A new take on causal mechanisms in social contexts: manipulability theory’s demands for mechanistic reasoning

# 1. Introduction

In a backlash against the pervasiveness of statistical methods ([cf. King, Keohane, and Verba 1994](#_ENREF_9)), in the last decade certain social scientists have focused on finding the causal mechanisms behind observed correlations ([Mahoney 2001](#_ENREF_10), [Tilly 2001](#_ENREF_12), [Hedström and Ylikoski 2010](#_ENREF_7), [Hall 2012](#_ENREF_6)). To provide evidence for such mechanisms, researchers increasingly rely on process-tracing, a method which involves contrasting the observable implications of several alternative mechanisms ([Bennett and Checkel forthcoming](#_ENREF_2), [Brady and Collier 2010](#_ENREF_3), [George and Bennett 2005](#_ENREF_5)).

The process-tracing methodology literature as of yet does not commit to any particular fundamental notion of causation. Process-tracing reacts to the statistical approach by arguing that finding a correlation between a potential cause and effect variable is not enough evidence for genuine causation. We should also investigate the intervening variables between the putative cause and effect. Process-tracing however does not solve the problem it set out to solve, but rather push the problem one step back. What, after all, is their evidence that the link between these intervening variables are cases of genuine causation? In this paper, I will show a way out of this problem. James Woodward’s manipulability theory of causation tells process-tracers how to find the evidence they need: not only must process-tracers study the intervening variables, but also the *intervention variables* of each link in the causal chain.

This paper is set up as follows. First, I analyse what process-tracing is and what it aims to do. Second, I set out the relevant aspects of Woodward’s theory, including my motivation for using his notion of causation rather than another. Third, I evaluate process-tracing in light of Woodward’s theory, conclude it indeed lacks evidence for genuine causation, and give recommendations for solving this problem.

# 2. A philosophical reconstruction of process-tracing

Process-tracing is a mechanism-based method for analysing causal relationships. To be precise, the term refers to two techniques ([cf. Bennett and Checkel forthcoming](#_ENREF_2)), bottom-up and top-down process-tracing.Bottom-up process-tracing involves surveying a situation of interest with as little preconceptions as possible, in order to then formulate a hypothesis about possible causal connections in that situation. For instance, a researcher may spend time in a post-conflict area in the process of nation-building, interview the population to get data on how secure people feel, and subsequently form a hypothesis about causal links between nation-building efforts and human security. Top-down process-tracing tests type-level causal hypotheses about the mechanism (the ‘process’) connecting an independent variable and a dependent variable using data collected in case studies. Bottom-up and top-down process-tracing are occasionally mixed; a researcher may start with a bottom-up study to formulate hypotheses, and continue with a top-down study to see if these hypotheses are corroborated or refuted by the evidence available. In what follows, I will look at the second type of process-tracing, i.e. top-down process-tracing, because I wish to evaluate how process-tracers *justify* causal claims.

First, let us consider top-down process-tracing more formally. The essence of top-down process-tracing is using a case study to contrast rival hypotheses about the causal connection between an independent variable and a dependent variable , that is, hypotheses which suggest rival causal mechanisms that have contradictory observable implications. Let us call the researcher’s own hypothesis . holds that a causal mechanism exists that connects and , i.e. a set of variables such that (where means that causes ). Besides hypothesis , process-tracers will also investigate any alternative hypotheses , , etc. that are postulated in the literature, that is, they also investigate the observable implications of chains or of intermediate variables or etc.

Methodologists Andrew Bennett and Jeffrey Checkel argue that the observable implications of mechanisms are “the facts and sequences within a case that should be true if each of the alternative hypothesized explanations of the case is true. Which actors should have known, said, and did what, and when? Who should have interacted with, worried about, or allied with whom?” ([Bennett and Checkel forthcoming, 39](#_ENREF_2)) According to Sharon Crasnow, process-tracing hypotheses about causal mechanisms are therefore *singular* causal claims, whereas the evidence from case studies consists of *general* causal claims ([Crasnow 2012](#_ENREF_4)). A philosophical account of process-tracing therefore needs to consider both singular and general causal claims, as well as the link between them. In this paper, I will focus on the general causal claims process-tracing makes. In my conclusion I will, however, indicate what implications Woodward’s theory has for the link between singular case study evidence and general hypotheses.

# 3. Woodward’s manipulability theory of causation

Let me now turn to the relevant aspects of James Woodward’s manipulability theory of causation, before explicating how we can apply the theory to process-tracing. Woodward argues that any successful description of a cause-effect relationship must refer to causal factors that can be manipulated to change the phenomenon under study. Specifically, a variable is a cause of a variable if there exists some ‘intervention variable’ which we can use to change , so that will then in turn change without any interference of other variables linked to . In other words, using we can ascertain that made the change in happen.

I have chosen to look at what would happen if the process-tracer committed to Woodward’s notion of causation, rather than others, because of two reasons. Firstly, Woodward’s theory provides an alternative to the probabilistic notions of causation that are taken for granted in the statistical approaches that process-tracers criticize, such as the one in social science methodology bible [King, Keohane, and Verba (1994)](#_ENREF_9). Secondly, Woodward’s notion is arguably more suited to studying causal mechanisms in social science than the energy-transfer notions of causation developed for causal mechanisms in natural sciences like biology. Thirdly, Woodward’s notion has not been widely applied to the social sciences, and therefore this analysis may contribute to the literature in philosophy of causation as well as to philosophy of social science.

Arguably, we could get similar results by accepting Judea Pearl’s manipulability framework for causation ([Pearl 2000](#_ENREF_11)). Though there are formal differences between Pearl and Woodward’s notion of an intervention, I believe that my general conclusion in this paper will hold no matter whether the process-tracer commits to Pearl or Woodward’s theory.

## 3.1 Manipulability theory

Let me now outline Woodward’s theory. The focal point of Woodward’s work is his formal set of necessary and sufficient conditions for to be a (type-level) cause of , which form his manipulability theory:

A necessary and sufficient condition for to be a (type-level) **direct cause** of with respect to a variable set is that there be a possible intervention on that will change or the probability of when one holds fixed at some value all other variables in . ([Woodward 2003, 59](#_ENREF_13))

To illustrate the use of the variable set , consider the following scenario: we are interested in a Scandinavian village, asking whether, for its villagers, eating citrus fruit () is a direct cause of an absence of scurvy (). To answer that question, we can’t just feed the villagers citrus fruit for a month to see what happens to their health. We need to take into account other variables that may influence this (lack of) scurvy. So, we investigate the villagers’ diet, and find out that they greatly enjoy eating liver; their liver consumption () is very high. What will happen in our experiments to determine the effect of citrus consumption is the following. If we *ignore* the liver consumption, , of the villagers, we will find that no possible intervention on their citrus consumption, , will change their developing scurvy or not, . Simply put, not eating citrus fruit won’t mean that the villagers get scurvy. However, *if we keep fixed at 0* the variable for these villagers, we will find out that there is an intervention on , i.e. making the villagers eat citrus fruit, that *will* change , i.e. whether they develop scurvy. We find that if , i.e. the villagers don’t consume the fruit, then , i.e. they develop the deficiency disease. If they *do* consume the fruit, i.e. , then they don’t develop the disease, i.e. .

The notion of a direct cause alone, however, is too basic for a complete theory of causation. Woodward calls our attention to the possibility of a variable which influences a variable along some route but has no total effect on because ’s influence is always cancelled out by other factors ([Woodward 2003, 50](#_ENREF_13))[[1]](#footnote-1). In that case, is not a direct cause of , but Woodward nevertheless wants to call a cause. Therefore he introduces the notion of a contributing cause:

A necessary and sufficient condition for to be a (type-level) **contributing cause** of with respect to variable set is that:

1. there be a directed path from to such that each link in this path is a direct causal relationship; that is, a set of variables such that is a direct cause of , which in turn is a direct cause of , which is a direct cause of … , which is a direct cause of ; and that
2. there be some intervention on that will change when all other variables in that are not on this path are fixed at some value.[[2]](#footnote-2)

If there is only one path from to , or if the only alternative path from to besides contains no intermediate variables (i.e., is direct), then is a contributing cause of as long as there is some intervention on that will change the value of , for some values of the other variables in . ([Woodward 2003, 59](#_ENREF_13))

As Woodward himself stresses, a direct cause is always a contributing cause, but a contributing cause is not always a direct cause.

## 3.2 Interventions

The notion of an ‘intervention’ is a crucial part of Woodward’s argument. Note that there is a difference between an intervention variable and a contributing cause variable; whereas a contributing cause variable is part of the situation one is trying to analyse, the intervention variable is the means by which one undertakes this analysis. Before I discuss Woodward’s rather technical definition of an intervention variable, I will introduce it with an example.

According to Woodward’s theory, introducing a microfinance institution in a country will be an intervention variable for investigating whether taking out microcredit loans () causes a reduction in household poverty () if and only if the following things hold. First, the introduction of the microfinance institution has to increase the probability that a microcredit is taken out. Second, there must be no other source of microcredit loans besides this microfinance institution (so that when we do not introduce the microfinance institution, no microcredits will be taken out). Third, and this is more difficult to ascertain in practice, the introduction of the microfinance institution should not reduce poverty in a way that is unrelated to microcredits. If it turns out, for instance, that opening a microsavings account also reduces households’ poverty, and such accounts are offered by the microfinance institution, the third demand will fail. We would not be able to tell whether the microcredit loan or the microsavings account made the difference. In general, overlooking other ways besides whereby may influence clouds our judgement about the relation between and . Fourth and last, introducing the microfinance institution must be statistically independent of all variables that reduce poverty by other means than microcredit loans. For instance, if we can only introduce the microfinance institution in regions that have a stable government, this clouds our judgement: the stability of the government could itself cause an eventual reduction in households’ poverty. So, we must ascertain that there are no other ways in which can influence ; if there were, that would mean that gives us a misguided picture of the connection between and . (To see the difference between the third and fourth requirement, consider the following. Both the third and the fourth requirement are violated if there is a factor causally connected to both and but not to . Requirement 3 only captures cases in which we have , whereas for requirement 4, the relation between and is unknown. It may, for instance, just as well be that .)

This brings us to the four requirements in Woodward’s definition of an intervention for type-level causation:

“ is an **intervention variable** for with respect to if and only if meets the following conditions:

1. causes .
2. acts as a switch for all the other variables that cause . That is, certain values of are such that when attains those values, ceases to depend on the values of other variables that cause and instead depends only on the value taken by .
3. Any directed path from to goes through . That is, does not directly cause and is not a cause of any causes of that are distinct from except, of course, for those causes of , if any, that are built into the connection itself; that is, except for
	1. any causes of that are effects of (i.e. variables that are causally between and ) and
	2. any causes of that are between and and have no effect on independently of .
4. is (statistically) independent of any variable that causes and that is on a directed path that does not go through .” ([Woodward 2003, 98](#_ENREF_13))

In short, is an intervention variable for with respect to when we can use to check whether is a (direct or contributing) cause of , i.e. when we can use to change , where after will change without interference from other variables causally related to . Using , we will be able to ascertain that made the change in happen.

Woodward claims that the intervention does not *actually* need to happen; we may devise a hypothetical experiment. What’s more, the intervention does not need to involve human action. A natural process can qualify as an intervention as well. In the microfinance case, it may well be that there are two regions in the world that are similar in all crucial respects[[3]](#footnote-3) except that one has microfinance institutions whereas the other does not. If we compared the two, taking into account all the requirements above, and found that in the country without microfinance institutions a larger proportion of households was below the poverty threshold than in the country with microfinance institutions, then this would corroborate the claim that there is a causal relation between taking out microcredits and reduction of the proportion of poor households.

Summing up the above, Woodward makes a distinction between contributing causes , intervention variables that we use to analyse whether a variable is in fact a cause, and intervening variables that are the means by which a contributing cause influences its effect .

# 4. Process-tracing evaluated from the perspective of the manipulability theory of causation

At first glance, Woodward’s notion of a *contributing cause* fits with the hypotheses in a top-down process-tracing study. In what follows I will argue that although the hypothesis has structural similarities with Woodward’s notion, nevertheless the proposed methods for testing the hypotheses are quite different. In Woodward’s framework, we need to show that all links of the chain connecting and are cases of direct causation, which means we need to show there exists some intervention on that will change . In contrast, all the process-tracing method outlined by methodologists like Bennett and Checkel requires is that we observe the deductive implications of the intervening variables of the mechanism in a case study.

To contrast the two approaches in more detail, consider a simple example. Imagine a social scientist has the type-level causal hypothesis that ‘an economic recession, , is a contributing cause of a drop in non-domestic violent crime, , via the intervening variable of a drop in participation in the night time economy, ’[[4]](#footnote-4). In this example, Woodward would urge the social scientist to answer the following questions:

1. Is a direct cause of ? In other words, is there a possible intervention on that will change or the probability of when one holds fixed all other variables in at some value?
2. Is a direct cause of ? In other words, is there a possible intervention on that will change or the probability of when one holds fixed all other variables in at some value?

(In practice, as we have seen, this scientist would also investigate the observational implications of alternative mechanisms, e.g. the hypothetical mechanism that causes by means of a rise in the number of people suffering from clinical depression, . For conciseness’ sake, I will not discuss the scientists’ study of this alternative mechanism; it will happen analogously to the study of her own proposed mechanism .)

Using Woodward’s definition of an intervention variable ([Woodward 2003, 98](#_ENREF_13)), we can now adapt our list of required information. To answer question 1, we need to know the following:

1. There exists a variable which
	1. causes ;
	2. acts as a switch for ;
	3. does not directly cause and does not cause any causes of except those on the path ;
	4. is statistically independent of any variable not on the path that causes .

and analogously for question 2.

So, concretely, what information does our social scientist need to gather in order to answer to demands 1\*) and 2\*)? For conciseness’ sake, I will focus only on 1\*) here, i.e. on finding . The social scientist must find a variable which, firstly, causes the economic recession; secondly, acts as a switch for the economic recession (i.e. makes whether the recession occurs independent of any other variables); thirdly, does not directly or through a path not on cause the drop in participation in the night time economy; fourthly, is statistically independent of any variable not on the path that causes the drop in participation in the night time economy. Thus, to find , the social scientist needs to ask herself: could we have prevented the economic recession from happening, in a way that is in no way connected to the night time economy through a different route? And would people participate more in the night time economy if we prevented the recession in this way?

We clearly see the connection between Woodward’s interventionist framework and the process-tracers’ method break down at this point. A process-tracer interested in the causal connection between and who follows methodologists like George, Bennett, and Checkel is not concerned with finding interventions. Rather, what she does is investigate whether there are observable implications of all three factors (economic recession, drop in participation in the night time economy, drop in total violence) present in some case study. So, she may ask the people living in e.g. the London borough of Hackney whether they go out more or less since the crisis (reasoning that if participation in the night time economy dropped, then these people would confirm they went out less); she may also ask them whether they have experienced violent behaviour (reasoning that if violent crime dropped, then these people would say that they experienced violent behaviour less often). What she is not requiredto do,if we take methodological advice from George, Bennett, and Checkel seriously, is come up with an intervention variable. Thus, she will not prove that , the economic recession, is a contributing cause to , the drop in total violence.

On the other hand, whereas a process-tracer can get away with merely noting that the intervening variables in the mechanism are ‘instantiated’ in some case study, Woodward requires one to supplement this with evidence for *intervention variables* for each link of the mechanism. If we commit to Woodward’s framework, this spells trouble for process-tracing methodologists like George, Bennett, and Checkel.

### 4.1 Woodwardian recommendations for process-tracers

Looking ahead, what advice would the Woodwardian philosopher give to a process-tracer? For both practical and ethical reasons, social scientists may wish to refrain from intervening in social science scenarios themselves. This however does not stand in the way of the Woodwardian process-tracer; as we have seen in section 2, an intervention does not actually need to involve human action. Therefore, one of the ways in which a process-tracer may find an intervention variable is by looking at two distinct case studies, which are as alike as possible, except in regards to independent variable .

So, thinking back to our last example, the social scientist should look for a region in which the economic recession did not happen (or perhaps, practically speaking, in which the recession had less of an impact), and another region in which the recession did occur. We must make sure that the reason these two regions differ with regards to the impact of the economic recession is not itself affecting the participation in the night time economy, except through affecting the impact of the economic recession; as an extreme example, if the economic recession had less of an impact on a region because the region was mainly populated by Amish, we would expect the Amish beliefs to be the reason the region’s inhabitants do not participate in the night time economy, regardless of the economic recession. Thus, the process-tracer should study at least *two* cases, a ‘control case’ and an ‘experimental case’, and justify that these two cases are sufficiently similar.

**4.1.1 Sufficiently similar? Causal homogeneity in process-tracing**

This brings us to one issue with process-tracing highlighted in the introduction that deserves further study. There is a tension between the methodology of using singular case study evidence on the one hand, and the aim of making general causal claims on the other. For the social sciences this tension is difficult to resolve: arguably, it requires a causal relevance assumption ([cf. Hitchcock 1995](#_ENREF_8)) for the set of events generalized over. Christopher Hitchcock argues that singular causal claims (here produced in case study research) can only be used as evidence for a general causal claim if one can demonstrate that the causal relevance of the cause for the effect is the same from individual to individual. Applying this causal relevance criterion to social science, to move from singular case studies to general theories one needs a ‘homogeneity assumption’: the assumption that in both the target cases and the case under study, ceteris paribus the causal relevance of the cause for the effect is the same.

Consider the following example ([following Bakke 2013](#_ENREF_1)). Whilst process-tracing political scientist Kristin Bakke looks for evidence for the singular causal claim that the presence of transnational insurgents in Chechnya during the Second Chechen War caused the radicalization of war tactics there, she wants to make the stronger claim that the presence of transnational insurgents in civil conflict more generally causes radicalization. Here, the homogeneity assumption is that in both the Second Chechen War and in all other civil conflicts, all other things being equal the causal relevance of transnational insurgents for the radicalization of war tactics would be the same. That is, the presence of transnational insurgents would raise the probability of radicalization by the same amount in all civil conflicts. This homogeneity assumption is difficult, if not impossible to defend. I do not have time to consider this assumption in more detail here, but note that it deserves further attention.

# 5. Conclusion

In this paper I have shown that process-tracers generally postulate causal hypotheses which relate a cause and effect by a path consisting of intervening variables . They then find a case study in which both and are present, and investigate whether are also present. Woodward defines that is a *contributing* cause of with respect to if and only if there was a set of intervening variables such that is a direct cause of , which in turn is a direct cause of , which is a direct cause of … , which is a direct cause of . As it stands, process-tracing does not establish the complete right hand side of this if and only if statement. Process-tracers show that a set of intervening variables exists, but they do not show that each link of the chain is a relation of direct causation. If they commit to Woodward’s notion of causation, process-tracers have to provide evidence that there is a possible intervention to show that the relations they hypothesize are genuinely causal.

References

Bakke, Kristin M. 2013. "Copying and Learning from Outsiders? Assessing Diffusion from Transnational Insurgents in the Chechen Wars." In *Transnational Dynamics of Civil War*, edited by Jeffrey T. Checkel, 31-62. Cambridge: Cambridge University Press.

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

Brady, Henry, and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. 2nd ed. Lanham, MD: Rowan & Littlefield.

Crasnow, Sharon. 2012. "The Role of Case Study Research in Political Science: Evidence for Causal Claims." *Philosophy of Science* no. 79 (5):655-666.

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

Hall, Peter. 2012. "Symposium: Tracing the Progress of Process Tracing." *European Political Science* no. 11:20-30.

Hedström, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* no. 36:49-67.

Hitchcock, Christopher. 1995. "The Mishap at Reichenbach Fall: Singular vs. General Causation." *Philosophical Studies* no. 78:257-291.

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

Mahoney, James. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* no. 16 (3):575-593.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge / New York: Cambridge University Press.

Tilly, Charles. 2001. "Mechanisms in Political Processes." *Annual Review of Political Science* no. 4:21-41.

Woodward, James. 2003. *Making Things Happen*. Oxford: Oxford University Press.

1. This issue is closely related to the notion of ‘faithfulness’, employed amongst others by Spirtes, Glymour, and Scheines. These ‘washing out’ cases are cases when faithfulness, defined as being able to read off all causal independence relations off probabilistic (conditional) independence, fails. [↑](#footnote-ref-1)
2. This second requirement is meant to sort out cases where transitivity of a causal relation fails. To illustrate, such a case, imagine that whilst having breakfast I spill coffee on my navy blazer ($C$), which causes me to wear a cream blazer instead ($B=c$ rather than $B=n$). Now, it turns out that at my job interview for a fashion editor position that afternoon, wearing a blazer rather than not wearing a blazer (i.e. in this scenario $B=c$ rather than $B=0)$ causes me to get the job $(J)$. However, despite requirement (i) being satisfied (after all, there is a directed path $C\rightarrow B\rightarrow J$), we would hardly say that my spilling coffee at breakfast ($C$) causes me getting the job ($J$). The causal relation is not transitive. This failure of transitivity is captured by requirement (ii): there is no intervention on my spilling coffee that will change whether I get the job. If I don’t spill the coffee, I will wear my navy blazer instead. [↑](#footnote-ref-2)
3. I concede that this requires one to specify what ‘similar’ would mean in this context, bringing up such issues as external validity and the reference class problem. I will discuss this in more detail in section 4.1.1. [↑](#footnote-ref-3)
4. That is, a drop in spending at e.g. pubs and nightclubs. The intuition behind this proposed mechanism is that if someone has less money to spend at pubs and nightclubs they will, all other things being equal, consume less alcohol, and therefore be less likely to partake in violent drunken behaviour, which is one of the main forms of non-domestic violence. [↑](#footnote-ref-4)