

# Concrete Counterfactual Tests for Process Tracing

Rosa W. Runhardt, Institute for Philosophy, Leiden University, [r.w.runhardt@phil.leidenuniv.nl](mailto:r.w.runhardt@phil.leidenuniv.nl)

## 1. Introduction

James Woodward's interventionist theory of causation is a methodological project (Woodward 2003; Woodward 2015), amongst others aimed at helping the social scientist discern between correlation and genuine causation. At its foundation, the interventionist theory relies on a counterfactual view of causation. While the interventionist theory has clear implications for quantitative research (cf. Woodward 2000; Woodward 2007a; Woodward 2007b; Waldner 2012), its implications for case study research are rarely explored.

Case study researchers, meanwhile, are increasingly impressed by process tracing as a method for discerning correlation and causation: process tracing promises to open the 'black box' of case study causation by focusing on finding the causal mechanisms behind observed correlations (cf. Bennett and George 1997; Mahoney 2001; George and Bennett 2005; Brady and Collier 2010; Hedström and Ylikoski 2010; Hall 2013; Bennett and Checkel 2015; Beach and Pedersen 2019). However, there is a clear lack of consensus about the appropriate underlying philosophy of causation, as evident in e.g. debates about the benefits of a counterfactual versus a systems approach (Beach 2016; Jacobs 2016; Runhardt 2016). One of the charges levelled against the counterfactual approach is that it lacks concrete evidential tests.

This paper combines interventionist theory with process tracing methodology, thus bringing the philosophical and methodological literature closer together and filling gaps in both. I will show that for mechanism-based case study research, interventionism has concrete implications; interventionism suggests evidential tests which can corroborate case studies' mechanistic causal claims. Since these tests rely on specifying and evaluating counterfactual claims, I will compare interventionism's technical demands with the often more intuitive positions in the social scientific literature on counterfactual claims in case study research (cf. Tetlock and Belkin 1996; Goertz and Levy 2007; Levy 2008; Harvey 2011; Nolan 2013; Harvey 2015; Levy 2015; Mahoney and Barrenechea 2019).

The structure of the paper is as follows. First, I suggest how one ought to apply interventionist theory of causation in case study research, based on the theory's technical definition of singular, actual causation. In particular, I introduce and discuss the notion of a so-called 'hypothetical intervention associated with a causal claim', and detail three pieces of evidence that one must gather to support interventions. Subsequently, I argue that the interventionist approach clarifies the logic behind more imprecise counterfactual approaches in the social sciences. This supports my claim that interventionism suggests concrete evidential tests, despite its counterfactual view of causation.

The second part of the paper supports the first by investigating a concrete example, Haggard and Kaufman's process-tracing work in *Dictators and Democrats* (Haggard and Kaufman 2016). In particular, I analyse their process tracing analysis of the Argentinian democratic transition. I argue that the support Haggard and Kaufman's need to corroborate their own hypothesis should be described in terms of hypothetical interventions. Evaluating Haggard and Kaufman's work by causal interventionist standards, this paper finds that the authors do not meet interventionism's strict demands for corroborating singular causal claims. Although the authors circumscribe the singular

causal claim being tested in line with interventionist demands, they do not deliver sufficient evidence for the associated counterfactual. In the last part of the paper, I frame this negative conclusion in Bayesian terms.

## 2. Interventionism

### 2.1 Interventionism

James Woodward's interventionism is a "set of methodological proposals" (Woodward 2015, 3577) which analyses a causal claim  $X \rightarrow Y$  by asking whether intervening on  $X$  would have an effect on  $Y$ . If appropriate changes in  $X$  are associated with changes in  $Y$ , and some specific technical requirements on the intervention and other variables connected to  $X$  and  $Y$  are met, the relation is causal. Interventionist evidence, in its simplest form, can include evidence from actual interventions (e.g. in randomized controlled trials) and natural experiments (Woodward 2003; Reiss 2005).

Woodward realises that one cannot always manipulate a putative cause  $X$  in practice; interventionism admits that in those cases, we should think of the intervention as purely hypothetical. In other words, in those cases the interventionist theory is a *counterfactual* approach to causation, in which we search for evidence of what *would* happen to  $Y$  if we intervened on  $X$  in a particular way. In social scientific case study research, it is rarely possible or desirable to manipulate putative causes and so this area of research is fitting for Woodward's analysis.

While interventionist theory gives us a clear set of technical requirements on what a hypothetical intervention ought to look like to support the claim that  $X \rightarrow Y$ , Woodward does not consider what evidence for the intervention *looks like* in qualitative social sciences. What sources of evidence are appropriate? What types of evidence? Woodward's most detailed examples are from the natural sciences and so of little help. Yet there also exists a body of literature specifically aimed at working out what counterfactual claims are acceptable in the social sciences (Tetlock and Belkin 1996; Schroeder 2004; Goertz and Levy 2007; Levy 2008; Harvey 2011; Nolan 2013; Harvey 2015; Levy 2015; Mahoney and Barrenechea 2019). Marrying these approaches and interventionist theory will thus be fruitful. I will come back to this below.

As stated in the introduction, I will focus on mechanistic claims in this paper. There exists a range of definitions of what a mechanism actually is (cf. Mahoney 2001). Here, I will follow my earlier analysis (Runhardt 2015), in which I describe process tracing as such: "In the simplest case (in which there is only one hypothesized mechanism), we may formalize process tracing as follows. Let us call the researcher's own hypothesis  $H_Z$ .  $H_Z$  holds that a causal mechanism  $Z$  is behind a process linking a putative cause,  $X$ , and the observed effect,  $Y$ . This mechanism has observable implications, that is, a set of variables  $Z_i$  such that  $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow \dots \rightarrow Y$  (where  $Z_i \rightarrow Z_j$  means that  $Z_i$  causes  $Z_j$ ). It is this chain of events that process tracers trace; note that there is a difference between causal mechanisms and the process that they produce." (Runhardt 2015, 1297)

In earlier work (Runhardt 2015; Runhardt 2016), I have shown that interventionism asks us to consider each step  $Z_i \rightarrow Z_{i+1}$  of the causal chain  $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow \dots \rightarrow Y$  in turn, providing evidence of what would happen to  $Z_{i+1}$  under an intervention on  $Z_i$ . Given this reconstruction, an interventionist approach to corroborating a singular mechanistic causal claim will consist of three main steps: (1) describing each causal 'step' under investigation (including the  $Z_i$  and  $Z_{i+1}$  in the causal relation, but also the wider network of potentially causally related factors) and the associated counterfactual; (2) choosing an appropriate intervention; (3) evaluating evidence of what would happen under this intervention, i.e. guidelines for evaluating evidence for or against the counterfactual. In the remainder of this section, I will detail each of these steps in turn.

### 2.1.1 Describing the causal claim and associated counterfactual

To start an interventionist analysis, one must carefully circumscribe the causal claim  $X \rightarrow Y$  and its associated counterfactual. Woodward gives the example of “Being female causes one to be discriminated against in hiring and/or salary” (Woodward 2003, 115) and asks what the associated counterfactual is. He argues this counterfactual should not be about physically changing someone's sex to see if they stand a better chance of getting hired or getting a higher salary; rather, it is about changing the perception of the person's gender by the hiring committee to see what happens to that person's hiring odds and salary level. Considering this associated counterfactual allows Woodward to carefully circumscribe the causal claim at issue here. ‘Being female causes one to be discriminated against in hiring and/or salary’ was too vaguely specified, but once the counterfactual is stated clearly,  $X$  and  $Y$  are described as well. In Woodward's example, the following counterfactual claim about an intervention outcome should be investigated: “if one were to intervene to change an employer's belief that a candidate is male (female) to a belief that the candidate is female (male), the result would be to decrease (increase) the probability that the candidate is hired and decrease (increase) the salary that the candidate is offered.” (Woodward 2003, 115)

In the above example, Woodward analyses a general causal claim. However, it is not difficult to see how his analysis would apply to a singular, actual causal claim. Say we wish to know whether Mary's gender was an actual cause of her being passed over for a particular sales manager job. The associated counterfactual then is: ‘had we intervened to change the hiring committee's belief that Mary is female to a belief that this candidate is male, the result would be to increase the probability that Mary is hired’.

### 2.1.2 Choosing an appropriate intervention

Once we have circumscribed the causal claim, our second step is to choose an appropriate hypothetical intervention (and subsequently, in the third step, provide evidence for it). The search for what is and is not an appropriate intervention is not new to Woodward's work; it comes up, albeit without mention of interventions, in the methodological literature on counterfactuals. In a thought experiment reminiscent of Woodward's analysis of the gender in hiring case, Jack Levy writes:

“With respect to the counterfactual world defined by the hypothetical non-assassination of Franz Ferdinand in June 1914, it would make a difference, in terms of Austria-Hungary's response and the likelihood of war, whether this outcome was the result of a missed shot by Gavrilo Princip, a shot by Princip that missed Ferdinand but killed his wife Sophie, no shot by Princip after the explosion of the first bomb on the bridge, or no assassination attempt at all. With respect to the outcome, does the hypothesized absence of war refer to the absence of a local Austro-Serbian war or of a world war, and in the summer of 1914 or during the next several years? How one interprets a counterfactual argument might depend on the precise specification of its antecedent and consequent.” (Levy 2015, 389)

Interventionist theory goes beyond the methodological literature because it provides a more precise logic behind the interventions that Levy here intuitively refers to. The intervention, here, is whatever prevents the assassination of Franz Ferdinand in June 1914. In interventionist theory,  $X = x$  causes  $Y = y$  iff:

“(AC\*1) The actual value of  $X = x$  and the actual value of  $Y = y$ .

(AC\*2) For each directed path  $P$  from  $X$  to  $Y$ , fix by interventions all direct causes  $Z_i$  of  $Y$  that do not lie along  $P$  at some combination of values within their redundancy range. Then determine whether, for each path from  $X$  to  $Y$  and for each possible combination of values

for the direct causes  $Z_i$  of  $Y$  that are not on this route and that are in the redundancy range of  $Z_i$ , whether there is an intervention on  $X$  that will change the value of  $Y$ . (AC\* 2) is satisfied if the answer to this question is "yes" for at least one route and possible combination of values within the redundancy range of the  $Z_i$ ." (Woodward 2003, 84)<sup>1</sup>

To illustrate, let's imagine a toy scenario much simpler than the state of the world in the Levy quotation, i.e. the complex causal network that describes the run-up to World War I. Say we investigate a local high school, asking whether its class size cap of 30 students in each mathematics class is an actual cause of the 41% students receiving a strong pass in their final school exams. The first step of intervention theory, that of careful circumscription, is met in our class size example. The hypothesis is sufficiently specific. Contrast this with a more vague hypothesis like "students do better in a small group", which would not be circumscribed carefully enough. We must describe what we mean by 'doing better'; in the case above we referred to exam performance, as is common in educational practice. One can imagine other interpretations; we may be after more active participation (raising hands more often, or doing more work in class, etc.), a greater sense of security and inclusion, etc.

The second step of intervention theory is finding an appropriate intervention. This is captured under requirement (AC\*2)<sup>2</sup>, and works as follows. Say that in this toy example the only other cause of attainment is the number of Saturday catch-up mathematics classes students are required to attend each month. We wish to avoid speculation about a scenario in which the number of Saturday classes disrupts the causal link between class size and attainment. Of course, if a particular number of Saturday classes would prevent a change in attainment (the benefit of attending these classes every week is so great for a student that it trumps however large their class is), then under that value of Saturday classes we could not say that class size is an actual cause. But if there is a number of Saturday classes that does not affect the causal link between class size and attainment (say, when we only hold them in April and May, just before the final school exam is sat), this is the number we should assume holds during our hypothetical intervention. We should then find evidence that there is a logically possible intervention that would reduce the class size cap from 30 to lower and thereby improve the percentage of students receiving a strong pass in their final school exam in mathematics, while the number of required Saturday classes is fixed. We could investigate, for example, whether *hiring further staff* could mean smaller classes and thereby bring attainment up. We can coherently describe what it would be like for a school to hire more staff, thus meeting the second step of intervention theory.

### 2.1.3 Evaluating evidence for the intervention

The third step of the intervention approach is to evaluate whether there is evidence for or against the specified intervention from step two. In the toy scenario above, this means we must ask: can we assess claims about what would happen to class size and attainment when we hire more staff in the counterfactual scenario? We investigate this link by collecting evidence about what would happen to the number of strong passes at the local high school if we set the cap to a lower number (allowing, for example, 20 students in each class). That would mean we can disentangle, amongst others, the effect on attainment of the smaller class size from the effects of e.g. the fresh exchange of teaching

<sup>1</sup> Here, the 'redundancy range' is defined as follows: "The values  $v_i \dots v_n$  are in what Hitchcock calls the redundancy range for the variables  $V_i$  with respect to the path  $P$  if, given the actual value of  $X$ , there is no intervention that in setting the values of  $V_i$  to  $v_i \dots v_n$ , will change the (actual) value of  $Y$ ." (Woodward 2003, 83).

<sup>2</sup> Here we have already completed (AC\*1); a simple data collection exercise can tell us the cap at this school is actually 30 students per class and 41% of students received a strong pass (a grade 9-5 in the UK).

techniques when a new member of staff is hired. And this is indeed possible: we may imagine a new member of staff who was trained in the same teaching school etc. and so does not bring a fresh perspective, a mere additional ‘body’ to put in front of the children.

According to Woodward, evidence for an intervention can come from “observation or from a combination of observation and experiment” (Woodward 2003, 35). Information about what happened when class size caps in similar schools (with a similar Saturday class schedule, an intake of students of similar ability, etc.) were introduced; external evidence that some teaching techniques, which can be used in smaller classes but not in larger, are more efficient; some psychological profiling which shows that the students at the high school would work better in smaller groups, etc. all work.

Note that none of these pieces of evidence alone are sufficient to confirm the counterfactual claim. We specify the counterfactual claim in order to clarify for ourselves which pieces of evidence we must collect to corroborate the causal relation (between the class size cap of 30 and attainment of 41% strong passes). We use these pieces of evidence to bolster our faith in the counterfactual, and thereby the causal relation. The fact that we can’t confirm, only corroborate a hypothesis, should not come as a surprise; this is not a drawback unique to interventionism. I will come back to this in the short Bayesian analysis in the final section of the paper.

## 2.2 Comparisons with the methodological literature on counterfactual analysis

So far, we have seen that interventionist theory defines singular causation in terms of hypothetical interventions. I have prescribed three steps for an interventionist investigation: to circumscribe the causal claim, to find an appropriate intervention, and finally to establish evidence for the intervention.

Evidence for a hypothetical intervention is by definition evidence for a counterfactual claim. As mentioned in the introduction, there exists a methodological literature on counterfactual analysis in the social sciences which can supplement interventionist theory with concrete evidential guidelines. So, before describing what evidence for interventions would look like in a concrete case study in section 3, I will finish this section by comparing the methodological and interventionist literature. For purposes of brevity, I will analyse three common requirements: the ‘demand for clarity’, the ‘minimal rewrite rule’, and the demand for ‘plausibility of the antecedent’. I show how an interventionist’s recommendations compare to these requirements, highlighting their agreements and differences.

### 2.2.1 The demand for clarity

The demand for clarity, which we can find in Tetlock and Belkin (Tetlock and Belkin 1996, 19–21) is the demand that counterfactuals have “well-specified antecedents and consequents” (Tetlock and Belkin 1996, 19): one ought to be specific enough in describing the antecedent that other variables that may influence the outcome are mentioned as well. For example, Tetlock and Belkin discuss the counterfactual “If the Industrial Revolution had not occurred, the British standard of living would have been lower than it was” (Tetlock and Belkin 1996, 20). ‘Clarity’ here means we must also specify the growth rate of the British population, since this variable is connected both to the occurrence or absence of the Industrial Revolution and to the standard of living.

At first glance, we may feel this demand is related to interventionist theory’s demands on interventions. However, this is not necessarily the case. Tetlock and Belkin’s worry is that the absence of the Industrial Revolution would have affected the population growth rate, which in turn would have affected the standard of living. But this, in interventionist theory, would be acceptable; we are asked to keep fixed at a certain value all variables connected to *Y* *not on the path between*

$X$  and  $Y$ . Interventionism specifies that “for each directed path  $P$  from  $X$  to  $Y$ , fix by interventions all direct causes  $Z_i$  of  $Y$  that do not lie along  $P$  at some combination of values within their redundancy range” (Woodward 2003, 84). In Woodward’s framework, we are not required to include in the description of the intervention or antecedent any factors  $Z_i$  that *are* on the directed path  $P$  from  $X$  to  $Y$ , unless these factors are independently influenced by the intervention. Yet that is not what is at stake here. When Woodward talks of being specific enough in describing antecedent and consequent, he means something different than Tetlock and Belkin do: he asks only that it is logically possible to change  $X$ , and that we can disentangle “the effect on  $E$  of changing just  $C$  from the effects on  $E$  of changes in other potentially confounding variables, including direct effects from the intervention process itself.” (Woodward 2003, 131) In the Industrial Revolution example, this means that we must distinguish the causal linkage  $I \rightarrow R \rightarrow P \rightarrow L$  (with  $I$  the intervention,  $R$  the industrial revolution,  $P$  the population growth,  $L$  the standard of living) from the influence of  $I \rightarrow P \rightarrow L$  bypassing  $R$ . We do not want the intervention we specify to change the population level except through preventing the industrial revolution.<sup>3</sup>

### 2.2.2 The minimal rewrite rule

The minimal-rewrite rule (Tetlock and Belkin 1996, 23) is the methodological literature’s equivalent of interventionist theory’s demands on appropriate interventions. The minimal-rewrite rule asks that in constructing the counterfactual, we ought to “(a) start with the real world as it was otherwise known before asserting the counterfactual; (b) not (...) unwind the past and rewrite long stretches of history; (c) not unduly disturb what we otherwise know about the original actors and their beliefs and goals” (Tetlock and Belkin 1996, 23). One way to do so is to only change “small events and contingent choices that could have easily turned out differently” (Mahoney and Barrenechea 2019, 333). In Levy, the minimal rewrite rule comes down to the argument that “counterfactual analysis ideally posits an alternative world that is identical to the real world in all theoretically relevant respects but one” (Levy 2008, 635). Interventionist theory shows us the logic behind the minimal rewrite rule; however, interventionist theory also establishes the minimal rewrite rule’s limitations.

In interventionist theory, an intervention ought to only change the putative cause,  $X$ , and not any other factors  $V_i$  related to  $Y$  but not on the path between  $X$  and  $Y$ . This is necessary for being able to distinguish the effects of  $X$ ’s change on  $Y$  from the effects of the other variables  $V_i$ ’s changes on  $Y$ . The whole ‘event’, if you will, is the combination of all variables  $V_i$ ,  $X$ , and  $Y$ ; to say that the ‘event’ ought to be very similar to the ‘counterfactual scenario’ as the minimal-rewrite rule does, is in essence to ask for the  $V_i$ ’s related to  $Y$  but not  $X$  to be fixed at some value. The interventionist requirements and those in the counterfactual methodology literature thus far agree.

However, interventionism says nothing about changing variables we know to be unrelated to  $Y$ . If the intervention affects such variables, this is unproblematic for the interventionist, but not for someone strictly adhering to the minimal rewrite rule. Moreover, while the minimal-rewrite rule can be seen as a way to exclude any effects of intervention  $I$  on effect  $Y$  that ‘circumvent’  $X$ , we should not see

---

<sup>3</sup> A second demand in Tetlock and Belkin, *cotenability*, is also directly related to describing the causal network, in particular the other variables in the causal network of which  $X$  and  $Y$  are a part. “Every counterfactual is a condensed or incomplete argument that requires connecting principles that can sustain, but not imply, the conditional claim. (...) The connecting principles specify, within reasonable limits, everything else that would have to be true to sustain the counterfactual.” (Tetlock and Belkin 1996, 21) So, in the industrial revolution example, we must specify that in order for the standard of living to be lower, the population growth must stay equal to its actual value between 1750 and 1850. The logic behind this is clear from the intervention demands given above; it asks us to be specific about the other variables in the network and their connections with  $X$ ,  $Y$ , and the hypothetical intervention  $I$ .



this rule as a description of what types of  $X$ 's are usually causes (e.g. as a claim that these  $X$ 's are usually small, easily altered events). That would conflate epistemological and ontological considerations. While Mahoney and Barrenechea write that "small and contingent events offer plausible antecedents in counterfactual analysis" (Mahoney and Barrenechea 2019, 318), it would be more appropriate to say that interventions ought to be aimed at changing small and contingent aspects of the whole 'situation', leaving other aspects fixed.

### 2.2.3 Plausibility of the antecedent

We have so far seen how the demand for clarity and the minimal rewrite rule compare to interventionism's technical requirements. The last methodological guideline for counterfactuals that I will discuss is closely related to the minimal rewrite rule: the demand for plausibility of the antecedent. While methodologists like Mahoney and Barrenechea, Levy, Tetlock and Belkin are emphatic that the antecedent of the counterfactual must be 'plausible', interventionist theory does not have a straightforwardly equivalent demand.

To understand the distinction between interventionism and the counterfactual methodologists here, consider the debate about what, if anything, one may change for the counterfactual, between Levy on the one hand (Levy 2008; Levy 2015) and Paul W. Schroeder (Schroeder 2004) on the other. Schroeder summarizes his position in the debate as:

"a common impression about counterfactuals is unsound, namely, that one can choose a particular spot to insert a counterfactual element into history and then trace the changes it might have made in developments subsequent to that point, without altering or affecting what had gone before. In other words, in counterfactual reasoning the path of history antecedent to the counterfactual stays the same; only the consequent future path is altered. This notion now seems to me untrue even for reasonable, plausible counterfactuals. To make them work, one has to change their antecedents as well as their consequents." (Schroeder 2004).

Interventionist theory goes against Schroeder and sides with those who argue we can disregard, to a certain extent, what went on before the counterfactual element's place in history. To see why, consider Woodward's example of the causal relation between the position of the moon with respect to the Earth and the motion of the tides (Woodward 2003, 129–131). This is a singular causal relationship which cannot be tested through an actual intervention or natural experiment (we cannot change the position of the moon, nor is there a similar enough planet with bodies of water and a single moon which is at a different distance from its planet). Thus, the intervention must be hypothetical. Woodward argues that it is not at all clear that such an intervention is even physically possible:

"there is some physically possible process (involving the occurrence at some earlier time of initial conditions that are different from the actual initial conditions prevailing at that time but involving no violation of physical law) that leads from the actual world to a situation in which the moon's orbit is twice its present value in a way that satisfies the conditions for an intervention" (Woodward 2003, 129)

An intervention on the causal relationship between the moon and the tides, Woodward argues, must be

"sufficiently fine-grained and surgical that it does not have any other effects on the tides besides those that occur through the change that it produces in the position of the moon, and it may well be that the laws of nature guarantee that all real causal processes will have

such additional effects. At the very least, it seems wildly optimistic to assume that appropriately surgical intervention processes must be available for all true causal claims.” (Woodward 2003, 130)

This example shows us that it is not within interventionist theory to make arguments along the lines that ‘we cannot change history’. Schroeder’s worry is that counterfactuals are difficult to describe and test because e.g. “Napoleon (...) comes to us historically in one piece. To change what he was capable of becoming and doing after 1805 is to change what he was and was capable of doing before then.” (Schroeder 2004), implying that we change what went on before 1805 even if our counterfactual antecedent is only aimed at that moment in time.

Importantly, interventionist theory is not limited to physically possible interventions. Woodward argues that he has a ‘logical possibility’ approach in mind instead. He finds the following much too strong: “On one notion, an event E is physically possible if and only if it is consistent with the laws of nature and the actually obtaining initial conditions. When conjoined with determinism, this notion of physical possibility implies that interventions on X will not be possible unless they actually occur.” (Woodward 2003, 128) Again, nowhere here is Woodward concerned with the plausibility of the antecedent in the ‘concrete’ terms of the methodologists. His interventionist approach allows him to be more ‘fine-grained’ as it were in choosing what interventions are informative, rather than having to use a more vague notion of ‘plausibility’.

### 3. Causal mechanisms in Haggard and Kaufman’s *Dictators and Democrats*

Thus far, we have seen that the interventionist theory of causation defines singular case causation in terms of hypothetical interventions, and that to corroborate singular claims, it prescribes three steps: circumscribing the causal claim, choosing an appropriate intervention, and finding evidence for this intervention. We have so far seen a brief toy scenario (the example of class size and attainment) to highlight what this last step consists of, and I have discussed how the interventionist theory’s demands compare to three common demands in the methodological literature on counterfactuals. I will now turn to a more detailed analysis on how to find evidence for counterfactual interventions, using a concrete case study. I focus on a recent example of process tracing in political science, Haggard and Kaufman’s analysis of democratic transitions and reversals (Haggard and Kaufman 2012; Haggard and Kaufman 2016).

#### 3.1 Context and method

In *Dictators and Democrats* (Haggard and Kaufman 2016), Haggard and Kaufman study which causal mechanisms played a role in transitions to and from democracy during the Third Wave (1980-2000). The authors analyse and test for several alternative causal processes to democratic transitions and reversions. For example, they study the mechanisms of distributive conflict theory<sup>4</sup> and conclude the distributive mechanisms are less common during the Third Wave than their own suggested mechanisms. Haggard and Kaufman hypothesize that during a number of cases in the Third Wave, institutional strength was a cause of democratic transitions, since institutions mobilized opposition to the government.

To compare the relative merit of different causal mechanisms, the authors use a combination of statistical work and within-case process tracing. As an example of the latter, the authors suggest that

---

<sup>4</sup> Distributive conflict theory argues that inequality and distributive conflict play a role in democratic transitions. Its main proponents are Carles Boix, Daron Acemoglu and James Robinson.



the strength of the unions in Argentina was a cause of mass mobilization there, which in turn led to democratization efforts in the early 1980s.

While Haggard and Kaufman's analysis is part of a more general project, which attempts to uncover the causal mechanisms behind transitions and reversals during the Third Wave in general, for purposes of brevity I will limit my discussion only to their within-case analysis of Argentina. The question of whether their methodology lends itself to subsequent comparisons across cases, or indeed to wholesale generalization, as well as the merits of Haggard and Kaufman's 'large-N qualitative testing' as a mixed method approach (Goertz 2017, chap. 7) is beyond the scope of this paper. In what follows, I analyse what Haggard and Kaufman are able to show in their case study of the Argentinian transition using interventionism's demands from section 2.

### 3.3 Argentina

Haggard and Kaufman focus on the history of the Argentinian transition from the military regime seizing power in 1976 to the transition to a competitively elected government in 1983. After the military regime seized power, they attempted to repress the country's strong union movement, amongst others by "banning parties and strikes and imposing censorship" (Haggard and Kaufman 2016, 111), "direct purges of labor-based Peronist adversaries" (Haggard and Kaufman 2016, 112), and "neoliberal economic reforms, including reforms of the labor market designed to curtail union power" (Haggard and Kaufman 2016, 112). After these reforms, the economy performed poorly, and this, together with structural dislocations, spurred the union movement to respond with a "wave of strikes and general strikes" (Haggard and Kaufman 2016, 113). The strikes put pressure on the regime, which made several changes in its leadership and economic policy, each unsuccessful and leading to more strikes. After this continued pressure, the government decided to invade the Falkland Islands, which ended unsuccessfully. The regime gave in and appointed a caretaker government, which together with amongst others the labour unions organized democratic elections.

To explain the democratic transition in Argentina, Haggard and Kaufman propose the "union collective action mechanism" hypothesis. This mechanism is nicely summarized in Beach and Pedersen (2019): "mass mobilization (strikes and demonstrations) by unions over distributive grievances and repression of unions" (Beach and Pedersen 2019, 153) leads to the "ruling elite responds with further repression" (Beach and Pedersen 2019, 153), after which "unions respond with further strikes and demonstrations that eventually gain support of human rights organizations and other societal groups" (Beach and Pedersen 2019, 153) which has "a compelling effect on [the] government because of [the] economic situation" (Beach and Pedersen 2019, 77). The mechanism contains has "multiple iterations" (Beach & Pedersen 2019, 153), back to further repression and subsequent further strikes, in a feedback loop, but eventually makes it so that the "ruling elite concedes to democratic transition" (Haggard and Kaufman 2016, 71).

Let us now consider what an interventionist approach to collecting evidence for this causal hypothesis would look like. This will consist of the three phases from section 2: (1) circumscribing the causal claim; (2) choosing an appropriate intervention; (3) evaluating evidence of what would happen under this intervention.

### 3.4 Circumscribing the causal claim

As argued briefly in section 2 and in more detail in Runhardt (2015), the counterfactual we are concerned with in this process tracing case study is not simply the general 'if union strength had been lower, the democratic transition in Argentina would not have occurred'. Rather, we must break the process up into its smaller steps. For example, one step in the Argentinian process (which is a token instantiation of the union collective action mechanism) is the causal hypothesis  $U \rightarrow S$  where

$U$  and  $S$  refer to union strength and the first strikes respectively. As Ruth Collier describes in her analysis of the role of labour in the Argentinian transition: “The repression [immediately after the coup in 1976] was extensive and effective, but it did not completely put an end to labor opposition. Even after the coup labor activity had continued.” (Collier 1999, 121). As James McGuire showed, “[b]elying the image of a society immobilized by repression, individual strikes began almost immediately after the 1976 coup” (McGuire 1995, f15).

The description of  $U \rightarrow S$  needs to meet very specific demands, as we have seen in section 2. For one, we must be able to specify both  $U$  and  $S$  in sufficient detail. Consider first  $U$ , the strength of unions in Argentina; Haggard and Kaufman measure this through the proxy of “the per capita membership in unions affiliated with the International Trade Union Confederation (ITUC)” (Haggard and Kaufman 2016, 71). Haggard and Kaufman argue this measurement is a good proxy, since “almost all of the countries with scores of zero are ones in which unions are in effect appendages of the state and/or where unionization rates are in fact low” (Haggard and Kaufman 2016, 113).

The description of  $S$  is more difficult. If we specify  $S$  as, say, the first general strike protesting the government’s economic policies, organized by the unions in April 1979, it is not obvious that  $S$  is conceptually distinguishable from  $U$ . On the other hand, if we specify  $S$  as some more general measure of the number of mass mobilization events in 1979, we are losing track of the motivation for using process tracing (case based) analysis in the first place: we no longer ‘open the black box’ behind a correlation between mobilization and democratic transition.

The authors will be able to specify  $S$  by considering two questions: firstly, what is the intended counterfactual that goes along with  $U \rightarrow S$ ? E.g., are the authors suggesting that if the unions were weaker, no mobilization would have occurred (e.g. not been organized by other groups)? Secondly, I would suggest that in choosing a specific  $S$  it is important to keep track of the *next* event in the process chain starting with  $U \rightarrow S$ .  $S$  must be specified in enough detail to also be a putative cause of the next step, in this case e.g. the repression by government forces.

### 3.5 Describing the intervention

Setting aside the issues with specifying  $S$  for the moment, let us turn to the second step of interventionist process tracing. What hypothetical intervention might lower the per capita membership in unions affiliated with ITUC, as above, in a way that accords with interventionism’s demands? We ought to find an intervention variable  $I$  that meets the following (Woodward 2003, 98; Runhardt 2015):

- 1) Firstly,  $I$  should decrease union strength  $U$ .
- 2) Secondly,  $I$  should act as a switch for  $U$ , that is, make the union strength independent of any other variables.
- 3) Third, the intervention  $I$  cannot itself lead to, or prevent,  $S$  in a way that is unrelated to the strength of unions in Argentina. This means, for example, that increasing the repressive capabilities of the government is most likely a poor intervention; in ‘labour repressive’ regimes like in Argentina, the repressiveness of the system itself is seen as part of the causal process from union strength to transition. The repression of unions was as much a part of the unions’ decision to strike as was, for example, the economic downturn.
- 4) Fourth and last, the intervention  $I$  must be statistically independent of all variables that increase/decrease  $S$  by other means than union strength  $U$ .

### 3.6 Evidence for the counterfactuals

The key question remaining is: have Haggard and Kaufman provided any evidence, directly or indirectly, for the existence of an intervention variable as described above? Is there any reason to believe that an intervention  $I$ , if performed properly on  $U$ , would lead to a change in  $S$ ?

In their analysis of the Argentinian transition, Haggard and Kaufman mention only a few things which corroborate the intervention variable. For example, looking at the demand that by lowering  $U$ , we lower the probability of  $S$ , one might ask whether there is any overdetermination or pre-emption going on; e.g. if other interested parties would have mobilized had the unions not organized strikes. Haggard and Kaufman do not explicitly rule this situation out, but their description of the years leading up to the transition could be interpreted as providing evidence against this specific case of overdetermination and pre-emption. For example, the authors state that while the unions had been organizing strikes since 1977, human rights organizations and political parties did not join in with the strikes until 1982. This arguably makes it more believable that without the unions, human rights organizations and political parties would not have demonstrated instead. Whether this counts as an appropriate counterfactual scenario depends, amongst other things, on their intended variable  $S$ .<sup>5</sup>

In short, my interventionist analysis here is negative. If we accept an interventionist view of causation, we require there to be evidence of intervention claims; such evidence however is never explicitly stated. As such, the causal claims in Haggard and Kaufman's work are far from established.

## 4. A Bayesian analysis of the above

In this final section, I wish to present a brief, alternative reading of the methodological advice in the rest of this paper. In particular, I wish to follow Macartan Humphreys and Alan Jacobs' Bayesian analysis of mixed method research and discuss "the varying likelihoods with which potentially probative pieces of evidence may be associated with causal effects" (Humphreys and Jacobs 2015).

Humphreys and Jacobs describe process tracing as: "a search for clues that will be observed with some probability if the case is of a given causal type and that will be observed with some differing probability if the case is of a different causal type" (Humphreys and Jacobs 2015, 657). They interpret process tracing with a Bayesian approach in the following way:

"In formalizing Bayesian process tracing, we start with a very simple setup, which we then elaborate. To return to our running example, suppose that we already have  $X, Y$  data on one authoritarian regime: we know that it suffered economic crisis ( $X = 1$ ) and collapsed ( $Y = 1$ ). We want to know if  $X$  caused  $Y$ . We answer the question by collecting one or more clues that we believe are related to the case-level causal effect of  $X$  on  $Y$ . We use the variable  $K$  to register the outcome of the search for a clue (or collection of clues), with  $K = 1$  indicating that a specific clue (or collection of clues) is searched for and found, and  $K = 0$  indicating that the clue is searched for and not found." (Humphreys and Jacobs 2015, 657).

---

<sup>5</sup> The authors provide more evidence for an intervention variable  $I \rightarrow S \rightarrow T$ , i.e. one testing the relation between the strikes and the eventual transition. Their section on Argentina is full of phrases indicating that the actions of the government were in response to mobilization. They say for instance that the decision to invade the Falkland Islands was "directly precipitated by growing pressure from below" (Haggard and Kaufman 2016, 113); that "faced with these political and economic challenges" (Humphreys and Jacobs 2015, 655) (i.e. the strikes), the military amongst others replaced leaders and later appointed a caretaker government. The unspoken assumption is that they would not have done so had there not been pressure from below. This counterfactual is never explicitly stated nor is it systematically discussed.

In the context of interventionism, the relevant ‘search for a clue’ is the search for evidence of the hypothetical intervention (step 3 as described in section 2). If we search for and find such evidence,  $K = 1$ . If we search for and do not find such evidence,  $K = 0$ . The relevant probabilities to establish are now the following:

1. The likelihood of the clue given the hypothesis:  $\Pr(K = 1|H)$ . What is the likelihood we find evidence of the hypothetical intervention, if  $X$  indeed caused  $Y$ ?
2. The likelihood we find evidence of the hypothetical intervention if  $X$  *did not* cause  $Y$ , i.e.  $\Pr(K = 1|\neg H)$ . The difference between the two likelihoods  $\Pr(K = 1|H)$  and  $\Pr(K = 1|\neg H)$ , Humphreys and Jacobs clarify, is the probative value of observing this ‘clue’, i.e. the probative value of the hypothetical intervention data.
3. The likelihood of the clue in general,  $\Pr(K = 1)$ , which is a function of the two likelihoods above as well as  $\Pr(H)$  and  $\Pr(\neg H)$ .

These likelihoods combine to give us  $\Pr(H|K = 1)$  in the usual way using Bayes’ theorem, i.e.

$$\Pr(H|K = 1) = \frac{\Pr(K=1|H) \times \Pr(H)}{\Pr(K=1)}.$$

In the interventionist theory, the causal hypothesis  $X \rightarrow Y$  is necessary and sufficient for the truth of the counterfactual describing the hypothetical intervention outcome. As such,  $\Pr(K = 1|\neg H)$  is 0. However, since it will be difficult to find evidence of the hypothetical intervention outcome, it may well be that  $\Pr(K = 1|H)$  is not very high either. This means that in the mapping of Humphreys and Jacobs, following Van Evera’s (1997) terminology, the presence of  $K$  is a ‘smoking gun’ for  $H$  (the causal relation  $X \rightarrow Y$ ), in that it would be sufficient to convince us of  $H$ . Had Haggard and Kaufman established the counterfactual of what would have happened to mobilization  $S$  under lower union strength  $U$ , this would have strongly supported their causal claim  $U \rightarrow S$ . On the other hand, the absence of  $K$  is a ‘hoop test’ for  $\neg H$  ( $X$  does not cause  $Y$ ), in that it would be necessary for  $\neg H$  to be true but not sufficient evidence to convince us of  $\neg H$ . We have seen that Haggard and Kaufman do not provide evidence of  $K$ . This does not mean that their causal claim  $U \rightarrow S$  is necessarily false; however, it also leaves open the possibility that  $U$  does not cause  $S$ .

I will finish by noting that the above is a simplification of how counterfactual evidence and the causal claim relate. We hardly ever have convincing evidence of the truth of a counterfactual; by this, I mean no more than the old Popperian belief that corroboration of a hypothesis is possible, but confirmation of a hypothesis is not. As such, the variable  $K$  which “register[s] the outcome of the search for a clue (or collection of clues), with  $K = 1$  indicating that a specific clue (or collection of clues) is searched for and found, and  $K = 0$  indicating that the clue is searched for and not found” (Humphreys and Jacobs 2015, 657) is an oversimplification of what we have to deal with in searching for counterfactuals. In future work, a more precise analysis of degrees of corroboration of a counterfactual claim is necessary.

## 5. Conclusion

In this paper, I have combined interventionist theory with the process tracing methodology. I showed that interventionist theory defines singular case study causation in terms of hypothetical interventions. Nevertheless, I argued that interventionism has concrete implications; interventionism suggests evidential tests which can corroborate case studies’ mechanistic causal claims. I have prescribed three steps for an interventionist investigation: to circumscribe the causal claim, to find an appropriate intervention, and finally to establish evidence for the intervention.

I spent most of the paper analysing the last of these steps, comparing interventionist theory's demands on such evidence with the already existing methodological demands for counterfactual reasoning in case study research. This led to a technical discussion of which other variables in the causal network can be altered under the intervention (in the counterfactual scenario) and which need to remain fixed.

I then turned to a concrete case study, Haggard and Kaufman's analysis of the Argentinian transition to democracy, and argued that this study does not (yet) meet the requirements of interventionist theory. I finished by showing what this means for the causal claims in Haggard and Kaufman using a Bayesian framework. I argued that the absence of real evidence for the associated counterfactuals that interventionist theory prescribes means that Haggard and Kaufman have failed to strongly support their own causal hypotheses, but it does not definitively establish the alternative hypotheses they discuss either.

## References

- Beach, Derek. 2016. What Are We Actually Tracing? Process Tracing and the Benefits of Conceptualizing Causal Mechanisms as Systems. *Qualitative & Multi-method Research*: 15–22.
- Beach, Derek, and Rasmus Brun Pedersen. 2019. *Process-Tracing Methods: Foundations and Guidelines*. 2nd ed. Ann Arbor: University of Michigan Press.
- Bennett, Andrew, and Jeffrey T Checkel. 2015. *Process Tracing: from Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.
- Bennett, Andrew, and Alexander L George. 1997. Process Tracing in Case Study Research. *MacArthur Foundation Workshop on Case Study Methods*. Belfer Center for Science and International Affairs, Harvard University.
- Brady, Henry E, and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. 2nd ed. Lanham: Rowman & Littlefield.
- Collier, Ruth Berins. 1999. Labor Action in Recent Democratization. In *Paths toward Democracy*, 110–165. Cambridge: Cambridge University Press.
- van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca: Cornell University Press.
- George, Alexander L, and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.
- Goertz, Gary. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies*. Princeton: Princeton University Press.
- Goertz, Gary, and Jack S. Levy. 2007. *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*. London: Routledge.
- Haggard, Stephan, and Robert R Kaufman. 2012. Inequality and Regime Change: Democratic Transitions and the Stability of Democratic Rule. *The American Political Science Review* 106. [American Political Science Association, Cambridge University Press]: 495–516.
- Haggard, Stephan, and Robert R Kaufman. 2016. *Dictators and Democrats: Masses, Elites, and Regime Change*. Princeton: Princeton University Press.

- Hall, Peter A. 2013. Symposium: Tracing the Progress of Process Tracing. *European Political Science* 12: 20–30.
- Harvey, Frank P. 2011. *Explaining the Iraq War: Counterfactual Theory, Logic and Evidence*. Cambridge: Cambridge University Press.
- Harvey, Frank P. 2015. “What If” History Matters? Comparative Counterfactual Analysis and Policy Relevance. *Security Studies* 24: 413–424.
- Hedström, 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: 653–673.
- Jacobs, Alan M. 2016. Introduction: Mechanisms and Process Tracing. *Qualitative & Multi-method Research*: 13–15.
- Levy, Jack S. 2008. Counterfactuals and Case Studies. In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press.
- Levy, Jack S. 2015. Counterfactuals, Causal Inference, and Historical Analysis. *Security Studies* 24: 378–402.
- Mahoney, James. 2001. Beyond Correlational Analysis: Recent Innovations in Theory and Method. *Sociological Forum* 16: 575–593.
- Mahoney, James, and Rodrigo Barrenechea. 2019. The Logic of Counterfactual Analysis in Case-Study Explanation. *The British Journal of Sociology* 70: 306–338.
- McGuire, James W. 1995. Interim Government and Democratic Consolidation: Argentina in Comparative Perspective. In *Between States: Interim Governments in Democratic Transitions*, ed. Yossi Shain, Juan J. Linz, and Lynn Berat, 179–210. Cambridge: Cambridge University Press.
- Nolan, Daniel. 2013. Why Historians (and Everyone Else) Should Care About Counterfactuals. *Philosophical Studies* 163: 317–335.
- Reiss, Julian. 2005. Causal Instrumental Variables and Interventions. *Philosophy of Science* 72: 964–976.
- Runhardt, Rosa W. 2015. Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward’s Manipulability Theory of Causation. *Philosophy of Science* 82: 1296–1307.
- Runhardt, Rosa W. 2016. Tracing the Productive Continuity of Social Mechanisms. *Qualitative and Multi-Method Research: Newsletter of the American Political Science Association’s QMMR Section* 14: 22–28.
- Schroeder, Paul W. 2004. Embedded Counterfactuals and World War I as an Unavoidable War. In *Systems, Stability, and Statecraft: Essays on the International History of Modern Europe*, ed. Paul W. Schroeder, David Wetzell, Robert Jervis, and Jack S. Levy. New York: Palgrave Macmillan.
- Tetlock, Philip E., and Aaron Belkin. 1996. *Counterfactual Thought Experiments in World Politics*. Princeton: Princeton University Press.



- Waldner, David. 2012. Process Tracing and Causal Mechanisms. In *The Oxford Handbook of Philosophy of Social Science*, ed. Harold Kincaid, 65–84. Oxford: Oxford University Press.
- Woodward, James. 2000. Explanation and Invariance in the Special Sciences. *The British Journal for the Philosophy of Science* 51: 197–254.
- Woodward, James. 2003. *Making Things Happen: a Theory of Causal Explanation*. Oxford: Oxford University Press.
- Woodward, James. 2007a. Causal models in the Social Sciences. In *Philosophy of Anthropology and Sociology*, ed. Stephen P. Turner and Mark W. Risjord, 157–210. Amsterdam: Elsevier.
- Woodward, James. 2007b. Interventionist Theories of Causation in Psychological Perspective. In *Causal Learning: Psychology, Philosophy, and Computation*, ed. Alison Gopnik and Laura Schulz, 19–36. Oxford: Oxford University Press.
- Woodward, James. 2015. Methodology, Ontology, and Interventionism. *Synthese* 192: 3577–3599.