Synthese (2022) 200:194
https://doi.org/10.1007/s11229-022-03676-0

ORIGINAL RESEARCH

Variable relativity of causation is good

Veli-Pekka Parkkinen[1]

Received: 19 August 2021 / Accepted: 25 March 2022

Abstract
Interventionism is a theory of causation with a pragmatic goal: to define causal concepts that are useful for reasoning about how things could, in principle, be purposely manipulated. In its original presentation, Woodward's (2003) interventionist definition of causation is relativized to an analyzed variable set. In Woodward (2008), Woodward changes the definition of the most general interventionist notion of cause, contributing cause, so that it is no longer relativized to a variable set. This derelativization of interventionism has not gathered much attention, presumably because it is seen as an unproblematic way to save the intuition that causal relations are objective features of the world. This paper first argues that this move has problematic consequences. Derelativization entails two concepts of unmediated causal relation that are not coextensional, but which nonetheless do not entail different conclusions about manipulability relations within any given variable set. This is in conflict with the pragmatic orientation at the core of interventionism. The paper then considers various approaches for resolving this tension but finds them all wanting. It is concluded that interventionist causation should not be derelativized in the first place. Various considerations are offered rendering that conclusion acceptable.

Keywords Causality · Interventionism · Manipulability · Variable relativity

1 Introduction

James Woodward's interventionist theory of causation aims to provide practicable definitions of various causal concepts when the primary aim of causal reasoning is understood to be predicting outcomes of interventions, and the primary medium of representing causal relations is directed acyclic graphs (DAGs). In particular, the theory aims to provide an explicit definition of the notion of direct cause, which is presupposed

✉ Veli-Pekka Parkkinen
veli-pekka.parkkinen@uib.no

[1] Department of philosophy, University of Bergen, Sydnesplassen 12-13, 5020 Bergen, Norway

Published online: 03 May 2022

in constructing and interpreting causal DAGs, but treated as an undefined primitive in the frameworks that formally define DAGs and the rules of their use.

In the original presentation of the theory in Woodward (2003), the interventionist definitions of direct causation and general causal relevance are given with reference to a variable set.[1] This has spurred a criticism that interventionism makes causation representation-relative: a variable X may be a cause of Y in variable set $\mathbf{V}$, but not be a cause of Y in a different set $\mathbf{V}^*$ (Strevens, 2007). In response, Woodward has made various clarifications. Firstly, Woodward acknowledges that the concept of direct causation does exhibit variable relativity: as DAGs are defined over a variable set, and a causal interpretation of a DAG rests on the notion of direct causal relation between two variables, any definition of direct causation suitable for causally interpreting DAGs must have this feature (Woodward, 2008). Secondly, according to Woodward (2008), while this is true of direct causation as defined by interventionism, it is not true of the concept of contributing cause, or in other words, of the interventionist concept of general causal relevance. According to Woodward, X is a contributing cause of Y, and therefore a cause *simpliciter*, if and only if there exists a variable set in which the interventionist definition of contributing causation is satisfied with respect to X and Y (Woodward, 2008). In sum, direct causation is relativized to the choice of variables of interest, but the concept of contributing causation or plain causal relevance is derelativized by qualifying that the existence of a variable set, known or unknown, where the definition is satisfied, is what is required for one variable to be a cause of another.

In this paper, I argue that when contributing causation is derelativized, interventionism entails a distinction between two notions of unmediated causal relation that are not coextensional. That is, there are circumstances in which some variable is an unmediated cause of another in one sense, but not in the other sense. An example is given in Sect. 3. There is however no difference in the conditions under which one can establish that one variable is a cause of another in either sense, or in what claims about manipulability relations they entail about the variables within the analyzed variable set. These concepts only differ in that one of them entails claims about manipulability with indirect reference to variables not included in the analyzed variable set, while the other never entails such claims. Thus, while these concepts are not coextensional, they do not differ in any completely specified manipulability claims that they entail. The distinction between the two concepts therefore violates the interventionist slogan ”*no causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference*” (Woodward, 2003, p. 61), hereafter referred to as the *Manipulability Thesis* for short. This is demonstrated in Sects. 2 and 3. In Sects. 4 and 5, various possible approaches to solving this redundancy in interventionism are considered, and found wanting.

It is then argued in Sect. 6 that variable relativity of causation should simply be accepted: if the purpose of causal reasoning is to discover dependencies that can be exploited for manipulation and control through exogenous interventions, this is

[1] The variable-set relativity of causal concepts, as defined in (Woodward, 2003), has come to be known simply as "variable relativity" of causation in the literature commenting on (Woodward, 2003). I adopt this shorthand usage in this paper in order to be consistent with other literature on the topic, such as (Strevens, 2007) and (Woodward, 2008).

how causal concepts should work. The reason is that for a manipulability account of causation, causal concepts apply to local systems of dependencies identifiable only in relation to some embedding environment from which the system can be intervened on (Hitchcock, 2007, Kuorikoski, 2014, Pearl, 2000, Woodward, 2007). In other terms, applying causal concepts requires distinguishing between (1), a set of factors that one considers to be controllable by interventions, i.e. the ”inside” of a system that one asks causal questions about, (2), processes not governed by the system itself that would count as interventions on parts of the system, and, (3), known and unknown background conditions that are not considered to be controllable by interventions. Choosing which factors are contained within the system of interest, i.e. which variables are taken to be the plausible targets of interventions in the first place, will on occasion affect conclusions about manipulability relations, and thus conclusions about causality. Arguably, interventionist definitions of causal concepts ought to be relative to a variable set in order to maintain a link between manipulability and causation. Derelativization undermines this, by entailing two concepts of unmediated causal relation that are not coextensional, but which nonetheless do not entail different conclusions about manipulability relations in any analyzed variable set. Therefore, variable relativity of causation is a good thing for the interventionist, as concluded in Sect. 7.

2 Interventionism and variable relativity of causation

In *Making things happen*, James Woodward presents a two-part definition of type level causality that he calls ”manipulability theory” or (M) for short (Woodward, 2003, p. 59). The first component is a definition of *direct cause* (DC), i.e. an unmediated causal relation between two variables:

Direct Cause (DC) A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in $\mathbf{V}$ besides X and Y are held fixed at some value by interventions (Woodward, 2003, p. 55).

The second component relies on (DC) to define a notion of *contributing cause* (CC):

Contributing Cause (CC) A necessary and sufficient condition for X to be a [...] contributing cause of Y with respect to variable set $\mathbf{V}$ is that (i) there be a directed path from X to Y such that each link in this path is a direct causal relationship: that is, a set of variables Z_1, ..., Z_n such that X is a direct cause of Z_1, which is in turn a direct cause of Z_2 which is a direct cause of ...Z_n, which is a direct cause of Y, and that (ii) there be some intervention on X that will change Y when all other variables in $\mathbf{V}$ that are not on this path are fixed at some value. If there is only one path P from X to Y or if the only alternative path from X to Y besides P contains no intermediate variables (i.e., is direct), then X is a contributing cause of Y as long as there is some intervention on X that will change the value of Y, for some values of the other variables in $\mathbf{V}$ (Woodward, 2003, p. 59).

Woodward's "manipulability theory" (of causation) comprises the conjunction of (DC) and (CC), here introduced and labeled separately for later reference. These definitions make use of the notion of intervention. Briefly, an intervention on a putative cause variable X with respect to a putative effect Y is an exogenous manipulation of X that replaces other causes of X so that X's value (or probability distribution) is caused by the intervention only, does not cause Y through any path that does not go through X, and does not cause or probabilistically depend on any such off-path cause of Y. Causal relations are required to be *invariant* under interventions to some degree, i.e. there must be at least one pair of values of the cause such that when interventions vary the value of the cause between those values, the value of the effect variable or its probability distribution will also change (Woodward, 2003, pp. 69–70, chapter 6)[2]. Since interventions are themselves causes, these definitions do not provide a reductive analysis of causation. For interventionism, the fact that some variables X and Y are causally related is not determined by any underlying non-causal fact like probabilistic dependence, transfer of energy, or instantiation of laws, but by other *causal* facts. That interventionism nonetheless avoids vicious circularity is because these other causal facts only consider the possibility of manipulating X through a process that is in a suitable way external to the rest of the structure that embeds X and Y, or in other words, causal facts are presupposed in characterizing what is an intervention on X relative to Y, but these include no presuppositions about whether X is a cause of Y.

While (DC) is conceptually more basic than (CC) in the sense that (CC) is defined in terms of (DC), the appeal of interventionism is in many ways due to (CC), which describes a minimal criterion for general causal relevance. Any dependence between variables that qualifies as causal must satisfy (CC); for some variable X to be a cause at all, X must be a contributing cause of something. Direct causes are also contributing causes, per the definitions of (DC) and (CC): a direct causal relation is a causal relevance relation with no mediating causes between the relata. Based on this minimal criterion captured in (CC), one can define other causal concepts in terms of the kinds of manipulability relations those concepts track. For example, the concept of total cause is defined as a variable X that makes a difference to an effect Y when only X and no other variable is intervened on (Woodward, 2003, p. 51). Furthermore, one can make detailed comparisons between causal relations in terms of various other properties like sensitivity to background conditions or the specificity of the mapping between values of the cause and the effect variables (Woodward, 2010). The reason that (DC) nonetheless is conceptually prior to (CC) is that (CC) makes use of the notion of *directed path*—a sequence of causally connected variables—that is defined in terms of sequential direct causal relations between the variables on the path.

[2] Throughout the paper, interventions are understood as "hard" interventions that break X off of its other causes, as defined in Woodward (2003). "Soft" interventions that preserve all causal connections sometimes have distinct advantages for inference (Eberhardt & Scheines, 2007), but would not work for Woodward's definition of causation as invariance under interventions, because effects of a common cause will be dependent both when unmanipulated and when manipulated by "soft" interventions, thus appearing to be causally connected if causation is understood as invariance under "soft" interventions. Qualifying the meaning of invariance so that the dependence between causes and effects is not only required to persist, but to not be weakened (in terms of e.g. degree of correlation) under interventions, may allow defining causation as invariance under "soft" interventions. This is not investigated here.

Woodward's theory builds on the idea that a causal structure is a network of direct causal relations between variables that can be represented and reasoned about graphically using directed acyclic graphs (DAGs). This idea originates in the theory of causal Bayes nets—a type of DAG that connects causal structure to the structure of probabilistic dependencies in a set of variables (e.g. Pearl, 2000, Spirtes *et al.*, 2000). Such causal DAGs comprise a set of variables as its nodes, and a set of arrows (directed edges) connecting pairs of variables. To construct a causal DAG, one draws an arrow between each pair of variables that are connected as direct cause and effect. A causal DAG then describes aspects of the joint probability distribution over the variables such that this distribution conforms to the causal Markov condition, according to which each variable is independent of its non-effects given its direct causes. One can then read off statements about conditional (in-)dependencies between the variables from the graphical representation of their causal structure, or, in cases where all independencies are due to the Markov condition, infer qualitative causal structure from information about conditional (in-)dependencies between variables.

All this obviously requires clarity about the concept of direct cause, and this is what (DC) intends to provide: (DC) is meant to describe exactly under which conditions one should draw an arrow between two variables in a causal DAG (Woodward, 2008, p. 198). Once all direct causal relationships are determined, the resulting structure, together with the functional forms of the dependencies between the directly causally related variables, determine all the facts about contributing causal relationships or general causal relevance between variables. The last point about functional dependencies is important, as interventionist causation is not transitive (Woodward, 2003, pp. 57–59). Consider a simple graph $X \rightarrow Y \rightarrow Z$ that depicts a causal chain in which X is a direct cause of Y, which is a direct cause of Z, in the sense described by (DC). Each direct causal relation is associated with a function that describes how the values of the effect variable change in response to changes in the cause, $Y = F(X)$ and $Z = G(Y)$. Here X is a contributing cause of Z if and only if the composite function $Z = G(F(X))$ is such that it makes the value of Z sensitive to changes in the value of X (Woodward, 2003, p. 58). If not, then X is not a cause of Z even though X is a cause of Y, which is a cause of Z, because no changes in the value of X map to changes in the value of Z. While the latter situation is perhaps atypical in real-world causal structures, it is not ruled out by the interventionist definition of causal relevance. Hence, transitivity is not entailed by the definition. In cases where it is known or assumed that the dependencies between direct causes and effects compose in a way that renders indirect causes and effects dependent under some combination of interventions, all contributing cause relationships can be read off the graphical structure of direct causal relations, *as if* causation were a transitive relation. Such an assumption is mentioned later in the ongoing section, and again in Sect. 4, but purely in order to illustrate unrelated points. This paper does not take a stand on any substantive issues related to transitivity of causation.

The idea that causal concepts are primarily used for predicting the outcomes of interventions is meant to characterize causal reasoning more broadly than just the explicit use of DAGs. DAGs are simply the canonical medium for representing such manipulability relations. For interventionism, any representation of causal structure codifies claims about the outcomes of actual or hypothetical interventions on the causal relata. Conversely, according to Woodward, "each completely specified set of claims about what will happen to each of the various variables in some set under various possible manipulations of each of the other variables, singly and in combination, will correspond to a distinct causal structure" (Woodward, 2003, p. 61).

What is meant by the claim that a (representation of a) causal structure corresponds to a "completely specified" set of manipulability claims requires some clarification. As is evident from the quote just above, whether a set of manipulability claims that corresponds to a causal structure is completely specified or not is relative to a variable set. That is, a set of claims about manipulability relations between variables in a variable set $\mathbf{V}$ may be completely specified relative to $\mathbf{V}$ even if there exists an expanded variable set $\mathbf{V}^*$, $\mathbf{V}^* \supset \mathbf{V}$, such that additional claims about manipulability of variables $\mathbf{V}$ can be made with reference to some variables that are included in $\mathbf{V}^*$, but not in $\mathbf{V}$.

I also take Woodward’s formulation to straightforwardly mean that a completely specified set of manipulability claims must state for each variable in a variable set $\mathbf{V}$, what would happen to the value of that variable under every combination of interventions on the other variables, where minimally one of the other variables is intervened on. This intepretation is roughly in line, by analogy, with uses of the notion in other contexts, for example when a function is said to be completely specified only if it defines an output value for every possible input value. Moreover, I take this to include the requirement that for each causal relation in a causal structure over variables $\mathbf{V}$, such a completely specified set of manipulability claims must include a claim that explicitly states all the variables that must be subjected to interventions, for example to hold their values fixed, in order for interventions on the cause to change the effect. Note that this does not mean that the manipulability claims must state every background condition that is required to obtain for a manipulability relation between some variables X and Y to obtain. It merely requires that every enabling condition for the manipulability relation that can only come about as a result of some combination of *interventions* on other variables than X and Y is described so that those other variables are directly referenced. In other words, the manipulability claims associated with specific causal relations in a structure over $\mathbf{V}$ cannot be elliptical in the sense that they mention variables that would have to be controlled by interventions in order to render effects manipulable by their causes, without stating what those variables are.

To illustrate the last mentioned point, consider a hypothetical causal structure where X causes Y via two separate paths such that the effect of X on Y through one path is exactly cancelled by the effect of X on Y through the other path. Y will hence not be manipulable by interventions on X unless one simulatenously interferes with one of the paths to prevent the cancelling of the effect through the other path, i.e. a further intervention is required on at least one intermediate variable on one of the paths from X to Y, for Y to be manipulable by interventions on X. Contrast this to a distinction between causes and "ordinary" background conditions. Let us say, for example, that the position of a light switch on the wall is a cause of the room being lit or not. The manipulability relation associated with this causal relation is dependent on background conditions, such as the main electricity switch of the building being on. But for the lighting of the room to be manipulable by interventions on the light switch, the main switch simply needs to be on, no matter how that condition came to be. In the cancelling paths case, by contrast, intermediate variables between X and Y are not background conditions in the same sense. Namely, Y is never, under any conditions, manipulable by interventions on X unless at least one of the intermediate variables on one of the paths is also intervened on. I assume the notion of "completely specified set of manipulability claims" to entail that the claims comprising the set excplicitly state all such variables that must be controlled by additional interventions in order to render the effect variables in the corresponding structure manipulable by interventions on their causes. I take it that this is a reasonable interpretation of Woodward, because such a requirement is needed to ensure that knowledge of causality reliably associates with knowledge of how things can be manipulated, which is the overarching aim of interventionism. Without such a requirement, knowledge of a causal relation between a cause X and an effect Y would not necessarily translate to understanding of how, exactly, Y could be controlled by interventions on X.

These commitments are summarized in the *Manipulability Thesis*: ”*No causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference*” (Woodward, 2003, p. 61). The *Manipulability Thesis* reflects the pragmatic goal of interventionism; as a causal structure corresponds to a completely specified set of claims about manipulability relations, knowledge of a causal relation between two variables entails knowledge about what exactly must be intervened on in order to control the effect. This idea will be revisited in what follows, as it will be shown that interventionism entails a distinction between two causal concepts that in no context of application will differ in what completely specified manipulability claims they entail.

Interventionism has drawn criticism according to which the definitions comprising (M), by defining causation relative to a variable set, make causation itself relative to an inherently subjective choice of variable set (Strevens, 2007). Woodward has replied to such criticism by clarifying that the intended meaning of the definition of contributing causation in (M) is to characterize

> what it is for X to be correctly represented as a contributing cause of Y with respect to $\mathbf{V}$. Understood in this way, [what] (M) says is that X is "correctly represented as a contributing cause of Y with respect to $\mathbf{V}$" if there is a chain of direct causal relationships (a directed path) leading from X to Y and if when one fixes variables that are off that path at some value, an intervention on X changes the value of Y. One can then go on to say that X is a contributing cause of Y simpliciter [...] as long as it is true that there exists a variable set $\mathbf{V}$ such that X is correctly represented as a contributing cause of Y with respect to $\mathbf{V}$ (Woodward, 2008, p. 209).

So, X is a contributing cause as long as there exists a variable set, whether known or not, in which X would be correctly represented as a contributing cause according to (CC). This can be simplified by adding the existential quantifier in the definition of (CC) itself:

Derelativized Contributing Cause ($\mathbf{CC}_{DR}$) A necessary and sufficient condition for X to be a [...] contributing cause of Y is that *there exists* a variable set $\mathbf{V}$ such that [...rest of the definition as in (CC)].

According to (CC_{DR}), X is a contributing cause of Y if and only if there exists a variable set in which a possible intervention on X would change Y when all off-path variables are fixed at some values.

As for direct causation, Woodward maintains that variable relativity is inevitable. No explicit argument is given, but one can be constructed roughly as follows, a more detailed argument is given in Sect. 4. Consider a causal chain $X \rightarrow Y \rightarrow Z$ where the forms of the dependencies between the direct causes and effects are such that the value of Z is sensitive to interventions on X via changes in Y. If we at one point in time can only measure X and Z but can intervene on X, X will be identified as a direct cause of Z against the variable set $\{X, Z\}$ and an arrow should be drawn between X and Z in the corresponding graph. If at a later point in time we find ways to measure and intervene also on Y, X is identified as not a direct but a contributing cause of Z, and the graphical representation is a chain. In this sense, a direct cause relationship in one variable set might not be a direct cause relationship in an expanded variable set. Woodward does not see this as a problem, as even if facts about direct causal relations change when the variable set changes, all correct representations of contributing causation are preserved in such scenarios: the variable relativity of direct causation cannot lead to false ascriptions of causal relevance *simpliciter*.

In sum, given these clarifications, the concept of direct cause remains relativized to a variable set, but the concept of contributing cause (or just "a cause") is derelativized by existentially quantifying over variable sets in its definition.

3 Direct cause versus unmediated contributing cause

Consider a structure where one variable causes another both directly and through a path involving a third variable, so that the direct effect is exactly opposite to the effect through the indirect route. Such an example is discussed by Woodward (Woodward, 2003, p. 49), and is reproduced below.[3] The example considers a causal structure over $\{P, Q, R\}$ where the relations between the three variables are governed by the following equations,

$$Q = aP + cR \tag{1}$$

$$R = bP \tag{2}$$

where the coefficients a, b and c satisfy the equality $a = -bc$. The associated graphical model is shown in Fig. 1.

Assume now that the mediating variable R is unknown, such that we are considering the variable set $\{P, Q\}$. What can we say about the relation between P and Q? On the one hand, P does not count as a direct cause of Q, even if it does count as one relative to an expanded variable set that includes R, since no intervention on P will change Q and there are no other variables to control for in $\{P, Q\}$. On the other hand, P does count as a contributing cause of Q, because there exists a variable set $\{P, Q, R\}$ such

[3] Variable names have been changed from the original to avoid confusion due to many references to X, Y and Z throughout the paper.

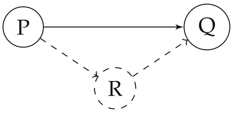

Fig. 1 The direct effect of P on Q, $P \rightarrow Q$, and the indirect effect via R, $P \rightarrow R \rightarrow Q$, cancel out. The dashed arrows indicate that the route through R is unknown.

that fixing R, interventions on P would change Q, and that is all it takes for P to be a contributing cause of Q independently of any particular variable set, according to (CC_{DR}). Moreover, P is an unmediated contributing cause of Q. So, relative to $\{P, Q\}$, P is not a direct cause of Q, but it is an unmediated cause of Q.

But what does ”direct cause” mean, if not an ”unmediated cause”? Yet here these two concepts come apart: P is not a direct cause of Q, but is an unmediated cause. However, for the practical purpose of predicting the outcomes of interventions in a variable set, the distinction between direct cause and unmediated contributing cause has no consequences. Given the variable set $\{P, Q\}$, asking whether either concept applies to the relation between the variables boils down to asking whether Q is manipulable by intervening on P, and the answer is no. In fact, there cannot be circumstances in which applying one or the other concept would lead to different, completely specified claims about manipulability relations. These two concepts simply do not track distinct types of manipulability relations. To be clear, (CC_{DR}) does entail additional claims to (DC) when the analyzed variable set is $\{P, Q\}$, just not ones that explicitly describe *how* things could be manipulated. Namely, affirming that P is a cause of Q in the sense of (CC_{DR}) commits one to the claim that there exists additional variables, minimally one, that are not contained in the analyzed variable set $\{P, Q\}$, such that under some combination of interventions on those variables, Q will be manipulable by intervening on P. But notice how the pragmatic aim of connecting causal knowledge to manipulability fails to be met for (CC_{DR}) unless these variables are included in the analyzed variable set; one could correctly believe that P is cause of Q in the sense of (CC_{DR}) without knowing about R and, hence, without knowing how Q can be manipulated. The only semantic differences between (DC) and the concept of unmediated cause in the sense of (CC_{DR}) consider variables not included in the analyzed variable set. They entail the same manipulability claims in all analyzed variable sets. For instance, for the variable set $\{P, Q\}$, both entail that neither variable can be manipulated by intervening only on the other one. If R is included in the analyzed variable set, both concepts entail the same positive manipulability claims: Q is manipulable by combinations of interventions on P and R. Thus in light of *Manipulability Thesis*, the distinction between (DC) and (CC_{DR}) should not be drawn.

Yet another complication arises due to the way facts about contributing causation depend on facts about direct causal relations. Namely, for a variable X to be a contributing cause of Y, there must, among other things, be a directed path from X to Y, and a directed path is but a chain of direct causal relations. But direct causation is, of course, defined relative to a variable set. How can one say that, for example, P in our example is a contributing cause of Q independently of any particular variable set, when this fact is constituted by facts that hold only relative to a particular variable set? In more detail, the underlying thought with derelativization is supposed to be that P causes Q independently of any particular variable set because there exists the variable set $\{P, Q, R\}$ in which $P \rightarrow R$ and $R \rightarrow Q$ form a path that, when held fixed by intervening on R, allows Q to be manipulated by interventions on P. But these facts about direct causation only hold relative to $\{P, Q, R\}$, not independently of any variable set. It is unclear what to make of this, given that it is perfectly straightforward that P is *not* a direct cause of Q relative to $\{P, Q\}$. It would seem that the existential quantification in the definition of (CC_{DR}) must be interpreted to range over the variable sets implicated in the underlying definition of (DC): (DC) must be derelativized as well.

Given all of the above, it may be natural to ask why is the concept of direct causation relativized to a variable set in the first place? If both direct and contributing causation were defined relative to the *existence* of a variable set, the superfluous distinction would not arise, as P would be both direct and contributing cause of Q regardless of variable choice. This unfortunately does not work, as is acknowledged by Woodward. Since the reasons are not entirely obvious, it is worth investigating in some detail why solutions for eliminating the distinction cannot be based on derelativizing direct causation. This is the topic of Sect. 4 below.

4 Why direct causation must be relativized to a variable set

Consider a causal chain $X \rightarrow U_1 \rightarrow Y \rightarrow U_2 \rightarrow Z$, where the functional forms of the direct causal relations are such that changes in upstream causes propagate to all downstream effects via the intermediary causes. Then consider the following sets of claims about the relations between X, Y, and Z:

1a X is a contributing cause of Z
1b Y is a contributing cause of Z
1c X is a contributing cause of Y

and

2a X is a direct cause of Z
2b Y is a direct cause of Z
2c X is a direct cause of Y

Claims *1a* to *1c* are unproblematic, albeit fairly uninteresting: all are straightforwardly true by the assumption that all effects are dependent on all indirect causes. Given these assumptions, there exists a variable set, e.g. $\{X, Y, Z\}$, such that relative to that variable set, the definition of contributing causation is satisfied as claimed by *1a* to *1c*. This makes *1a* to *1c* true independently of any particular variable set, given the derelativized concept of contributing causation (CC_{DR}).

With respect to the second set of claims, matters are different. None of these claims are straightforwardly false by interventionist standards, in that there is a sense, elaborated below, in which none of them make claims about manipulability relations that contradict the claims entailed by the assumed ground truth. More specifically, the answers depend on the choice of variable set. *2b* and *2c* will be true if one considers e.g. variable set $\{X, Y, Z\}$. Against that variable set, *2b* claims that intervening on Y and holding fixed every other variable than Z in $\{X, Y, Z\}$ will change Z, which is true according to the assumed ground truth. Claim *2c* makes the analogous claim about X and Y. By similar considerations, claim *2a* is true if one considers variable set $\{X, Z\}$, as the only claim made with respect to those variables is that intervening on X and nothing else will change Z, which is a claim entailed by the ground truth also. By contrast, all claims *2a* to *2c* will be false in any variable set that contains the claimed causal relata plus variables U_1 and U_2. One such variable set is, of course, the set that contains all the variables in the ground truth. In any such variable set, *2a* to *2c* will entail claims that some manipulability relations hold between some of $\{X, Y, Z\}$ when U_1 and U_2 are held fixed. All such claims are entailed to be false by the ground truth, where U_1 is between X and Y, and U_2 between Y and Z, so that when U_1 and U_2 are fixed, no intervention on X or Y will change any variable located further down the path. One might now ask, why not derelativize the notion of direct causation in the same way as contributing causation? That way, claims of the form "A is a direct cause of B" would be either true or false independently of any particular variable set, depending on the existence or not of a variable set in which (DC) is satisfied with respect to A and B.

To see why the analogy to (CC_{DR}) does not work, let us assume such a derelativized notion of direct causation: what makes a claim of direct causation true is the existence of a variable set in which (DC) is satisfied for the variables of interest. All claims *2a* to *2c* appear true in this absolute, derelativized sense, since there exists subsets of $\{U_1, U_2, X, Y, Z\}$ in which each of the claimed direct causal relations hold. Now recall that a causal structure comprises variables and direct causal relations between them; a representation of a causal structure is constructed by connecting every pair of directly causally related variables with an arrow, and (DC) describes just what it means for two variables to be directly causally related. For the assumed chain from X to Z, every other qualitative fact about a causal structure is then determined by the distribution of direct causal relations among the variables, given the assumption about the functional forms of the direct causal relations. Given our derelativized reading of direct causation, all claims *2a* to *2c* are judged to be true independently of any particular variable set, and thus instruct that the variables X, Y and Z should be connected as direct cause and effect as shown in the simple, two-variable models below:

model 2a $X \rightarrow Z$
model 2b $Y \rightarrow Z$
model 2c $X \rightarrow Y$

The idea is, then, that the structure of causal relations over $\{X, Y, Z\}$ could be constructed by combining these simple or partial models that describe established facts about direct causal relations. The corresponding model over these variables is shown in Fig. 2.

But this is simply wrong in a way none of the individual claims are. Recall the *Manipulability Thesis*, which says that each completely specified set of claims about manipulability relations corresponds to a causal structure, and vice versa. The model in Fig. 2 entails that intervening on X while holding Y fixed would change Z. But this claim is false according to the ground truth structure $X \rightarrow U_1 \rightarrow Y \rightarrow U_2 \rightarrow Z$;

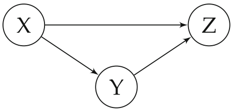

Fig. 2 A complex model built by combining direct causal relations that obtain in some subsets of $\{U_1, U_2, X, Y, Z\}$, where $X \rightarrow U_1 \rightarrow Y \rightarrow U_1 \rightarrow Z$ is the true structure that determines manipulability relations among all the variables.

if Y is held fixed, no intervention on X will change Z. There is no subset of $\{U_1, U_2, X, Y, Z\}$ in which the manipulability claims entailed by Fig. 2 hold. Yet all the direct causal relations that comprise the faulty model in Fig. 2 do hold in some subset of $\{U_1, U_2, X, Y, Z\}$, according to the ground truth. This example establishes that one cannot build a correct description of the causal structure over a variable set $\mathbf{V}$ based on knowledge of direct causal relations established in variable sets other than $\mathbf{V}$: there simply are no facts about direct causation that are independent of $\mathbf{V}$ that would determine the causal structure over $\mathbf{V}$, unless one makes an unrealistic demand that (DC) only applies at the finest possible granularity at which causal relata can be described (see Sect. 5.1 below for why this will not work). As Woodward notes, any notion of direct causation suitable for interpreting and regimenting how directed graphs represent causal structure must be relativized to a variable set (Woodward, 2008, p. 208).

5 Possible solutions

5.1 Define direct causation relative to finest granularity of causal dependence

Section 3 described a problem for interventionism: if contributing causation is derelativized but direct causation is not, it follows that the concept of direct cause is not coextensional with unmediated contributing causation. The obvious solution would be to derelativize direct causation as well (which Woodward rightly does not attempt), but that solution would face the problem described in Sect. 4: given a causal chain of at least three variables where effects are sensitive to interventions on indirect upstream causes, pairs of variables on the chain will be directly causally related in some proper subset of the whole variable set, but judgments about these direct causal relations obviously cannot be combined to a correct representation of the whole chain. Hence, the concept of direct causation can only apply relative to a variable set, if it is to regiment causal reasoning with DAG-like representations. Thus we are seemingly left with the problem described in Sect. 3—if contributing causation is derelativized but direct causation is not, a direct cause is not the same as an unmediated (contributing) cause.

This reasoning overlooks one simple solution. This solution would involve insisting that two variables are directly causally related only if it is impossible to interpolate other causes between them: A is a direct cause of B if and only if there exists a variable set in which A and B are directly causally related as dictated by (DC), and there is no variable set in which there are other causes between A and B. As a consequence, a variable X being a contributing cause of Y would now mean that there is a contributing causal relation between X and Y in a variable set that is most fine-grained or fundamental in the sense that it specifies every mediating variable between X and Y, as well as every variable on every path that needs to be controlled for in order to render Y manipulable by interventions on X. The problem with causal chains would not arise—claims like *2a* in Sect. 4 would be false, because they claim that two variables are directly causally related even though there exists a variable set in which they are not. Consider now the consequences of this for the problematic example from Sect. 3, where the concepts of direct cause and unmediated cause come apart. Given this new definition, P would be a direct cause of Q independently of any particular variable set, since there exists a variable set $\{P, Q, R\}$, in which P is a direct cause of Q, assuming that there are no expanded variable sets with other causes between P and Q. The concept of direct cause would thus be coextensional with the concept of unmediated contributing cause, as intended. If there exists a variable set with variables causally in between P and Q, then P is not a direct cause of Q and the model in Fig. 1 is incorrect, but the fact remains that P is a contributing cause of Q.

The problem with this proposal is that it would entail that a representation of a causal relation between two variables is correct only if it includes every mediating variable between them (cf. Eberhardt, 2014). This is neither a realistic nor a useful requirement. All causal claims made in the special sciences, e.g. social or biological sciences, would be judged to be false out of hand even if they correctly identify manipulability relations, as they always involve some degree of coarse-graining or abstracting away from underlying causal detail. In cases where the process underlying a causal dependence between variables is continuous it would even be unclear what such a requirement really means. Moreover, nothing in the guiding idea of interventionism—that the purpose of causal knowledge is to predict outcomes of interventions—motivates imposing such a requirement of correctness on representations of causal structures.

5.2 Restrict the scope of interventionism

The structure characterised by Eqs. (1) and (2) and the model in Fig. 1 exemplifies a violation of faithfulness (Woodward, 2003, p. 49). Faithfulness is the assumption that all probabilistic independencies present in a set of variables are consequences of the causal Markov condition (Eberhardt, 2009; Spirtes *et al.*, 2000, p. 13). In other words, it is assumed that every causal relation is accompanied by probabilistic dependence between the cause and effect variables given the direct causes of the former, and no two variables are causally related without being so dependent. Assuming a probabilistic version of Eqs. (1) and (2) with independently distributed random errors added for each variable, faithfulness fails for $\{P, Q, R\}$: P is (unconditionally, for it has no causes) independent of Q, $p(Q) = p(Q|P)$, even though P is a cause of Q, and R is independent of Q conditional on its direct cause P, $p(Q|P) = p(Q|P, R)$, even though R is a cause of Q.

Faithfulness is a standard assumption in applications of directed graphs for causal inference or discovery. The usual argument for this assumption is that violations of faithfulness would require that the functional dependencies between the variables satisfy a specific constraint, such as the equality $a = -bc$ that is assumed to hold for the coefficients in Eqs. (1) and (2), which can only be the case in particular, contrived circumstances. Formal results establish that in the space of arbitrary parametrizations of causal DAGs, the set of points that corresponds to faithfulness violations is the null set (Spirtes *et al.*, 2000, pp. 42-43, pp. 382-385; Meek, 2013). Thus, assuming that the causal dependencies found in real-world causal structures take any arbitrary functional form with equal probability, it would be extremely unlikely that we would ever be presented with causal structures that violate faithfulness. In the absence of a specific reason to think otherwise, faithfulness could then be assumed as a sensible default.

In Sect. 3 it was shown that the notions of direct cause and unmediated (contributing) cause come apart when these concepts are applied to systems where the unmediated effect of a variable is exactly cancelled by the effect it has through a route mediated by other causes, amounting to a faithfulness violation. One way to avoid the problem that not all unmediated causal relations in the sense of (CC_{DR}) are direct causal relations in the sense of (DC), or at least to reduce the number of instances that belong in the extension of the former but not the latter[4], would thus be to restrict the scope of interventionism to systems that satisfy faithfulness, or to make faithfulness a part of the interventionist definition of causation itself (see Eberhardt (2014) for a longer discussion of similar issues). One would then argue for one or the other move by either claiming that faithfulness violations are corner cases that require drawing on altogether different causal intuitions than difference-making, or by claiming that faithfulness violations simply do not arise in real systems and therefore are not genuine examples of causal systems that an analysis of causation needs to handle.

Neither of the above options is satisfactory. Causal relations in the real world do not exhibit arbitrary functional forms in equal frequencies, but are parts of causal structures that are often subject to external constraints that favor certain sets of functional dependencies over others. It is well known that faithfulness violations can arise, and indeed must arise, in systems that have a capacity to adapt to internal or external changes in order to maintain a particular internal state; e.g. evolved or engineered control systems (Andersen, 2013). For example, in endothermic organisms such as humans, physical activity both raises body temperature through its direct kinetic effect, and lowers body temperature by triggering physiological mechanisms like sweating. Outside extreme circumstances these effects cancel out over time to maintain normal body temperature. Systems like these are subject to a selection process that makes faithfulness violations much more common than expected by chance (Andersen, 2013). Moreover, interventionism has no problems in handling cases like the above example from Sect. 3, at least as long as one considers the complete set of variables $\{P, Q, R\}$. Ruling such structures to be non-causal or beyond the scope of interventionism feels unnecessary.

[4] Whether there are causal structures that satisfy faithfulness that include unmediated causal relations in the sense of (CC_{DR}) that are not direct cause relations in the sense of (DC) is not investigated here. But note that if there are, this would only make the derelativization of contributing causation even more problematic: (DC) would fail to be coextensional with the concept of unmediated cause in the sense of (CC_{DR}) even if faithfulness is made a part of the definition of causation, or unfaithful structures excluded from the scope of interventionism.

5.3 Treat direct causation as a technical term

While the distinction between direct causes and unmediated contributing causes may feel unintuitive, one could argue that this is no problem, as (DC) is meant to be a technical term specifically crafted to explicate the primitive notion of direct cause that is presupposed in representing causal structures with DAGs. Any such notion in turn must be relativized to a variable set, but this is a pure technicality—the real appeal of (DC) is that it connects the technical apparatus of DAGs to other interventionist causal notions like contributing and total cause, which are more intuitive. Or one could be more radical and argue that it does not matter at all that interventionism's technical definitions of causal concepts have unintuitive consequences, as interventionism explicitly aims to be pragmatic: interventionism aims to explain how various causal concepts are used and should be used when the goal of using them is manipulation and control, and interventionism should be judged based on its success in giving such explanations and norms. By and large interventionism does a good job in explaining what the content of various causal concepts must be like in order to play such a role.

Nonetheless, I don't think these arguments suffice for ignoring the problem. The fact remains that one is left with two concepts of unmediated causal relation that are not coextensional, but this distinction has virtually no pragmatic consequences for causal inquiry. Given a variable set to be analyzed, these concepts do not dictate distinct strategies of inquiry in the same way as, say, one would require different analyses to determine total cause as opposed to contributing cause relations. Maintaining such a distinction goes against the pragmatic motivation of interventionism, and thus cannot be justified by appealing to pragmatism.

6 Variable relativity should be accepted

One can of course make the problem go away by giving up the derelativization of contributing causation[5], and instead defining contributing causation relative to a variable set, as it is literally defined in (M). A direct cause would then be a special case of contributing cause, as intended. For the canceling-paths -example from Sect. 3, this would mean that P is neither direct nor a contributing cause of Q in $\{P, Q\}$. The reason for resisting such a move is perhaps equally obvious. The definition of contributing cause describes a minimal criterion of general causal relevance, and relativizing this to a variable set would mean that there is no concept of causation that is entirely free of a subjective element: in some cases the presence or not of a causal relation between variables would partly depend on how we delineate the causal systems that embed those variables, i.e. on how we choose which variables to consider together as putative causal relata. I argue that this is as expected and as it should be, when causal concepts

[5] In (Woodward, 2008), Woodward explains that the definition of contributing causation in (M) was never meant to be relative to a variable set in the same sense as the definition of direct causation, and references to a variable set in the definition should be interpreted as referring to the existence of a variable set. In this sense there are no two distinct versions of the definition. This however means that the problem would arise for Woodward's original theory as described in *Making Things Happen*, and not just for a modified theory that explicitly derelativizes contributing causation.

are viewed as tools for parsing locally observable dependencies into those that support manipulations in virtue of being invariant under interventions, and those that do not.

Recall that for interventionism, causal concepts apply to a system of variables, and the fact that X causes Y (or not) depends on other causal facts, facts about interventions. The concept of intervention is meant to render nonantropomorphic the idea that X being a cause of Y is equivalent to X and Y being dependent if values of X were set by manipulations that override the normal causes of X, e.g. by randomizing X. Any variable I that is a cause of X can take the role of such a manipulation, as long as I does not cause Y through a route that does not include X and is independent of such off-path causes of Y, and for some values of I, X takes particular values or assumes a particular probability distribution.[6] Graphically, for there to be a causal relation between X and Y, the relations between $\{I, X, Y\}$ must thus be $I \rightarrow X \dashrightarrow Y$, where the dashed arrow connecting X and Y indicates that $\{X, Y\}$ is circumscribed as a (rather minimal) causal system with respect to which I takes the role of a manipulator that determines that X is a cause of Y. For a slightly more complex case that includes another cause of Y, Z, we may have a structure like $I_1 \rightarrow X \dashrightarrow Y \dashleftarrow Z \leftarrow I_2$, where $\{X, Y, Z\}$ is now taken to be the system whose causal structure is determined by the existence of interventions I_1 and I_2. Here $\{X, Y, Z\}$ is the analyzed system, and I_1 and I_2 represent manipulations that originate from outside that system. In general, identifying causation with manipulability implies a distinction between a system and its environment: causal concepts apply to systems that are suitably circumscribed from their (causal) environment, but not completely isolated, so that they can conceivably be manipulated by interventions that originate from outside of the system itself (Hitchcock, 2007, pp. 52-53; Pearl, 2000, pp. 349-350 ; Woodward, 2007, p. 68, pp. 91–92).

Nothing of course precludes one from including into a model variables that represent interventions, as was done in the toy examples above, where I_1 and I_2 are included in a model that describes the causal relations between $\{X, Y, Z\}$. Applying causal concepts to $\{X, Y, Z\}$ considers what would happen if variables like I_1 and I_2 would take certain values; per assumption, I_1 determines that X causes Y and I_2 determines that Z causes Y. But the boundary between the system whose causal structure is determined by interventions, and the environment where the interventions originate, could as well be drawn differently. For example, if we take the last mentioned structure $I_1 \rightarrow X \dashrightarrow Y \dashleftarrow Z \leftarrow I_2$, but focus on just X and Y as our target system, Z now counts as an intervention that determines that Y does not cause X. Moreover, since interventions are causes, then for the interventions that determine the causal relations among variables in variable set $\mathbf{V}$, there must be an expanded variable set $\mathbf{V}^*$ that includes those interventions as variables $I_1, \ldots, I_n$, and there must be possible interventions on $I_1, \ldots, I_n$ in terms of which $I_1, \ldots, I_n$ are causes in the first place. The same, then, must hold with respect to the interventions on $I_1, \ldots, I_n$, and with respect to any interventions on those interventions, and so on (Baumgartner, 2009). But at some point this expansion of the variable set must stop. While no causal reasoner has to worry about encountering such a situation, at some point everything that exists will have been included as a variable in the target system. At this point it becomes unclear

[6] To relate back to the idea of intentional manipulation, think of I as a binary variable that indicates a researcher's decision to randomize or not randomize X.

what determines the causal structure of the system *qua* a structure of manipulability relations, since the distinction between the system and the environment from which it can be manipulated by interventions dissolves.

With this in mind, let us again revisit the motivating example from Sect. 3, described by Eqs. (1) and (2) and the model in figure 1. In this example, when we take $\{P, Q\}$ as the target structure, there is no manipulability relation; the variables are not dependent under any intervention on either one. What then is the motivation for insisting that P nonetheless is a contributing cause of Q, as we would if we use the derelativized concept of contributing cause (CC_{DR})? It cannot be pragmatic, since claiming that P causes Q does not entail any true claims about manipulability by intervening on the analyzed variables, $\{P, Q\}$, and would hence be in conflict with *Manipulability Thesis*. To avoid this problem, one could say that here it is the variable choice itself that is incorrect or inadequate, as the relation between P and Q is only identifiable as a manipulability relation relative to the third variable R. An adequate or correct variable set would be one that includes all variables that, according to the underlying causal facts, are needed to demonstrate the causal relations that obtain between the chosen variables, independently of any particular variable set. Omitting R from the set violates such a requirement. Variable choice would then merely affect which causal facts one can correctly represent in a model, whereas the facts themselves obtain independently of any attempts to represent them (Woodward, 2008, p. 209, footnote 8).

The problem with this strategy is that it presupposes that the world as a whole has a determinate and unambiguous causal structure, that a totality of causal facts exists independently of any causal reasoner's perspective, such that for any particular choice of variables one asks causal questions about, those facts determine whether the variable set is adequate or not. This assumption is in tension with defining causation as manipulability, since it is unclear what it means for the world as a whole to have a determinate causal structure *qua* a structure of manipulability relations, as noted earlier.[7] Without reference to some particular viewpoint or target system short of the whole universe, it is not clear how a manipulability definition of causation can be applied. Therefore it is unclear how one should understand the claim that variable relativity only has to do with what parts of the totality of causal facts one can correctly represent in a model.

But describing what causation amounts to when analyzed from a completely universal perspective is not what interventionism is designed to do anyway. According to Woodward, interventionism is best understood as a functional project that describes norms that causal reasoning ought to conform to in order to guide reasoning about manipulation and control (Woodward, 2014). To be useful for this purpose, interventionism arguably does not need an analysis of causation that is completely general, as causal reasoning typically considers systems defined in some local context. What is needed for interventionism to be applicable in such contexts are (causal) assumptions about what kinds of processes would count as interventions on the variables of interest. Once this is clear, interventionism can be used to assess whether or not one

[7] Authors such as Pearl (2000) argue that the concept of causation ceases to apply when the whole world is considered as the target system. Woodward is more cautious and notes that it is perhaps not impossible but nonetheless unclear how to apply causal concepts when there is nothing outside the system being studied that can serve as a source of interventions (Woodward, 2010, pp. 92–93).

is in a position to draw causal conclusions about the relations between the variables of interest, and if not, what kinds of procedures would produce evidence that licenses such conclusions. On the one hand, in most scientific and other contexts where causal reasoning takes place, it is sufficiently clear what would count as an intervention. To take the most obvious example of randomized experiments, it is sufficiently clear that randomization causes the probability distribution of the randomized variable in a manner that approximates an intervention, and it is possible to use this knowledge to pose questions and reason about causal relations (*qua* manipulability relations) between a limited set of variables, even if it is unclear what the totality of causal facts in the universe supervenes on. On the other hand, in contexts where it is unclear what would count as an intervention on a particular variable, such that it is unclear whether that variable can be a target of causal inquiry in the first place, having a completely general philosophical theory of causation is unlikely to help.

When interventionism is understood as such a functional theory, and it is acknowledged that applying interventionist concepts presupposes that a decision is made regarding how to delineate the target system, variable relativity appears less worrysome. In the example from Sect. 3, variable relativity is required for interventionism to give correct prescriptions about when causal concepts apply to the relation between P and Q and when they do not. No manipulation of P alone would change Q, hence no causal conclusions ought to be possible when the target system comprises only those two variables. That the causal status of P becomes relative to a variable set is thus no more than a reflection of the *Manipulability Thesis* and the pragmatic orientation of interventionism. Also, even if the problems inherent to derelativization could be solved, such that it would be straightforwardly true that P causes Q independently of any variable set, this would have problems of its own. For the interventionist, every true causal claim must be associated with some true claims about the outcomes of interventions. In the case of $\{P, Q, R\}$, a claim that P causes Q is associated with the claim that intervening on both P and R would change Q. Derelativization has the consequence that one could correctly believe the claim "P causes Q" without knowing about R, and thus without believing the associated true claim about interventions. In general, one could in theory have any number of correct causal beliefs, without those beliefs amounting to understanding about manipulation and control.

None of the above implies that one must accept a free-for-all relativism about the truth of causal claims. Variable relativity is perfectly compatible with a view according to which the world has a determinate and objective structure of some kind, perhaps characterizable in terms of some metaphysical category like Humean regularities or constellations of powers, as long as this structure involves or gives rise to dependencies between some magnitudes that we can describe with variables in a model. Interventionist causal concepts then apply when a part of that total structure is conceptually isolated from the rest, such that the distinction between the inside of a system and its outside environment becomes meaningful (Kuorikoski, 2014). In such a picture, the function of causal concepts is to serve as tools that agents situated in the world use to parse the structure of dependencies into ones that support manipulations of such locally defined systems and ones that do not. This amounts to a species of perspectivism about causation, but not relativism: causal relations supervene on a basis comprising the objective structure of dependencies plus a distinction drawn by causal reasoners between a target system and the environment from which it can be intervened on. Variable relativity is simply a reflection of the fact that the manipulability of a particular relation between variables within a target system may depend on how the boundaries of the system are drawn; i.e. on the local perspective of a causal reasoner. But once the boundaries of the target system are settled, facts about manipulability relations between the variables included in the system are perfectly objective, insofar as the underlying fundamental structure of the world is objective. If one opposes variable relativity (and hence, interventionism) on the grounds that such objectivity is still in some sense too weak or concedes too much to relativism, one presumably then opposes the very idea that causal concepts primarily function to guide reasoning about manipulability relations.

7 Conclusions

Interventionism is an explicitly pragmatic theory of causation: it aims to define a number of causal concepts that are useful for the purpose of identifying and modeling dependencies that can be exploited for manipulation and control. That these definitions, especially that of direct causation, reference a variable set, is an upshot of this commitment to the pragmatic goal of the theory. To ward off accusations of relativism, Woodward has issued clarifications that change the literal meaning of the definition of contributing causation from a variable relative definition to one that is meant to apply independently of any particular variable set, thus aiming to capture the idea that causal relations *qua* manipulability relations are objective and independent of any attempts by causal reasoners to represent them.

I have argued that derelativizing contributing causation this way entails a distinction between two concepts of unmediated causal relation that are not coextensional, but do not track distinct types of manipulability relations within any given variable set. These concepts differ only in that the derelativized concept of unmediated causation entails claims about manipulability relations with elliptical reference to variables not included in the analyzed variable set, that is, claims about factors that need to be controlled for by interventions to render a putative effect manipulable by interventions on the putative cause, without explicitly stating what those additional factors are. By contrast, the concept of direct causation does not entail such claims. The distinction between the two concepts therefore goes against the *Manipulability Thesis*, which summarizes the idea that every causal structure corresponds to a set of completely specified claims about manipulability. As a consequence, the distinction amounts to breaking the connection between knowledge of causality and knowledge about manipulability and control, and therefore conflicts with the pragmatic goal of interventionism. There is no obvious way to resolve the conflict in a way that preserves a derelativized notion of contributing causation, without simultaneously creating more problems. The concept of direct causation cannot be derelativized to align with a derelativized concept of contributing causation, unless one also insists that the concept of direct causation only applies at the finest possible grain of description, which is an unrealistic and methodologically useless requirement. Excluding the problematic structures from the scope of interventionism would rule some *prima facie* causal structures as non-causal. Biting the bullet and treating the distinction as a mere technicality is not an option if one wants to preserve the pragmatic orientation of interventionism.

I have taken these problems with the derelativized notion of contributing causation as an indication that variable relativity is a required feature of a theory that has the goal that interventionism has. A definition of causation as manipulability is only applicable when a distinction is drawn between a system of which we ask causal questions about, and an environment from which it can in principle be manipulated. This distinction is drawn by an agent that engages in causal reasoning, is influenced by the interests and background knowledge of the agent, and in the formal machinery of interventionism it amounts to a decision to focus on one variable set rather than another. As demonstrated by the example of cancelling paths from Sect. 3, a manipulability relation may obtain between two variables in one variable set, but not another; how the boundaries of the target system are drawn can affect what can truthfully be concluded about manipulability. Hence, if causal concepts are to guide reasoning about manipulability, they should be allowed to exhibit the same sensitivity to variable choice. Even if the problems with derelativization mentioned in this paper could be avoided, it would remain the case that relativization to a variable set is required for causal knowledge to reliably associate with knowledge of manipulability relations. Namely, without relativization, one could possess causal knowledge without possessing knowledge of the associated manipulability relations, which again contradicts the pragmatic goal of interventionism.

Lastly, accepting variable relativity does not imply some deep, metaphysical relativism about the truthmakers of causal claims. The dependencies that decide whether a pair of variables in some variable set exhibit a manipulability relation can be, and are assumed by interventionism to be perfectly objective. Coupled with variable relativism, this does however imply that causal concepts are perspectival. It is not only the objectively existing dependencies, but also the way the boundaries of the target system are drawn, that decide whether causal concepts are correctly applied in some instance or not. If such perspectivism is too much of a concession to relativism, it still does not follow that the definitions of interventionist causal concepts are inadequate given the goal that interventionism sets for itself—to describe causal concepts that are useful tools for manipulation and control—but that this goal is not worth pursuing as the most important one. Criticisms of interventionism that focus on variable relativity should thus be accompanied by reflection about what alternative goals causal reasoning may serve that are more important, and how a philosophical theory of causation should address those goals.

Acknowledgements I would like to thank Michael Baumgartner for extensive feedback on earlier drafts of the paper, and the Toppforsk program of University of Bergen and Trond Mohn Foundation for providing financial support. I also want to thank two anonymous reviewers who provided numerous comments and helpful, concrete suggestions for how to improve the paper. If there are any mistakes left, I am solely responsible for those.

Funding The research towards this article was funded by the Toppforsk program of University of Bergen and Trond Mohn Foundation, grant ID 811866. Open access funding provided by University of Bergen.

References

Andersen, H. (2013). When to expect violations of causal faithfulness and why it matters. *Philosophy of Science, 80*(5), 672–683.

Baumgartner, M. (2009). Interdefining causation and intervention. *Dialectica, 63*(2), 175–194.

Eberhardt, F. (2009). Introduction to the epistemology of causation. *Philosophy Compass, 4*(6), 913–925.

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*(5), 981–995.

Hitchcock, C. (2007). What Russell got right. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited* (pp. 45–66). Oxford University Press.

Kuorikoski, J. (2014). How to be a Humean interventionist. *Philosophy and Phenomenological Research, 89*(2), 333–351.

Meek, C. (2013). Strong completeness and faithfulness in Bayesian networks. arXiv:1302.4973.

Pearl, J. (2000). *Causality*. Cambridge University Press.

Price, H., & Corry, R. (2007). Causation, physics, and the constitution of reality: Russell's republic revisited. Oxford University Press

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT Press.

Strevens, M. (2007). Review of woodward,"Making Things Happen". *Philosophy and Phenomenological Research, 74*(1), 233–249.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2007). Causation with a human face. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited* (pp. 66–105). Oxford University Press.

Woodward, J. (2008). Response to Strevens. *Philosophy and Phenomenological Research, 77*(1), 193–212.

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy, 25*(3), 287–318.

Woodward, J. (2014). A functional account of causation; or, a defense of the legitimacy of causal thinking by reference to the only standard that matters—usefulness (as opposed to metaphysics or agreement with intuitive judgment). *Philosophy of Science, 81*(5), 691–713.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.