



Causal discovery and the problem of psychological interventions

Markus I. Eronen

Theory and History of Psychology, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, Netherlands

ARTICLE INFO

Keywords:
Causality
Causal discovery
Robustness
Interventions
Interventionism

ABSTRACT

Finding causes is a central goal in psychological research. In this paper, I argue based on the interventionist approach to causal discovery that the search for psychological causes faces great obstacles. Psychological interventions are likely to be fat-handed: they change several variables simultaneously, and it is not known to what extent such interventions give leverage for causal inference. Moreover, due to problems of measurement, the degree to which an intervention was fat-handed, or more generally, what the intervention in fact did, is difficult to reliably estimate. A further complication is that the causal findings in psychology are typically made at the population level, and such findings do not allow inferences to individual-level causal relationships. I also discuss the implications of these problems for research, as well as various ways of addressing them, such as focusing more on the discovery of robust but non-causal patterns.

1. Introduction

A key objective in psychological research is to distinguish causal relationships from mere correlations (Kendler & Campbell, 2009; Pearl, 2009; Pearl & Mackenzie, 2018; Shadish & Sullivan, 2012). For example, psychologists want to know whether having negative thoughts is a cause of anxiety instead of just being correlated with it: If the relationship is causal, then the two are not just spuriously hanging together, and intervening on negative thinking is one way of reducing anxiety in patients suffering from anxiety disorders. However, to what extent is it actually possible to find psychological causes?

In this paper, I will point out and discuss several obstacles to discovering psychological causes that have not been adequately discussed in the psychological and philosophical literature. First, I will highlight a crucial but often neglected distinction: The distinction between non-psychological and psychological interventions, which create very different contexts for causal inference. Second, I will argue that the latter context, discovery of psychological causes, is deeply problematic.

I will formulate my arguments in the framework of the interventionist theory of causation. It is a theory of causation that aims at elucidating the role of causal thinking in science, and defining a notion of causation that captures the difference between causal relationships and mere correlations in a way that reflects scientific practice (Pearl, 2000, 2009; Spirtes, Glymour, & Scheines, 2000; Woodward, 2003; Woodward & Hitchcock, 2003). Several authors have also argued that interventionism adequately captures the role of causal thinking and reasoning in psychological research (Campbell, 2007; Kendler &

Campbell, 2009; Rescorla, 2018; Woodward, 2008a). Due to its conceptual clarity regarding the notions of causation and intervention, interventionism is exceptionally well suited for highlighting the problems that are the focus of this paper. However, as I will argue, these problems are not just restricted to interventionism, but should be seen as general challenges to the discovery of psychological causes.

The structure of this paper is as follows. I will start by giving a brief introduction to interventionism, and then turn to problems of interventionist causal inference in psychology: First, to problems related to psychological interventions (section 3), and then to problems arising from the requirement to “hold fixed” possible confounders (section 4). After this, I will show how these problems appear in some of the most common approaches to causal discovery in psychological research (section 5). In the last section, I consider different ways forward and various implications that my arguments have for psychology.

2. Interventionism

The guiding idea of interventionism is that causes are “difference-makers” for their effects: Causal relationships (unlike correlations) are such that we can manipulate or intervene on the cause to bring about a change in the effect (Woodward, 2003). As an intuitive example, if giving feedback to students is a cause of improved learning outcomes (and the two are not merely correlated), then interventions that increase feedback in a given population should result in improved learning outcomes.

Interventionism comes in two main forms: On the one hand, there is

E-mail address: m.i.eronen@rug.nl.

<https://doi.org/10.1016/j.newideapsych.2020.100785>

Received 20 June 2019; Received in revised form 11 February 2020; Accepted 13 February 2020

Available online 4 March 2020

0732-118X/© 2020 Elsevier Ltd. All rights reserved.

the approach developed by [Spirtes et al. \(2000\)](#) and [Pearl \(2000, 2009\)](#), which provides a theory of representing causal relationships as directed acyclic graphs (DAGs) and modeling them as causal Bayes nets. This framework provides tools for estimating causal effects between variables, and for inferring causal graphs from probability distributions. On the other hand, there is the more philosophical approach developed by James [Woodward \(2003\)](#), which focuses more on elucidating the conceptual relationship between various causal notions and capturing the basic methodological requirements for inferring causal knowledge ([Woodward, 2015, b](#)). I will mainly rely on Woodward's version of interventionism in this paper, as its conceptual focus allows for clearly spelling out the problems that I will discuss.

More specifically, Woodwardian interventionist causation is defined as follows:

(M) X is a cause of Y (in variable set V) if and only if it is possible to intervene on X to change Y when all other variables (in V) that are not on the path from X to Y are held fixed to some value ([Woodward, 2003](#)).

Thus, the idea is that in order to establish that X is a cause of Y , we need evidence that it is possible to intervene on X to change Y , when off-path variables are held fixed.¹ Importantly, it is not necessary to actually perform such an intervention in order to establish that X is a cause of Y : Knowledge about the behaviour the system under interventions can also be gained indirectly, for example based on observational data (I return to this in section 5).

The notion of an intervention plays a fundamental role in the account. According to [Woodward \(2003\)](#), interventions have to satisfy the following conditions.

Variable I is an intervention variable for X with respect to Y if and only if:

- (I1) I causes X ;
- (I2) I acts as a switch for all other variables that cause X . That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I .
- (I3) There is no directed causal path² from I to Y that does not go through X ;
- (I4) I is statistically independent of any variable Z that causes Y and is on a directed causal path that does not go through X .

The rationale behind these conditions is that if the intervention does not satisfy them, then one is not warranted to conclude that the change in Y was (only) due to the intervention on X . In a nutshell, the conditions state that the intervention should change the value of the putative cause variable X in such a way that the change in Y is *only* due to the change in X and not any other influences ([Woodward, 2015, b](#)). Following standard terminology in the literature, I will call interventions that satisfy the conditions I1–I4 *ideal* interventions.

In order to make the idea of interventionism more concrete, consider the following example (adapted from [Woodward, 2003](#)). A drug trial is

¹ More precisely, this is the definition for a *contributing* cause. X is a *direct* cause of Y if and only if it is possible to intervene on X to change Y when all other variables (in V) are held fixed to some value ([Woodward, 2003](#)). Thus, the definition of a contributing cause allows there to be other variables on the causal path between X and Y , whereas the definition of a direct cause does not. This does not reflect any substantive metaphysical distinction, as the question whether X is a direct or contributing cause is relative to what variables are included in the variable set. Importantly, the notion of a contributing cause is *not* relative to a variable set in any strong sense – if X is a cause of Y in some variable set, then X will be a cause of Y in all variable sets where X and Y appear ([Woodward, 2008b](#)). This is because the definition of an intervention is not relativized to a variable set.

² A directed causal path from A to B is a chain of causal relationships from A to B that all point in the same direction, for example $A \rightarrow P \rightarrow Q \rightarrow R \rightarrow B$, where the arrows represent causal relationships.

conducted with the aim to discover whether a new drug causes patients to recover from a psychotic episode. Let us denote administering the drug with variable D (value 1 = drug administered, value 0 = no drug administered) and recovery from the psychotic episode with variable R (value 1 = recovery, value 0 = no recovery). Following (M), we need to find out whether there is a change in R when we intervene on D , while holding fixed all other variables that are not on the path from D to R . In practice, this is typically done through a randomized controlled trial (RCT): Patients are randomly assigned either to a treatment group where they receive the drug ($D = 1$), or to a control group where they receive a placebo ($D = 0$). When the groups are large enough and the randomization is done correctly, the groups are on average similar to each other, the only difference being that one group received the drug and the other not. This corresponds to “holding fixed” all variables except D and R (and the variables on the causal path between them). The intervention I on D , that is, administering the drug, should satisfy conditions (I1)–(I4) – for example, the patients should not take the drug on their own accord (which would violate I2), and giving the drug (e.g., in syrup form) should not affect recovery in ways that are not due to the drug itself, but some factors (e.g., other healthy ingredients in the syrup; this would violate I3).

In psychology, interventionist theories of causation have not yet found their way to the mainstream. More popular approaches to causation include Rubin's causal model (RCM) and Campbell's causal model (CCM), of which the later has been particularly influential in psychology ([Shadish & Sullivan, 2012](#); [West & Thoemmes, 2010](#)). However, these approaches are to a large extent compatible and complementary to interventionism, as they have a different focus: Campbell's causal model is mainly concerned with principles of experimental design and dealing with threats to validity, whereas the Rubin's causal model provides a mathematical framework for causal inference based on (primarily) RCTs. In section 5, I will return to these approaches and show how the problems discussed in this paper also apply to them.³

Before moving on to problems of psychological interventions, some important distinctions need to be made. The first one is the distinction between *individual-level* causes and *population-level* causes. The first refers to causal relationships that hold for a particular individual: for example, John's negative thoughts cause John's problems of concentration. The latter refers to causal relationships that obtain in a population as a whole: for example, negative thoughts cause problems of concentration in a population of university students. It is widely thought that the ultimate goal of causal inference is to find individual-level causes, and that a population-level causal relationship should be seen as just an average of individual-level causal relationships ([Holland, 1986](#)). For example, the causal relationship between negative thoughts and problems of concentration in a population of university students is only interesting insofar as it *also* applies to at least some of the individual students in the population.⁴ I will return to the relationship between population- and individual-level causal inference in section 4.

Importantly, the distinction between population-level and individual-level causation is different from the distinction between type

³ There are of course many other approaches to causation in philosophy (see, e.g. [Beebe, Hitchcock, & Menzies, 2009](#)). For example, process theories take the distinction between causal processes and non-causal processes to be the key to understanding causation (e.g., [Dowe, 2000](#)); the counterfactual theory of David Lewis aims at defining causation in terms of non-causal counterfactuals (e.g., [Lewis, 1973](#)); and capacity theories treat the notion of causal powers or capacities as more fundamental than causal relationships (e.g., [Cartwright, 1989](#); [Mumford & Anjum, 2011](#)). For an overview of the relative weaknesses and problems of interventionism, see for example [Reutlinger \(2013\)](#).

⁴ [Borsboom, Mellenbergh, and Van Heerden \(2003\)](#) argue that there can be genuine between-subject causes, such as IQ differences between individuals, that are not causes that act within individuals (see also [Weinberger, 2015](#) for a rebuttal). However, this is a special category of causes that I will not discuss further in this paper.

and token causation, even though the two distinctions are sometimes mixed up in the philosophical literature (see also Illari & Russo, 2014, ch. 5). Token causation refers to causation between two actual events, whereas type causation refers to causal relationships that hold more generally. Individual-level causes can be either type causes or token causes. An example of an individual and type causal relationship would be “John’s pessimistic thoughts cause John’s problems of concentration”: This is a general relationship between two variables, and not a relationship between two actual events. An example of an individual and token causal relationship would be “John’s pessimistic thoughts before the exam on Friday at 2 p.m. caused his problems of concentration in the exam”. As interventionism is a type-level theory of causation, and the aim of psychological research is primarily to discover regularities, not explanations to particular events, in this paper I will only discuss the discovery of type causes.

I will now go through various problems in performing (ideal) interventions in psychology, starting from problems related to conditions I2 and I3 (section 3), and then turn to problems related to I4 and the “holding fixed” part of the definition of causation (section 4). Although these problems will be described in a way that is directly related to the interventionist conditions, they should not be seen as just internal problems for interventionism. Interventionism provides an intuitive and well-structured framework for describing and discussing these problems, but as I will show, they also arise in different forms for other approaches to causal inference in psychology.

3. Psychological interventions

In order to explain the problem of psychological interventions, one more key distinction has to be introduced first: The distinction between *psychological* and *non-psychological* causes. That is, there are on the one hand relationships where (1) the cause is *non-psychological*, and the effect is psychological, and on the other hand relationships (2) where the cause (and possibly also the effect) is *psychological*.⁵ Many classical experimental setups in psychology involve relationships of the first kind. In these experiments, an intervention targets a non-psychological variable, and the psychological effect of the manipulation of this non-psychological variable is tracked. Consider, for example, the Stroop task, the Wason task, or inattentive blindness experiments (e.g., the missing gorilla; Simons & Chabris, 1999). In educational psychology, experiments often involve comparing different learning materials and measuring their effects on learning. In clinical psychology, we find experiments that aim at discovering the causal influence that different medications have on recovery (e.g., from depression). In all of these examples, the putative causal relation is between a non-psychological cause variable (X) and a psychological effect variable (Y). Therefore, making the right kinds of interventions on the putative cause variable X is in principle not more difficult than in other fields (although of course far from trivial). As regards the psychological effect variable (Y), there is no need to intervene on it; it is enough to measure the change in Y (which, again, is far from trivial, but faces just the usual problems in psychological measurement, which will be discussed below). The fact that many psychological experiments involve this kind of causal relationships may have contributed to the recent optimism on the prospects of (interventionist) causal inference in psychology.

However, psychological research also often concerns relationships of the second kind, that is, relationships where the *cause* is psychological. This is, for example, the case when the aim is to uncover psychological mechanisms that explain cognition and behavior (e.g., Bechtel, 2008;

⁵ The line between psychological and non-psychological variables is likely to be blurry. However, for the present purposes it is not crucial where exactly the line should be drawn: My arguments apply to cases where it is clear that the cause variable is psychological (such as the examples in the main text), and such cases abound in psychological research.

Piccinini & Craver, 2011), or to find networks of causally interacting emotions or symptoms (e.g., Borsboom & Cramer, 2013). The reason why these relationships are crucially different from relationships of the first kind is that now the variable intervened upon is psychological, so the conditions on interventions now have to be applied to psychological variables.

Ideal interventions on psychological variables are rarely if ever possible. One reason for this has been extensively discussed by John Campbell (2007): Psychological interventions seem to be “soft”, meaning that the value of the target variable X is not completely determined by the intervention (Eberhardt & Scheines, 2007; see also Kendler & Campbell, 2009; Korb & Nyberg, 2006). In other words, the intervention does not “cut off” all other causes of X and fully take control of the value of X . As a non-psychological example, when studying shopping behaviour during one month by intervening on income, an ideal intervention would fully determine the exact income that subjects have that month, whereas simply giving the subjects an *extra* 5000€ would count as a soft intervention (Eberhardt & Scheines, 2007). Similarly, if we intervene on John’s psychological variable *alertness* by shouting “WATCH OUT!”, this does not completely cut off the causal contribution of other psychological variables that may influence John’s *alertness*, but merely adds something on top of those causal contributions (Campbell, 2007). As most or all interventions on psychological variables are likely to be soft, Campbell proposes that we should simply allow such soft interventions in the context of psychology. Campbell argues that these kind of interventions can still be informative and indicative of causal relationships (Campbell, 2007), and this conclusion is supported by independent work on soft interventions in the causal modelling literature (e.g., Eberhardt, 2014; Eberhardt & Scheines, 2007; Korb & Nyberg, 2006).

However, the problem of psychological interventions is not solved by allowing for soft interventions. There is a further, equally important reason why interventions on psychological variables are problematic: Psychological interventions are typically causes for *several psychological variables*, not just the intervention target. For example, suppose we wanted to find out whether *pessimistic thoughts* cause *problems in concentration*.⁶ In order to do this, we would have to find out what would happen to *problems in concentration* if we were to intervene just on *pessimistic thoughts* without perturbing other psychological states with the intervention. However, how could we intervene on *pessimistic thoughts* without changing, for example, *sadness* or *feelings of guilt*? As an actual scientific example, consider a network of psychological variables that includes, among others, the items *alert*, *happy*, and *excited* (Pe et al., 2015). How could we intervene on just one of those variables without changing the others?

One reason why performing “surgical” interventions that only change one psychological variable is so difficult is that there is no straightforward way of manipulating or changing the values of psychological variables (as in, for example, electrical circuits or drug trials). Interventions in psychology have to be done, for example, through verbal information (as in the example of John above) or through visual/auditory stimuli, and such manipulations are not precise enough to manipulate just one psychological variable. Also, state-of-the-art neuroscientific methods such as transcranial magnetic stimulation affect relatively large areas of the brain, and are not suited for intervening on specific psychological variables. Currently, and in the foreseeable future, there is no realistic way of intervening on a psychological

⁶ One could also argue that these variables cannot be part of a well-formed causal variable set, because they are not conceptually independent or independently manipulable (cf. Woodward (2015a,b)). This is an additional obstacle to causal inference in psychology, but one that has already been discussed in the literature (e.g., Campbell 2007; Woodward (2015a,b)). The problem of psychological interventions that I discuss below is more general, as it also applies to psychological variable sets where the variables are in fact independently manipulable.

variable without at the same time perturbing some other psychological variables.

An additional complication is that it is difficult to check what a psychological intervention precisely changed. In fields such as biology or physics there are usually several independent ways of measuring a variable: for example, temperature can be measured with mercury thermometers or radiation thermometers, and the firing rate of a neuron can be measured with microelectrodes or patch clamps. However, measurements of psychological variables, such as emotions or thoughts, are based on self-reports, and finding independent ways of verifying that these reports are correct is extremely challenging. Moreover, only a limited number of psychological variables can be measured at a given time point, so an intervention may always have unforeseen effects on unmeasured (and potentially confounding) variables.

Thus, it is likely that most or even all psychological interventions are not specific enough to only cause the target variable X , but are also causes for other variable(s) in the system. I will call such interventions *fat-handed*⁷ interventions (see also Baumgartner & Gebharter, 2016; Eberhardt & Scheines, 2007; Scheines, 2005).⁸ For example, an intervention on *pessimistic thoughts* that also causes *feelings of guilt* would be fat-handed.

Why are fat-handed interventions problematic for causal inference? The reason becomes clear when looking at condition I3: The intervention should not change any variable Z that is on a causal pathway that leads to Y (except, of course, those variables that are on the path from X to Y). This condition can sometimes be satisfied also for fat-handed interventions, but showing that it is satisfied requires knowing the causal structure of the system under study, as well as the changes that the intervention causes. The problem is that in the context of intervening on psychological variables, neither the causal structure nor the exact effects of the interventions are known. Thus, when the intervention is fat-handed, it is not known whether I3 is satisfied or not, and in many cases it is likely to be violated. In other words, we cannot assume that the intervention was an unconfounded manipulation of X with respect to Y , and cannot conclude that X is a cause of Y .

Fat-handed interventions have been recently discussed in philosophy of science, but mainly in the context of mental causation (e.g., Baumgartner & Gebharter, 2016; Romero, 2015). It is widely accepted that mental states depend on or “supervene” on brain states (i.e., there can be no difference in mental states without there being some difference in the underlying brain states), but this means that it is not possible to intervene on mental states without also changing some brain states, and therefore interventions on mental states are systematically fat-handed. One proposed solution for dealing with this kind of fat-handedness is to modify condition I3 (and other relevant conditions) so that they allow for changes in variables that are *conceptually* or *non-causally* related to X , for example due to supervenience (Woodward, 2015, a). However, it is important to emphasize that the fat-handedness that I discuss in this paper is different in kind. It is not due to the relationship between the mind and the brain (e.g., supervenience), but due to the fact that there is no straightforward way of manipulating single psychological variables. Therefore, even if we modify condition I3 to allow for conceptual relationships between the variables, this problem remains.

As an illustration of how the problem of fat-handedness plays out in psychological research, consider experimental designs where feelings of loss of control are manipulated (e.g., Whitson & Galinsky, 2008). The theory behind these experiments is that the feeling of loss of control is causally related to the perception of illusory patterns. Basically, the theory posits that when someone feels that she has no control over her life, she unconsciously compensates for this by perceiving patterns that are not actually there (van Elk & Lodder, 2018; Whitson & Galinsky, 2008). In order to test this, experiments have been conducted where the feelings of loss of control of the participants are manipulated (Whitson & Galinsky, 2008). This can be done in several different ways: For example, in one experiment Whitson and Galinsky (2008) asked participants to recall an experience in which they lacked control over a situation, and in another experiment, gave participants inconsistent feedback on a task that they were performing. However, it is likely that such interventions on feelings of loss of control also change other feelings or mental states. In other words, it is likely that they are fat-handed. For example, recalling an experience in which one lacked control over the situation may also induce feelings of anxiety or hopelessness, and receiving inconsistent feedback might result in frustration or anger.

It is common to apply “manipulation checks” (e.g., questionnaires) to control whether the manipulation did what it was intended to do (in this case, whether feelings of loss of control actually increased), but such checks only probe whether the targeted variable (and sometimes a limited number of other variables) actually changed. Thus, these experiments face the problem of psychological interventions: The intervention is likely to have been fat-handed, and the degree to which it was fat-handed is difficult to measure. Therefore, one should be very careful to draw any causal conclusions from these experiments (see also van Elk & Lodder, 2018, for a failed conceptual replication of the results of Whitson & Galinsky, 2008).

4. The problem of “holding fixed”

I will now turn to a second problem for causal inference in psychology. Whereas the previous problem concerned the requirements imposed on interventions, this one is related to the way causation is defined in (M): X is a cause of Y (in variable set V) if and only if it is possible to intervene on X to change Y when all other variables (in V) that are not on the path from X to Y are held fixed to some value. The motivation for this requirement of “holding fixed” is to make sure that the change in Y is really due to the change X , and not due to some other cause of Y .⁹ Fat-handedness is one way in which this condition can fail to be satisfied, as fat-handed interventions may change variables that are not on the path from X to Y . However, as I will now show, it is problematic in psychology also for reasons that are not due to fat-handedness.

In psychology, it is impossible to hold psychological variables fixed in any concrete way: We cannot “freeze” mental states, or ask an individual to hold her thoughts constant. Thus, the same effect has to be achieved indirectly, and randomized controlled trials (RCTs) are usually considered to be the gold standard for this (Woodward, 2003; 2008b; see, however, Cook, Scriven, Coryn, & Evergreen, 2010). As we saw

⁷ According to Scheines (2005), this term was coined by Kevin Kelly.

⁸ Some of these authors define fat-handed interventions more narrowly as interventions that change the putative effect Y not just through X but also through some other causal path(s). However, if we adopt this more narrow definition, we need another term for the broader notion that I discuss in this paper, that is, interventions that change multiple variables at the same time. Moreover, in practice it is often difficult to determine whether an intervention was fat-handed just in the broad sense or also in the narrow sense. Therefore, I use the broader notion in this paper. The narrower notion can then be seen as forming a subset of this broader notion: namely, those fat-handed interventions that are problematic for causal inference.

⁹ In recent publications, Woodward often gives a shorter definition of causation that does not include the “holding fixed” part, for example: “ X causes Y iff (i) it is possible to intervene on X and (ii) under some such possible intervention on X , changes in the value of X are associated changes in the value of Y ” (Woodward, 2015, b). This is understandable, as the definition of intervention already contains conditions I3 and I4, which effectively imply holding fixed potential causes of Y that are correlated with the intervention and are not on the path from X to Y . However, there are also good reasons why the full definition has to include the second component as well. For example, consider a situation where we intervene on X with respect to Y , and Y changes, but this change is fully due to a change in variable Z , which is a cause of Y that is *uncorrelated* with the intervention variable. In this situation, without the “holding fixed” requirement we would falsely conclude that X is a cause of Y .

above, the basic idea of RCTs is that when the groups are large enough and the randomization is done correctly, any difference observed in the outcome between the treatment and the control groups should be due to the treatment. This in principle has the same effect as holding fixed all variables that are not on the path from X to Y .

However, this methodology has an important limitation. As the effect of “holding fixed” is based on the difference between the groups on average, it only applies at the population level, and not at the level of individuals. For this reason, results of RCTs do not necessarily hold for particular individuals in the population (cf. Borsboom, 2005; Hamaker, 2011; Molenaar & Campbell, 2009). For example, if we discover in an experiment that feelings of loss of control are causally related to the perception of illusory patterns, it does not follow that this causal relationship holds in John, Mary, or any other specific individual in the study population. This is related to the “fundamental problem of causal inference” (Holland, 1986): Each individual in the experiment can belong to only one of the two groups (control or treatment group), and therefore half of the data is always missing, and instead of individual causal effects, only an (population) average causal effect can be estimated. What this implies for causal inference in psychology is that when a causal relationship is discovered in an RCT-style experiment, we cannot infer that this relationship holds for any specific individual in the population (see also Illari & Russo, 2014, ch. 5).

Population-level findings based on RCTs are certainly not uninformative or useless; the main point is rather that we currently have little understanding of when, to what extent and under what circumstances they also apply to specific individuals in the population. A tempting solution might be to simply look at the data more closely and find those individuals for whom the intervention on X actually corresponded with a change in Y . However, it would be a mistake to conclude that in those individuals the change in Y was caused by X . It might very well have been caused by some other cause of Y , as possible confounders were not held fixed at the individual level at all.¹⁰

This problem is not unique to psychology, but also naturally applies to other fields where RCTs are used, such as the biomedical sciences. Indeed, the problem has not gone unnoticed there: especially in the context of personalized medicine, the fact that RCTs are as such not enough to establish individual-level causal relationships has recently become a matter of discussion (e.g., de Leon, 2012). There is also increasing literature on single-case ($n = 1$) research designs, where the focus is explicitly on uncovering individual causal relationships through measuring individuals over time (e.g., before and after an intervention; Lillie et al., 2011; Shadish & Rindskopf, 2007). However, insofar as such designs aim at discovering *psychological* causes, they face the challenge that it is very difficult to achieve the effect of holding fixed psychological states in any other way than comparing population averages.

5. Psychological interventions and causal inference in psychology

To summarize, what I have argued so far is that interventionist causal inference in psychology faces several obstacles: (1) Psychological interventions are typically fat-handed (and soft): They change several variables simultaneously, and do not completely determine the value(s) of the variable(s) intervened upon. It is not known to what extent such interventions give leverage for causal inference. (2) Due to the nature of psychological measurement, the degree to which a psychological

¹⁰ Would it be possible for a causal relationship to hold at the population level, but not for any individual in the population? Probably not, if the relationship is genuine: Weinberger (2015) has argued that there has to be at least *one* individual in the population for whom the relationships holds. However, in the context of discovery, it is possible that a causal *finding* at the population level is just an artefact of heterogeneous causal structures at the individual level, and therefore does not apply to any individual in the population.

intervention was fat-handed, or more generally, what the intervention in fact did, is difficult to reliably estimate. (3) It is not clear how possible confounders could be held fixed at the individual level, and it is not known under what conditions population-level causal relationships also apply to individuals. Taken together, these issues amount to a formidable challenge for finding psychological causes.¹¹

To show that these problems are not just internal to interventionism, I will now discuss how they manifest in approaches to causal inference that are more familiar to psychologists. I will start with the two frameworks mentioned in section 2: Campbell’s causal model (CCM) and the Rubin’s causal model (RCM). After this, I consider the possibility of inferring causal relationships from observational data.

CCM is usually considered to be the most influential perspective to causal inference in psychology (Shadish & Sullivan, 2012; West & Thoemmes, 2010). It was introduced by Donald Campbell (1957) in the 1950s, and has since been further developed over many decades by Campbell and other scholars (e.g., Thomas Cook and William Shadish; Shadish, Cook, & Campbell, 2002). In contrast to interventionism, the focus in CCM is not on characterizing interventions and the conditions that warrant causal inferences, but rather on the *threats to the validity* of causal inferences. The idea is that the validity of causal conclusions is potentially threatened by a broad range of factors that can be related to the design, the sample, the statistical analyses used, and so on. It is the task of the researcher to identify and assess these threats, and reduce and eliminate them as far as possible. This can in the first place be done by designing good experiments, where also where the main emphasis of CCM lies. The starting point of CCM is randomized controlled trials, but (quasi-)experimental designs that allow causal inferences also include designs with no control group (e.g., repeated measures designs), designs where the control group is not randomly formed (e.g., selected from a population of untreated individuals), designs where participants self-assign to groups, and so on. The approach can also be extended to single-case experimental designs, where, for example, a person is measured repeatedly over time, before and after an intervention (see, e.g., Shadish & Rindskopf, 2007 for more).

One thing that is common to all such (quasi-)experiments, and a crucial point in the context of this article, is that they involve an intervention or manipulation of an independent variable. Therefore, the problem of psychological interventions is directly relevant to them: when the independent variable is psychological, it is very likely that the intervention on that variable will be fat-handed. Moreover, due to the problems in measuring and tracking changes in psychological variables, it is difficult to establish or estimate the degree to which the intervention was fat-handed. In the terminology of Shadish et al. (2002), the possibility of fat-handed interventions can be seen as a threat to validity. Likewise, the problem of “holding fixed” discussed in section 5 is a threat to the validity of causal conclusions, when conclusions about individuals are drawn based on group designs. Therefore, these challenges should be taken into account when the CCM framework is applied to look for psychological causes.

Similar considerations apply to Rubin’s causal model (RCM; Holland, 1986; Rubin, 1974, 2005), which is probably the most influential

¹¹ Baumgartner (2009; 2013; 2018) has argued that mental-to-physical supervenience makes it impossible to satisfy the Woodwardian conditions on interventions, and that if interventionism is modified to accommodate supervenience relationships (as in Woodward, 2015, a), the result is that any causal structure with a psychological cause becomes empirically indistinguishable from a corresponding structure where the psychological variable is epiphenomenal. If this reasoning is correct, it leads to a further (albeit more theoretical) problem for interventionist causal inference: Any empirical evidence for a causal relationship with a psychological cause is equally strong evidence for a corresponding epiphenomenal structure, and it is not clear which structure should be preferred and on what grounds. See also Eronen (2020) for more discussion.

approach to causation in psychology after CCM (Shadish & Sullivan, 2012; West & Thoemmes, 2010). RCM is more mathematically oriented than CCM, and provides tools for estimating the (average) causal effect from experimental data. Like CCM, also RCM takes the randomized controlled trial as the starting point for understanding causal inference, and the core of the model consists of tools for calculating the causal effect of the treatment (the experimental manipulation) based on potential outcomes of the treatment. As the problem of fat-handedness arises when the treatment targets a psychological variable, the distinction between psychological and non-psychological variables is crucially important when RCM is applied in psychology. If interventions are fat-handed and this is not accounted for, the observed change in the effect might not be due to the variable that was targeted (i.e., the treatment), in which case causal estimates calculated with RCM will be incorrect or biased. In discussions of applying RCM to psychology, this problem is not mentioned (e.g., Shadish & Sullivan, 2012; West & Thoemmes, 2010). Note that it is different from the problem of non-manipulable causes that has been widely debated in the context of RCM and elsewhere (e.g., race, species, the gravitational constant; Holland, 1986; Woodward, 2003): Psychological states are in principle manipulable (there is nothing conceptually inconsistent in manipulating thoughts, beliefs, etc.), the problem is rather that the fat-handed way in which they are typically manipulated results in problems for causal inference.¹²

So far, I have discussed causal inference based on experimental designs, but there are also methods to infer causal relationships from purely observational data, and such methods are receiving increasing attention in psychology.¹³ There is a broad range of techniques available, including the use of propensity scores (i.e., propensity scores are calculated and individuals with similar scores are matched and compared; Rubin, 2005; Rohrer, 2018), instrumental variables (e.g., Hogan & Lancaster, 2004) or causal discovery algorithms (e.g., Malinsky & Danks, 2018; Spirtes et al., 2000).

Prima facie, observational approaches may seem to avoid the problems of fat-handedness and “holding fixed”, as they do not involve interventions. However, this is not the case. A widely accepted principle is “no causes (or causal assumptions) in, no causes out” (e.g., Cartwright, 1989; Woodward, 2003). What this means is that reliable causal inference based on observational data needs as a starting point some initial knowledge of the causal structure of interest (e.g., in order to know which variables to statistically control for), or more general assumptions about the causal structure (e.g., the causal Markov condition or causal sufficiency; Malinsky & Danks, 2018), or preferably both. When it comes to finding psychological causes from observational data, both of these requirements are problematic: We usually have little if any initial knowledge of the causal structure among the variables of interest, and we cannot assume that the more general causal assumptions will hold.

As an example of the latter, a key assumption that comes up in different forms in all approaches to observational causal inference is *causal sufficiency*: that is, that there are no missing (i.e., latent or unmeasured) common causes for any pair of variables in the set (Malinsky & Danks, 2018; Zhang, 2006). The reason for this is that if there are such missing common causes, inferences concerning the causal relationship between those variables are likely to be incorrect. However, missing common causes is probably the norm rather than the exception when it

comes to inferences involving psychological causes. For example, if the variable set consists of, say, 16 emotion variables, how likely is it that *all* relevant emotion variables have been included? And even if all emotion variables that are common causes to other emotion variables are included, is it plausible to assume that there are no further cognitive or biological variables that could be common causes to some of the emotion variables? As similar questions can be asked for any context involving psychological variables, causal sufficiency is a very unrealistic assumption for psychological variable sets.¹⁴ The core of the problem is that observational methods require causal assumptions, and when the aim is to find psychological causes, these causal assumptions have to be about (other) psychological causes, which are difficult to reliably establish (as I have argued in sections 3 and 4).

To sum up, the problems of psychological interventions and “holding fixed” arise whenever the aim is to find psychological causes, regardless of what framework of causal inference is used. These problems need to be more explicitly tackled and discussed, both in psychology and its philosophy.

6. Discussion

In this paper, I have used the interventionist theory of causation to highlight problems in causal inference involving psychological causes. I have also argued that these problems are not just internal to interventionism, but challenges for psychological research in general, insofar as it aims at uncovering psychological causes.

The results of this paper have several implications for psychological research. First, the importance of the difference between psychological and non-psychological causes needs to be more widely acknowledged, and the problem of psychological interventions needs to be studied further in the context of causal modeling. More specifically, we need a better understanding of fat-handed interventions, and new models and formal tools for dealing with interventions that are fat-handed. This is a task for philosophers of science and causal modelers, and there is some recent work that may prove to be a useful starting point (see, e.g., Peters, Bühlmann, & Meinshausen, 2016).

Second, when individual methods or sources of evidence are insufficient or unreliable (as is the case with finding psychological causes), what is often needed is a broader perspective. In other words, more attention should be paid to seeking *robust evidence* for causal relationships, combining multiple *independent* sources of evidence. Such evidence can lead to a high degree of confidence even if the sources are individually relatively weak in their evidential strength (Eronen, 2015; Kuorikoski & Marchionni, 2016; Munafò & Smith, 2018; Wimsatt, 1981; 2007). For example, there is no single method or source of evidence that would be individually sufficient to establish that the anthropogenic increase in carbon dioxide is the cause for the rise in global temperature, but there is so much converging evidence from many independent sources that scientists are confident that this causal relationship exists. Similarly, if a causal relationship can be inferred with several independent methods (e.g., different experimental setups), this gives a degree of confidence to the relationship. In psychology, robustness has a venerable theoretical tradition going back to Cronbach and Meehl (1955) and Campbell and Fiske (1959), and it is also emphasized in the

¹² In RCM, the problem of “holding fixed” is discussed in the context of the “fundamental problem of causal inference” (see also section 4): An individual can only belong to the treatment or control group, but not both at the same time. Therefore, in the RCM approach it is acknowledged that individual causal effects can only be estimated if we make strong assumptions, for example, that the treatment effect was the same for all participants (West & Thoemmes, 2010).

¹³ Also the frameworks of RCM and CCM can be extended to cover observational settings; see Rubin, 2005 and Shadish et al., 2002.

¹⁴ There are also causal discovery algorithms that do not require assuming sufficiency. These algorithms take the possibility of missing common causes into account and explicitly represent the uncertainty when the direction or presence of an arrow cannot be reliably inferred due to hidden common causes (see, e.g., Colombo, Maathuis, Kalisch, & Richardson, 2012; Zhang, 2006). However, the more unmeasured common causes there are, the more uncertainty there will be. Thus, when there are many missing common causes, as is often likely to be the case with psychological variable sets, algorithms that allow violations of sufficiency will simply result in graphs where most or all of the causal relationships are unknown.

CCM framework (Shadish et al., 2002). However, in practice, it currently plays a minimal role in psychological research compared to criteria such as statistical significance or reliability (see Bringmann & Eronen, 2016 for more).

It is also an open question to what extent this integration of evidence can actually lead to sufficient evidence for psychological causal relationships. Therefore, we should also seriously consider a more radical conclusion: When it comes to relationships between psychological variables, a more realistic and fruitful aim might be the discovery of robust but *non-causal* patterns. This is supported by the history of psychology. On the one hand, there are very few uncontroversial examples of causal discoveries that involve psychological cause variables. Well-established causes in psychology (e.g., the visual illusions, psychotherapy, medication) are typically non-psychological variables that have psychological effects. On the other hand, many important discoveries in psychology are discoveries non-causal *patterns* or *phenomena* (Haig, 2013; Rozin, 2001; Tabb & Schaffner, 2017). Consider, for example, the celebrated discovery that people often do not reason logically when making statistical predictions, but rely on shortcuts, for example, grossly overestimating the likelihood of dying in an earthquake or terror attack (Kahneman & Tversky, 1973). In other words, when we reason statistically, we often rely on heuristics that lead to biases. The discovery of this phenomenon had nothing to do with methods of causal inference (Kahneman & Tversky, 1973), and its significance is not captured by describing causal relationships between variables. In fact, the causal mechanisms underlying the heuristics and biases of reasoning are still unknown. Similar examples abound in psychology: Consider, for example, groupthink or inattentive blindness. Of course, there are likely to be causal mechanisms that give rise to these phenomena, but the phenomena are highly relevant for theory and practice even when we know little or nothing about those underlying mechanisms (which is the current situation).

This, in combination with the challenges discussed in this paper, suggests that the discovery of psychological causes should perhaps not be seen as central for making progress in psychology. Instead, a more realistic and fruitful goal would be the discovery of patterns and phenomena that are robust and useful for theory and practice.¹⁵ In any case, if causal claims involving psychological causes are made, it is essential to be aware the problems and challenges involved.

Finally, one might wonder whether the problems I have discussed here are restricted to just psychology. Indeed, I believe that the arguments I have presented are more general, and apply to any other fields where there are similar problems with causal complexity and fat-handed interventions. There is probably a continuum, where psychology is close to one end of the continuum, and at the other end we have fields where ideal interventions can be straightforwardly performed and variables can be easily held fixed, such as engineering science. Fields such as economics and political science are probably close to where psychology is, as they also face deep problems in making (ideal) interventions and measuring their effects. Same holds for cognitive neuroscience: The problems of soft and fat-handed interventions and holding variables fixed apply just as well to brain areas as to psychological variables (see also Northcott, 2019). Thus, appreciating the challenges I have discussed here and considering possible reactions to them could also benefit many other fields besides psychology.

To conclude, I have argued in this paper that there are several serious obstacles to the discovery of psychological causes. As it is widely assumed in both psychology and its philosophy that the discovery of causes is a central goal, these obstacles need to be explicitly discussed, taken into account, and studied further.

¹⁵ In addition, as pointed out in section 3, the discovery psychological effects of non-psychological causes has led to many important findings and does not require making psychological interventions.

CRediT authorship contribution statement

Markus I. Eronen: Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing.

References

- Baumgartner, M. (2009). Interventionist causal exclusion and non-reductive physicalism. *International Studies in the Philosophy of Science*, 23, 161–178.
- Baumgartner, M. (2013). Rendering interventionism and non-reductive physicalism compatible. *Dialectica*, 67, 1–27.
- Baumgartner, M. (2018). The inherent empirical underdetermination of mental causation. *Australasian Journal of Philosophy*.
- Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, 67, 731–756.
- Bechtel, W. (2008). *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.
- Beebe, H., Hitchcock, C., & Menzies, P. (2009). *The Oxford handbook of causation*. Oxford: Oxford University Press.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203.
- Bringmann, L. F., & Eronen, M. I. (2016). Heating up the measurement debate: What psychologists can learn from the history of physics. *Theory & Psychology*, 26(1), 27–43.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297.
- Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik, & L. Schulz (Eds.), *Causal learning. Psychology, philosophy, and computation* (pp. 58–66). Oxford: Oxford University Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Clarendon Press.
- Colombo, D., Maathuis, M. H., Kalisch, M., & Richardson, T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *Annals of Statistics*, 294–321.
- Cook, T. D., Scriven, M., Coryn, C. L., & Evergreen, S. D. (2010). Contemporary thinking about causation in evaluation: A dialogue with. *American Journal of Evaluation*, 31(1), 105–117.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281.
- Dowe, P. (2000). *Physical causation*. New York: Cambridge University Press.
- Eberhardt, F. (2014). Direct causes and the trouble with soft interventions. *Erkenntnis*, 79(4), 755–777.
- Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. *Philosophy of Science*, 74, 981–995.
- van Elk, M., & Lodder, P. (2018). Experimental manipulations of personal control do not increase illusory pattern perception. *Collabra: Psychology*, 4(1).
- Eronen, M. I. (2015). Robustness and reality. *Synthese*, 192(12), 3961–3977.
- Eronen, M. I. (2020). Interventionism for the intentional stance: True believers and their brains. *Topoi*, 39(1), 45–55.
- Haig, B. D. (2013). Detecting psychological phenomena: Taking bottom-up research seriously. *American Journal of Psychology*, 126(2), 135–153.
- Hamaker, E. L. (2011). Why researchers should think “within-person”. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: Guilford Press.
- Hogan, J. W., & Lancaster, T. (2004). Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Statistical Methods in Medical Research*, 13(1), 17–48.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960.
- Illari, P., & Russo, F. (2014). *Causality. Philosophical theory meets scientific practice*. Oxford: Oxford University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kendler, K. S., & Campbell, J. (2009). Interventionist causal models in psychiatry: Repositioning the mind-body problem. *Psychological Medicine*, 39, 881–887.
- Korb, K. B., & Nyberg, E. (2006). The power of intervention. *Minds and Machines*, 16, 289–302.
- Kuorikoski, J., & Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, 83, 227–247.
- de Leon, J. (2012). Evidence-based medicine versus personalized medicine: Are they enemies? *Journal of Clinical Psychopharmacology*, 32(2), 153–164.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70, 556–567.
- Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *Personalized Medicine*, 8(2), 161–173.
- Malinsky, D., & Danks, D. (2018). Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(1), e12470.
- Molenaar, P., & Campbell, C. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18, 112–117.

- Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford: Oxford University Press.
- Munafò, M. R., & Smith, G. D. (2018). Robust research needs many lines of evidence. *Nature*, 553, 399–401.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.
- Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96–146.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.
- Northcott, R. (2019). Free will is not a testable hypothesis. *Erkenntnis*, 84(3), 617–631.
- Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., ... Kuppens, P. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, 3(2), 292–300.
- Peters, J., Bühlmann, P., & Meinshausen, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society B*, 78, 947–1012. <https://doi.org/10.1111/rssb.12167>.
- Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3), 283–311.
- Rescorla, M. (2018). An interventionist approach to psychological explanation. *Synthese*, 195, 1909–1940.
- Reutlinger, A. (2013). *A theory of causation in the social and biological sciences*. Dordrecht: Springer.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42.
- Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, 192(11), 3731–3755.
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2–14.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469), 322–331.
- Scheines, R. (2005). The similarity of causal inference in experimental and non-experimental studies. *Philosophy of Science*, 72(5), 927–940.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation*, 2007(113), 95–109.
- Shadish, W. R., & Sullivan, K. J. (2012). Theories of causation in psychological science. In H. Cooper, et al. (Eds.), *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 23–52). Washington, DC: American Psychological Association.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception*, 28(9), 1059–1074.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. New York: Springer.
- Tabb, K., & Schaffner, K. F. (2017). Causal pathways, random walks and tortuous paths: Moving from the descriptive to the etiological in psychiatry. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry IV: Nosology* (pp. 342–360). Oxford: Oxford University Press.
- Weinberger, N. (2015). If intelligence is a cause, it is a within-subjects cause. *Theory & Psychology*, 25(3), 346–361.
- West, S. G., & Thoenes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, 15(1), 18.
- Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science*, 322(5898), 115–117.
- Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination. In M. Brewer, & B. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124–163). San Francisco, CA: Jossey-Bass.
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University.
- Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford: Oxford University Press.
- Woodward, J. (2008a). Mental causation and neural mechanisms. In J. Hohwy, & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation* (pp. 218–262). Oxford: Oxford University Press.
- Woodward, J. (2008b). Response to Strevens. *Philosophy and Phenomenological Research*, 77, 193–212.
- Woodward, J. (2015a). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91, 303–347.
- Woodward, J. (2015b). Methodology, ontology, and interventionism. *Synthese*, 192, 3577–3599.
- Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, Part I: A counterfactual account. *Noûs*, 37(1), 1–24.
- Zhang, J. (2006). *Causal inference and reasoning in causally insufficient systems*. Doctoral dissertation. Carnegie Mellon University.