an attempt to systematize a narrative in order to provide stronger evidence of the important causal features.

Underlying the defence of process tracing is an account of the explanation of social outcomes in terms of causal mechanisms usually seen in opposition to generalizations that are

often the focus of high-*n* studies. It is clear, however, that whilst invariant (or law-like) generalizations or laws can perform explanatory roles, empirical generalizations are the outcomes that need to be explained. Explanation using mechanisms is not in conflict with the search for empirical generalizations. Quite the opposite: we should expect mechanisms to explain the empirical generalizations we find, and even to explain those examples that do not fit the generalization. I have argued elsewhere that empirical generalizations are produced by mechanisms which themselves are underlain by invariant generalizations (Dowding 2016; see also Waldner 2012). Nevertheless, an important element of social explanation is understanding the mechanisms that lead to types of outcomes. I have also argued elsewhere that single case studies cannot test – as 'testing' is normally understood – theories seen as either generalizations or mechanisms (although, of course, they can provide evidence that goes into a broader set of evidence that constitutes a test as ordinarily understood). Eddington's famous demonstration of relativity during a solar eclipse persuaded nobody who did not already hold Einstein's theory (Dowding 2016, pp. 113-14). Even 'crucial case studies' generally fail to establish that theories are false. What they can do is provide evidence as to whether a given mechanism *applies* to a specific case and fill in some of the details of that mechanism as it applies to that case (Dowding 2020). So they can show claims about the application of a given theory or mechanism to a specific case is false, but that does not show that the mechanism or theory is not true in other cases

Whilst there might be various non-rival approaches to causation in science, some have argued that process tracing has an *ontology* and understanding of causation different from those employed by scholars who operate with high-*n* studies (Beach and Pedersen 2016, for example). If so, that would be unfortunate if social science is to be one science rather than forever a set of competing paradigms. Here I defend process tracing as a systematic way of examining causal mechanisms, making direct analogies with the understanding of causation in experimental and

high-*n* statistical approaches, to explore both its strengths and weaknesses. To understand those strengths and weaknesses, we need to see that low-*n* qualitative work and high-*n* quantitative work are directed at different sorts of research questions posed at different levels of generality. The latter is directed at uncovering the ultimate or structural causes of *types*; the former aims to explain the proximate or historical causes of *tokens*.[2] A type is a class of objects composed of tokens and a token is an example of a type. In either case we can model the mechanism that is thought to be the causally important explanation. The idea is that indeed process tracing looks at different causal processes, but in a non-rival manner.

Process tracing is generally tied to testing hypotheses derived from theories. The recommendation is that a clear theory of some process is proposed and evidence collected to defend that theory, often in comparison with rival explanations. The single case can be used as a test of the theory, but is often used as a paradigm example of the proposed mechanism being operant in other examples. Researchers look for causal inferences from the evidence collected about a single case. Producing evidence for or against any hypothesis will involve claims about the counterfactuals that enable us to make the causal inferences.

Process tracing seems to be defined by the four tests specified in Van Evera (1997), developed by writers such Bennett (2010), Collier (2011) and Waldner (2012), with a further completeness test suggested in Waldner (2015). The tests are classified according to whether passing them provides necessary and/or sufficient conditions for accepting the causal inference.

[2] I use 'structural' and 'ultimate' interchangeably. In Dowding (2016) I adopted the term 'ultimate' from biology, as 'structural' can have rather different meanings in social science, and I thought 'ultimate' came with less verbal baggage. Then about the time my book came out I read Kitcher (2003, ch. 4, loc 2213), who recommends to biologists that they ought to use the term 'structural' rather than 'ultimate', since that latter term seems to downgrade the worth of 'proximate explanation' or what he prefers to call 'historical explanation'! What really matters is the distinction, not how we label it.

Collier (2011) gives the most complete and simple explanation of Van Evera's four tests, using the Sherlock Holmes story *Silver Blaise* to illustrate them. As Collier and many others point out, though, in real examples it is much harder to specify simply what form evidence actually takes in terms of the tests.

Briefly, the tests are as follows. *Straw in the wind* tests are neither distinctive nor precise and do not supply strong evidence for or against competing hypotheses; they merely provide evidence that is consistent with a hypothesis. *Hoop* tests use evidence that is precise, but not distinctive.[3] Failing a hoop test means a given hypothesis is wrong, so they are used to exclude potential explanations, but do not provide strong evidence for a given hypothesis. A standard example is the ability of a murder suspect to provide an alibi. If the suspect has an alibi, then the test is failed; having no alibi means they are still suspect but does not confirm guilt. *Smoking gun* tests provide evidence distinctive to the case and thus support the hypothesis; however, failing the test does not really undermine the explanation. A suspect having a smoking gun in their possession following a shooting seems strong evidence they are guilty; but not having the gun does not prove they are not guilty. *Doubly decisive* tests provide evidence that is precise and distinctive to the case – supplying direct evidence for a given claim – and thus can either support or undermine alternative explanations. Cabinet minutes, for example, provide very strong evidence of what a government thinks it was doing, and sometimes why, with some controversial measure.

[3] Van Evera (1997, p. 31) uses the term 'certain' rather than precise, but means by that 'unequivocal forecast'. I prefer 'precise', since 'certain' could be taken to mean 'determined'. He also uses the term 'unique' where I use 'distinctive'. He says 'A *unique* prediction is a forecast not made by other known theories. The more unique the prediction, the stronger the test' (Van Evera 1997, p. 31). I prefer to use the term 'distinctive' since below, with regard to case studies, I discuss the nature of uniqueness in a rather different sense.

These tests all seem rather clear when we have simple examples. However, the social sciences offer few real examples where the tests are so clearly laid out. Bennett (2008) uses two examples, the Fashoda crisis and Heinz Goemans's (2000) research on Germany and the end of the First World War. But even his careful article does not precisely lay out the arguments of the research at issue in terms of the four tests. The precise relationship between the evidence he discusses and the four tests is left to the reader to work out. Whilst the process-tracing tests are an interesting justification for the sorts of evidence that we use when making inferences in case-study research, in practical terms, qualitative researchers narrate their arguments presenting the evidence they see fitting their findings. Indeed Van Evera (1997, p. 32) suggests his four tests are the types of evidence we usually work with in historical analysis, but does not argue that we need to use them systematically. Later writers have suggested that qualitative researchers need to think more systematically about the nature of their evidence and establish it in the context of the four tests in order to argue their case.

We can add to the four tests Waldner's (2015) 'completeness standard'. The completeness standard says process tracing gives a causal explanation when it can provide (a) a causal graph whose nodes are jointly sufficient for some outcome and (b) an event-history map establishing a correspondence between the nodes and the events, together with (c) a theory about the causal mechanisms, in which (d) rival explanations are discounted by failing (a)–(c). Waldner's completeness standard is a very demanding test, which probably over-formalizes the nature of descriptive narrative history.

Some have represented the nature of additional qualitative evidence in Bayesian terms (Bennett 2008, 2015; Beach and Pedersen 2016, pp. 83–99; Humphreys and Jacobs 2015). The general idea of the Bayesian representation of process tracing is that sometimes thick, highly granular description provides more confidence in updating our prior beliefs about some causal process than simply adding an extra example to some high-*n* regression. That is correct.

However, without some systematic method of measuring the difference that different types of evidence provide, the Bayesian justification is simply a metaphorical way of defending process tracing (Fairfield and Charman 2017, 2019; Gerring 2017). It does not offer a way of actually quantifying case-study evidence in probabilistic terms.

3. Mechanisms and Causation

Process tracers have emphasized that their approach looks for evidence of mechanisms, initially proposing it as an alternative form of explanation to the search for generalizations (George and Bennet 2005). At times the advantages of process tracing have been highlighted by its ability to find intervening variables between some causal factor C and some outcome E. The general idea seems to be that the precise details of the mechanism (or process) is contained in these intervening variables. Now, of course, if some factor C always results in E, then C can be said to cause E. The fact that there might be several routes from C to E does not impugn the causal claim (unless the claim is that there is only one route, or only one mechanism). Showing in a given example that one rather than another route was present shows only that one mechanism was present in this case rather than another. Process tracing can test which of several different mechanisms operate in any given case. A single case study can only show which mechanism operates in that case; one would have to study every extant case to see whether that mechanism is always the causal process. One cannot test across rival mechanisms with one case study; one can only see which mechanism is present for the case under study (Dowding 2020).

It is this type of analysis that leads some process tracers to claim that qualitative methods operate with a different ontology and understanding of causation from quantitative ones. Beach and Pedersen (2016) suggest that process tracers are interested in 'actual causation' rather than counterfactual accounts of causation. What they seem to mean is that process tracers are interested in the actual conditions of a particular case, rather than the conditions that cause

outcomes of that type. In my terms, they track proximate causes of tokens rather than structural (or ultimate) causes of types.[4] Beach and Pedersen (2016) also suggest that process tracers are interested in the arrows in diagrams of causation (c → e), rather than the cause (C), which is what quantitative writers are concerned with.[5] George and Bennett (2005) had likewise suggested process tracers are interested in the detailed stages between some purported cause and outcome. I have termed this 'bump-bump' causation (Dowding 2016). There are often many bumps between the outcome and the initial causes – the fourth car in line hits the third, which hits the second, which hits the first. But, if we want to know why the first car is damaged, it is the speed of the fourth that is relevant, even though the damage to the first is mediated by that caused to all the others. It is true we might well be interested in the intervening cars and how they acted and were damaged (especially if one of them was ours). However, this is not a rival account of causation, nor does it impugn any claim about the initial cause. It just adds more detail to a specific case.

George and Bennett suggest the same initial conditions and final outcome can have different processes between them, and social science is interested in that process. We are indeed often interested in the details. 'Process' here is taken to mean 'mechanism' – and I have been using that term in that way in the two paragraphs above. And what process tracers claim is that different mechanisms can lead to the same outcome. They are surely right. However, there is some ambiguity in the term 'mechanism' here. George and Bennett (2005) mean by 'mechanism' the different details of the story, using the idea of equifinality as many different paths to the same outcome. So different processes (mechanisms) can lead to the same outcome.

[4] The type/token and the structural (or ultimate)/proximate distinctions are not identical. One can provide proximate accounts of types through their tokens, whilst structure always backgrounds proximate explanation of tokens. Cases are always invoked under a description, and no descriptions can ever avoid general categories. I am making a distinction here to explain the different level at which questions are asked and the sorts of answers different methods deliver, but the distinction can be pushed too far. Another way thinking of type and tokens is in terms of levels of explanation. A type can be a token of a higher-level type, and token a type for lower level tokens.

[5] I'm using lower case 'c' and 'e' for tokens and upper case 'C' and 'E' for type here.

However, equifinality is usually thought of as system predictability – a system whose behaviour is predictable from more than one preceding system (or starting point). That predictability is seen in terms of one mechanism – natural selection, for example. The mechanism is the structural conditions that lead to the same type of outcome. Hereafter, I will use the term 'process' for description of the proximate causes of token outcomes, and 'mechanism' for what structures those processes to the type outcome.[6] MOVE

Also the fact that different routes can be determined by some overarching mechanism (eg natural selection) or might be contingent - there are different possible ways in which one could get from a to z, but there is no mechanism that applies to all them.

All we need to understand by this ambiguity over the term 'mechanism' is that if we are interested in high-granularity accounts of a token case, then our research questions are different from those who are attempting to explain types of outcome. Proximate token explanation is not the same as structural type explanation. They are not necessarily rival. Case studies can tell us the details of different processes, and might tell us whether one purported type mechanism or another operates in a given case, but they cannot test whether or not either mechanism describes any case at all. EG PEASANT REVOLT

I have not formally defined what a mechanism is. There are numerous competing definitions (for a review, see Hedstrom and Ylikoski 2010 or Beach and Pedersen 2016). We can think of a model as something that describes a mechanism. Woodward (2003) provides the

[6] The dictionary definitions of process and mechanism are very similar. According to the Oxford English Dictionary, the most common use of 'process' is 'A continuous and regular action or succession of actions occurring or performed in a definite manner, and having a particular result or outcome; a sustained operation or series of operations', whilst a mechanism is 'a system of mutually adapted parts working together in a machine or in a manner analogous to that of a machine'.

most complete account of this, summarized here by Hedstrom and Ylikoski (2010, Table 1, p. 51):

> A model of a mechanism (a) describes an organized or structured set of parts or components, where (b) the behavior of each component is described by a generalization that is invariant under interventions, and where (c) the generalizations governing each component are also independently changeable, and where (d) the representation allows us to see how, by virtue of (a), (b), and (c), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves

So we can this that the claim that invariant generalizations underlie the counterfactual effects of manipulation. But an important aspect of a mechanism or a process is that it is composed of entities and activities – the latter being the things that entities do (Craver 2006, p. 371). We can define activities, at least in part, by the manipulability of the variables – that is, we can alter the value of one variable in the description of the mechanism by manipulating another (Woodward 2003; Pearl 2009). Process tracing tries to systematically study a given historical process in order to identify the important aspects or variables in a process, to allow us to isolate the important components that lead to some social outcome. However, we must bear in mind that we can have different intervening variables – different processes – across different systems, that are still processes of the same mechanism at higher level of generality.

History is a narrative. What we choose to include in the narrative is, generally speaking, what we think important for the narrative structure. There might be asides that include interesting facts or snippets of information. There might be elements where we make normative comments on the action or, especially in books, set up background conditions for another part of the story. Generally speaking, however, what we think is important for the narrative structure is what we think is important in the causal story for the outcome that we are narrating. One

problem with having a hypothesis about a particular case is that it creates 'the honest detective's problem' (Dowding 2017a). Once the prime suspect has been identified, it is efficient for the police to look for evidence that would convict that suspect. That means evidence that might convict someone else is less likely to be discovered. This is a form of bias in all case-study research. Examining rival hypotheses can help mitigate this bias – we examine two suspects – though there is still the danger that as one becomes front-runner, evidence in its favour is emphasized.

Causation is much discussed, and here is not the place to discuss different accounts of causation in any detail. A very general account of causation, and that which most historians seem to have in mind, is a 'but for' account. 'But for' accounts are elucidated in detail in various forms of necessary and sufficient conditions, such as INUS (insufficient but non-redundant components of unnecessary but sufficient condition) (Mackie 1974) or NESS tests (necessary element of a sufficient set of conditions) (Wright 1985, 1988). Described at some level of granularity, no INUS condition is either individually necessary or individually sufficient for an outcome, but each is a non-redundant element of the sufficient set for that outcome. Any set of conditions sufficient for an outcome is composed of either necessary or INUS conditions. A NESS test looks for the necessary element of a sufficient set of conditions for the outcome. Both of these characterizations focus upon each element of a cause. When we analyse any outcome, however, we treat all these INUS or NESS conditions equally. Some we background, and some we concentrate on, for our explanation, as the cause. In a historical narrative, the choices we make in terms of what we background and what we foreground determine what we consider to be the important causal story in the process that we are addressing.

4. Analogy with Experiments

If we do not accept that process tracing requires a different notion of causation, we can discuss it in direct relation to those approaches which provide the cleanest way of making causal inferences. That is, experiments where we can manipulate the variables – in experimental terminology, provide interventions – to directly measure causal effects.[7]

We know that for any outcome what happened earlier was sufficient for that outcome to occur. However, many of those conditions were not necessary for the outcome we are studying. Some are necessary for that outcome, but also necessary for other outcomes that we wish to distinguish from the outcome we are studying. Elections are necessary for any party to win an election, but we are not interested in 'the election being called' for why the Labor Party rather than the Liberal Party wins a given election; rather, we are interested in the conditions that led to that victory. To be sure, choosing the date might sometimes be important (if they had called it earlier the governing party might not have lost, for example). However, the fact of elections, whilst necessary, is irrelevant to our consideration of who wins. We thus ignore necessary conditions that are of too high a level of generality for the outcome as denoted, as well as those not necessary for our outcome. Social science then concentrates upon the important elements of those that are necessary for the outcome specified by our research question. In other words, our specific research question denotes both the factors that are important to the explanation we offer and, importantly, the level of analysis.

Process tracing is not an alternative to high-*n* studies because a case study cannot answer the sorts of questions addressed by high-*n* quantitative analysis, addressed at explaining types of phenomena (Dowding 2016). Case studies can only examine whether a given mechanism applies to that case or trace the proximate elements (the specific process) in the type mechanism (Dowding 2020). But process tracing can offer superior explanations of token phenomena, as it

[7] Gerring and McDermott (2007) draw analogies between case studies and experiments, but in a much more general manner than I do here.

answers questions that address that phenomenon in much more detail. However, even at the finest-grained level of explanation, we still need to choose the important elements that enter our proffered explanation of an event or institution.

But how do we choose the 'important elements'? We can think about them in terms of the stability of the elements that enter into the explanation. Some elements will be more robust or stable across changing context than others, and we tend to be interested in these. First, we ignore background circumstances – that is, those conditions not explicitly represented in the c–e relationship we are studying. In experimental conditions, for example, c is the intervention and e the outcome we observe, and the experimental set-up can be considered as the background circumstances held constant during the experiment. Outside of the laboratory, we have no sharp distinction between the background conditions and the intervention. However, for any analysis we have to make some assumptions about what we are considering background conditions and what we are thinking of as the causes (interventions) that create the outcomes we observe. Sometimes debates in social science revolve around whether a researcher has made the right choices here. Such debates might concern the correct level of analysis for a specific outcome or the stability of the background itself.

The correct level of analysis is concerned with whether or not we are attempting to explain some general phenomenon – the type – or whether we are examining the precise process in a given (token) example. Tokens are examples of a given type; importantly, though, not only will all tokens *not* share all of their features with each other, there need be *no* feature that is distinctive of a type that is common to all token examples (Wetzel 2009; Dowding 2016). Sheep can be born with only three legs, while typically genotypes vary more than phenotypes. We might think democracy requires some form of voting system, though some deliberative democrats argue that voting besmirches deliberation that should reach agreement without the need for votes. (Something that many British prime ministers claimed to achieve in their

cabinets.) The fact that we can demonstrate for a particular case study that some key decisions made by some agents were important to the precise outcomes of that case does not mean that more general claims made about cases of that type are not true of the type including that token example. Claims about the type which are foregrounded (seen as the important causal features) might be only background conditions of the token when considered in detail. Descriptions of tokens are much more detailed, of a higher granularity, than of types. And it is the detail we are interested in that distinguishes our case study example from others of that type.

After all, predictions can be about types or tokens. A prediction about a token item will give us our expectations about what we will see *in that case*. It will establish the probability of what we expect to see in that case. But if there are many cases, then it will tell us what we expect in each of those cases and, if it is probabilistic, the distribution of expected outcomes over those cases. We might examine a token in order to see what we can learn about the type, and how it transfers to other tokens within that type. The claim made about highly detailed studies of such tokens is that this evidence is of higher quality and should be given more weight in our considerations. However, it might have more weight in consideration of the proximate cause of that particular example, but it does not follow that it is true of the type, unless it can be shown that all members of the type share the same feature with that case study. Keeping in mind the level of the analysis, and how evidence bears in higher levels (less granular descriptions of types), can allow us to see that some claims in social science are not incompatible as they are sometimes thought to be. CASE STUDIES EFFECT IS SO WEAK NEED LARGE N

Reflecting on the stability of the background conditions, consider again interventions in the laboratory. We often make the distinction between the internal and the external validity of an experimental set-up. An experiment is said to be internally invalid if a poor experimental arrangement impugns causal inference. External validity concerns the generalizability of findings. Given that an experiment is internally valid, how far do the results apply more broadly

outside the experimental setting? A given internally valid experimental result is externally invalid at least where (1) the outcome is unstable with any slight change in the background conditions of the experiment and/or (2) when those conditions rarely hold precisely outside the laboratory. With any given case study, the precise conditions only apply to the case itself, so the second condition is not important for claims made about the proximate causes in that particular case (unless, of course, claims are extended from that case to other cases of that type). However, the first condition, the stability of the outcome regardless of varying background conditions, is important to any claim about the specific causal claims of that case.

The importance of any causal claim, C, varies with the stability of E, with regard to changes in the background conditions. The narrower the set of changes in the background conditions and the degree of their contingency will condition E's stability. Thus, when we are making causal claims in a case study, we are claiming that we have identified a cause c, where the counterfactual dependence of e on c is stable within a range of background circumstances that differ from the specific circumstances of c. Often when debate rages over some claim to have identified some c, the critics are pointing out that c's relationship to e is unstable with regard to some other factors c^{*}, c^{**} and c^{***} they have identified. Without wanting to dismiss such debates, we can point out that they address the importance of the c's *relative to one another*. They are not questioning the causal implication of C with regard to E; rather, they are querying its importance in this case. And that importance can be represented by claims about the stability of the background conditions. After all, saying that a given background condition only needs to vary slightly in order to reach a different outcome is to make the claim that it is an important causal condition. This means, slightly paradoxically, that a given causal claim for a specific case study might not have external validity for the very case that is being studied.[8]

[8] The originator of the internal/external validity distinction of experiments, Donald Campbell (in Campbell and Stanley 1963), later regretted making the distinction in the manner he did as he felt it takes on too much

Now the C factors we are interested in when considering a given case are ones that concern the proximate causes of an event. General theories, including those couched in the form of mechanisms, concern the ultimate or structural causes of outcomes. They concern what cause outcomes of this type, in the knowledge that the precise working through of the mechanism *might* have varying proximate causes. That is what is entailed by the idea of equifinality. Equifinality suggests that, for a given causal process, from different initial conditions, the same type of outcome will occur (Bertalanffy 1968). The precise elements – the bump-bump causal process – will vary across different token examples of the type, but the structure of the mechanism ensures that the same type of outcomes will accrue. Of course, at finer-grained description, that same type of outcome will have different characteristics – that is, different processes – but at the coarser description they are of the same type. Questions about case studies are often interested in the differences at the finer level of description, and for that reason they are interested in the proximate, bump-bump, causal process. But even here, however, we are still interested in the stability of the background conditions when we are describing what is important in proximate causal process. We have to pick out what we think are the important elements in our historical description; we do not describe everything.

REVOLTUTION EXAMPLE WOULD HELP

So historical process tracing is an explanation given at the token level to provide a proximate causal explanation of a given case. The mechanics of process tracing as a methodology are Van Evera's four tests, designed to demonstrate how strong our causal inferences are, given the evidence we procure. They are designed to explain the unique case we are studying and to highlight these stability issues. However, we also need to see that

importance in many discussions of causal inference from experimental result (Campbell 1986); this is my version of the reason. We make assumptions about the stability of backgrounds, both in the experimental set-up and in application. My slightly paradoxical way of setting up the issue brings that out.

uniqueness can mean different things with regard to any case study. For example, some people claim that all human events are unique and hence we cannot explain them as we do natural phenomena. If by that they mean that every social outcome has its own proximate (bump-bump) causal process, they are right: each such event is individuated differently. But then, so is every individuated event in natural processes. All events can be uniquely individuated by space–time coordinates (Fetzer 1975; Tucker 1998, p. 62). So, if they mean that this trivial truth about individuation entails that we can give no structural (or ultimate) causal story of types of which this example is a token, they are simply wrong. They are confusing claims made at different levels of analysis. However, they might mean that human events are unique in a stronger sense. For some events the claim might be that there are no actual individuated events of that type other than the one under consideration.

What does uniqueness in this sense mean? And what does it entail for our explanatory claims? If a case study is of a 'unique' event in this second sense, then the proximate explanation will also be unique. No other actual events are like this one. However, it is does not follow that there is no structural or ultimate explanation of the type of event of which this token constitutes an example. We can still consider such unique events to be tokens of a given class, where the other tokens in the class are not actual. Here the class would be given in terms of the stability of the background conditions to the specific C that is being investigated with regard to the purported explanation of E.

Of course, we might be claiming that the important element of some outcome E is that that specific outcome is highly unlikely and only came about because of the convergence of an (actually) unique set of precursors. If we could play that event over and over again, the actual outcome would rarely come about. Imagine repeating a specific scenario time and again in some computer simulation, especially where we do not fully understand the process or mechanism going on because of the complexity of the model. (Not 'fully understand' here

means we cannot mathematically model it at the granularity required to produce stable predictions.) Some of the interactions in the model at the limits of appropriate granularity are probabilistic, such that the same outcome will not occur if the model was played time and again. At best, we can give some probability distributions over given outcomes. Having done so, we find that the actual outcome is highly unlikely – say it comes about only 1/10,000 runs. All we can say is that this outcome is highly unlikely. What we might claim, following a very detailed examination of the case, is that specific, highly risky decisions, made by specific actors, led to the outcome.

Social scientists generally try to provide type-level explanations that are highly likely. They look for empirical generalizations and they theorize mechanisms to explain those generalizations. One form of historical explanation is to look for the unlikely. Some historians are specifically interested in individuals who make a difference – in Napoleon, Churchill, Stalin. These people might be special and therefore important in any causal story. Indeed, it seems that strong leaders are often those who take risks early in their careers, often against the interests of their core supporters (Dowding 2017b, ch. 8). The very fact that early on they took risky decisions that cemented their positions demonstrates the unlikeliness of the process described. The counterfactual question often takes such individuals out of the picture and asks how history would otherwise be. Those who see history as determined by social and economic structures tend to think history would not be so different without these characters; others, especially those interested in the details, believe such individuals are key elements in historical change. Historians tend to be interested in detail – high granularity of the token – whilst social scientists tend to be interested in the general form – low granularity of the type. These are two different forms of analysis, asking different questions, and both are important in their own right.

One danger of detailed story telling that defends some general mechanism at play is that we end up telling 'just so stories' (Evangelista 2015, loc 3917). What is the problem with

'just so stories'? Stephen Jay Gould's criticism was applied to evolutionary thinking where we find some characteristic of a creature and tell some story about how that characteristic confers some evolutionary advantage or is some holdover from some previous fitness advantage. Gould suggests that there might be 'spandrels' that never served any purpose, but also never detracted enough from the animal's fitness to be eliminated (Gould and Lewontin 1979; Gould 1997). In that sense, a 'just so story' is simply another name for a hypothesis – but one that cannot be adequately tested. The tests of process tracing are intended to enable us to make judgements about hypotheses, and to suggest where we might look for evidence. The problem with unique events such as 'the end of the Cold War', is that hypotheses about which particular contributing events are necessary and jointly sufficient for the outcome can only be indirectly tested. We cannot run the ending of the Cold War again. The four process-tracing tests, along with Waldner's (2015) 'completeness standard', are there to see which mechanism seems to best fit the narrative.

5. Critical Junctures

The idea of critical junctures is not, strictly speaking, part of the methodology of process tracing. To some extent, it is part of a theoretical tradition associated with historical institutionalism that argues that societies take on their different forms because of path-dependencies given specific decisions or historical moments. Critical junctures mark the change from one institutional form to another. They are usually thought to involve major institutional changes, which are distinct from what went before and result in an enduring legacy. They are relevant to this discussion because a critical juncture is, by definition, an important causal factor in some broader claim about the nature of what we see as an important outcome. A great deal of literature is concerned with defining precisely what is meant by a critical juncture (for example, Soiffer 2012; Collier and Munck 2017; Stark 2018). How do we distinguish the critical juncture

from other aspects of the historical process? delineate it from other events? separate antecedent conditions from the juncture itself? and how do we bound the concept?

In fact, I do not think we need worry too much about these elements at all. Any distinction between the antecedent conditions and the critical juncture itself, or discussion of how long was the critical juncture between the former institutional arrangements and the new ones, are simply verbal disputes over coding decisions (Dowding and Bosworth 2018). To be sure, some changes – for example, constitutional amendments – can be given precise dates and forms, though their impact on society might be less susceptible to precision. However, most critical junctures have fluid boundaries. It is rare that any critical juncture would be recognized by those living through it. Interest in critical junctures, and enduring debates over claims about them, usually derive from the way in which they are used to defend theoretical claims about types of mechanisms. Criticisms are usually at lower levels of granularity concerning the precise details of the purported mechanism and nature of the specific critical juncture in different societies. The theoretical claims are made at high generality, the critics concentrating upon details. Or the claim is that a specific episode, identified as a critical juncture leaving an enduring legacy, is not the full story and in fact change was slower and more incremental. Questions of determinacy and contingency are often involved in these disputes. Here the issue is the underlying theoretical basis of historical institutionalism. Path dependency suggests critical junctures that create path dependency, rather than incremental changes which suggest greater continency.

All of these issues can be seen in terms of the stability of the background conditions relative to the foregrounded explanation. There is really little point in trying to carefully define what constitutes a critical juncture, what makes something incremental rather than a radical change, or whether the process really was determined from some point in time or was more contingent. As long as we are aware of the relevant levels of analysis and use our evidence to

suggest, for any given narrative, what should be foregrounded in terms of the stability of the background conditions, we can keep in mind the general claims being made. Far too often, competing theories or narratives are pitted against each other, when they are not rival at all.

6. Conclusion

Process tracing is designed to make qualitative historical description more systematic and better able to make secure causal claims. Historical narrative necessarily involves making causal linkages, but the four tests of process tracing are designed to make researchers carefully consider the nature of the evidence they offer. We can represent this in a Bayesian framework; but without actually measuring our probabilistic inferences, that framework is simply justificatory and not methodological. Process tracing is designed to examine the process of changes. It should not be confused with the idea of type-level mechanism – though, of course, the process can be a token example of such a mechanism. I suggest the term 'process' be used to describe the detailed examination of token cases, and 'mechanism' be reserved for the type-level low-granularity description.

There is no need to think of process tracing as using a rival account of causation to that of high-*n* statistical analysis. The decision to forefront some aspects of the causal process as 'causes' and others as background conditions can be made in terms of the stability of the outcomes to changes in those foregrounded and those backgrounded. We can draw an analogy with the internal validity and external validity of experiments. Researchers should not waste too much time defining what aspects of their study are critical junctures or how incremental processes really are. Many of these debates are really about whether the correct decisions have been made to foreground some aspects of the narrative and what level of granularity we think appropriate to the research question being addressed.

Process tracing is a valuable methodology for the social sciences, but has attracted too much attention to ontological and epistemological questions over its explanatory status. These issues can be more easily seen in terms of the type–token distinction and of the historical-proximate or structural-ultimate nature of the explanations offered in different accounts. Different answers to similar questions are not necessarily rival once it is realized that those questions are addressed to different levels of analysis and thus require answers appropriate to that level.

References

Beach, Derek, and Rasmus Brun Pedersen. 2016. *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Abor, MI: University of Michigan Press.

Bennett, Andrew. 2008. "Process Tracing: A Bayesian Perspective." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 1–23. Oxford: Oxford University Press.

Bennett, Andrew. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, edited by Henry E. Brady and David Collier, 2nd ed., 207–19. New York: Oxford University Press.

Bennett, Andrew. 2015. "Appendix: Disciplining Our Conjectures: Systematicing Process Tracing with Bayesian Analysis." In *Process Tracing: From Metaphor to Analytic Tool*, in Andrew Bennett and Jeffrey T. Checkel (eds) Cambridge: Cambridge University Press.

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

Bennett, Andrew, and Jeffrey T. Checkel. 2015b. *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.

Bertalanffy, Ludwig von. 1968. *General Systems Theory: Foundations, Development, Applications*. New York: George Braziller.

Campbell, D. T. 1986. "Relabelling Internal and External Validity for Applied Social Scientists." In *Advances in Quasi-Experimental Design and Analysis*, edited by W. M. K. Trochim. San Francisco, CA: Jossey-Bass.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin.

Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44 (4): 823–30.

Collier, David, and Gerardo L. Munck. 2017. "Building Blocks and Methodological Challenges: A Framework for Studying Critical Junctures." *Qualitative and Multi-Method Research* 15 (1): 2–9.

Craver, Carl F. 2006. "When Mechanistic Models Explain." *Synthese* 154: 355–76.

Dowding, Keith. 2016. *The Philosophy and Methods of Political Science*. London: Palgrave.

Dowding, Keith. 2017a. *Power, Luck and Freedom: Collected Essays*. Manchester: Manchester University Press.

Dowding, Keith. 2017b. "So Much to Say: Response to Commentators." *Political Studies Review* 15: 217–30.

Dowding, Keith. 2020. "Can a Case-Study Test a Theory? Types and Tokens in Social Explanation." In *Handbook of Methods for Comparative Policy Analysis*, edited by B. Guy Peters and Guillaume Fontaine. Aldershot: Edward Elgar.

Dowding, Keith, and William Bosworth. 2018. "Ambiguity and Vagueness in Political Terminology: On Coding and Referential Vacuity." *European Journal of Political Theory* online first (doi: 10.1177/1474885118771256).

Evangelista, Matthew. 2015. "Explaining the Cold War's End: Process Tracing All the Way Down?" In *Process Tracing: From Metaphor to Analytic Tool*, in Andrew Bennett and Jeffrey T. Checkel (eds). Cambridge: Cambridge University Press.

Fairfield, Tasha, and Andrew E. Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25 (3): 363–80.

Fairfield, Tasha, and Andrew E. Charman 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17 (1): 154–67.

Fetzer, James H. 1975. "On the Historical Explanation of Unique Events." *Theory and Decision* 6 (1): 87–97.

George, Alexander L. 1979. "Case Studies and Theory Development: The Method of Structured, Focused Comparison." In *Diplomacy: New Approaches in History, Theory and Policy*, edited by Paul Gordon Lauren, 43–68. New York: Free Press.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

Gerring, John. 2017. "Qualitative Methods." *Annual Reviews of Political Science* 20: 15–36.

Gerring, John, and Rose McDermott. 2007. "An Experimental Template for Case Study Research." *American Journal of Political Science* 51 (3): 688–701.

Goemans, Heinz. 2000. *War and Punishment: The Causes of War Termination and the First World War*. Princeton, NJ: Princeton University Press.

Gould, Stephen Jay. 1997. "The Exaptive Excellence of Spandrels as a Term and a Prototype." *Proceedings of the National Academy of Sciences USA* 94: 10750–5.

Gould, Stephen Jay, and Richard C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society London* 205 (1161): 581–98.

Hedstrom, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49–67.

Humphreys, Macartan, and Alan M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109 (4): 653–73.

Kitcher, Philip. 2003. *In Mendel's Mirror: Philosophical Reflections on Biology*. Oxford: Oxford University Press.

Mackie, John L. 1974. *The Cement of the Universe: A Study in Causation*. Oxford: Clarendon Press.

Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. 2nd Edition, Cambridge: Cambridge University Press, .

Soiffer, Hillel David. 2012. "The Causal Logic of Critical Junctures." *Comparative Political Studies* 45 (12): 1572–98.

Stark, Alistair. 2018. "New Institutionalism, Critical Junctures and Post-Crisis Policy Reform." *Australian Journal of Political Science* 53 (1): 24–39.

Tucker, Aviezer. 1998. "Unique Events: The Underdetermination of Explanation." *Erkenntnis* 48: 59–80.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.

Waldner, David. 2012. "Process Tracing and Causal Mechanisms." In *Oxford Handbook of the Philosophy of Social Science*, edited by Harold Kincaid. Oxford: Oxford University Press.

Waldner, David. 2015. "What Makes Process Tracing Good? Causal Mechanisms, Causal Inference, and the Completeness Standard in Comparative Politics." In *Process Tracing:*

From Metaphor to Analytic Tool, in Andrew Bennett and Jeffrey T. Checkel (eds), 126–52. Cambridge: Cambridge Univeristy Press.

Wetzel, Linda. 2009. *Types and Tokens: On Abstract Objects*. Cambridge: Cambridge University Press.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wright, Richard. 1985. "Causation in Tort Law." *California Law Review* 73: 1735–1828.

Wright, Richard. 1988. "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts." *Iowa Law Review* 73: 1001–77.