Woodward's Interventionism as an Approach to Causation. An Overview of the Criticism

Horia Lixandru
TU Delft, The Netherlands
Lixandru-1@student.tudelft.nl

April 1, 2021

Abstract

Interventionist approaches to causation have enjoyed popularity within the scientific community, being regarded as *the* standard of causal inference. This methodological success has not translated well into consensus regarding the metaphysical question of causation, namely *What is causality?* Both manipulationists and non-manipulationists have raised criticism of the interventionist view, arguing that it is unacceptably anthropocentric, circular, methodologically fallible, or that another theory is more suitable to answer the metaphysical question, such as the Agency Theory or Causal Pluralism. In this literature review I aim to survey the main points raised against the interventionist theory of causation, as it was formulated by Woodward (2003), with a focus on the methodological and metaphysical criticisms.

Keywords — Woodward, Interventionism, Causation, Price, Agency

1 Introduction

Interventionist approaches to causation, which support the claim that causes are *handles* by which we can obtain their effects, are *intuitive* (Woodward (2016)). Whenever a certain result B is wanted, and thought that there is a causal relation between B and another event A, such that A causes B, A would be done in order to obtain B. It is also known that interventions are a powerful move in causal discovery (i.e., the process of discovering causal relations between events) and causal inference (i.e., the process of inferring knowledge from a causal structure) (Holland (1986), Pearl (2009), Spirtes et al. (2000)). The methodological advantages of an interventionist approach to causality, e.g., precision in isolating the target variable and proximity to the experimenter's intuitions (Woodward (2015)), have been held as a

standard of scientific[1] practice, taking on the mantra *No causation without manipulation* (Holland (1986), Rubin (1986)). Due to its methodological appeal, some contemporary philosophers (e.g., Menzies & Price (1993), Price (1991), Woodward (2003)) have also adopted manipulability as a metaphysical claim about causation (*What is causation?*), not just as a methodological one (*How can we use interventions?*).

However, there are, as with any philosophical theory, criticisms from both sides of the debate. There are some (e.g., Price (1991), Menzies & Price (1993)) that support, in essence, the claim that causes are *handles* that we can wiggle, but that place an emphasis on agency (i.e., the capacity of an actor to *act* in a given environment) rather than on the intervention. On the other side, of course, there are those that would rather see interventionism not be a part of the metaphysical definitions of causation altogether (Reutlinger (2013), Reutlinger (2012)). Their main points are that interventionism presents unacceptable weaknesses, such as anthropocentrism and circularity, or that another theory is better at answering the metaphysical question. Cartwright (2007b) and Baumgartner (2009) go even further and attack interventionism on a methodological level as well, thus challenging the scientific status quo itself.

My aim in this literature review is to present an overview of the main internal[2] and external[3] criticisms to the interventionist theory of causation. Should the reader require a more detailed presentation to Woodward's theory, I have provided a more detailed exposition in Section 3 of this paper. I will present the main raised criticisms to Woodward: The reply that Price offered to Woodward (2003) in his Price (2017) or, in other words, the reply of the agency theorists of causation to the interventionist criticism, focusing on how Woodward's theory itself is inescapably circular and anthropomorphic. Then, I will discuss the criticism offered by those that do not support manipulationism, mainly focusing on the criticism of Reutlinger (2012), as well as the one raised by causal pluralists (i.e., those that hold that causation is not one single concept, but rather a collection of concepts under the same umbrella-term) such as Cartwright (2007a). I will end with a brief discussion of the interventionist approach as a whole and state my own opinion on the debate, which gives preference to causal pluralism as a metaphysical stance, and interventionism as a methodological one.

[1]To clarify, I take *Science* as encompassing disciplines that use the Scientific Method as their main tool. While this might be an absurd oversimplification, it helps in setting at least a blurry line of distinction.

[2]Non-interventionists, but manipulationists.

[3]Non-manipulationists

2 Methods

2.1 Search

The search of the articles has been done in digital libraries. The ones used were JSTOR, Scopus, Elsevier, as well as the Stanford Encyclopedia of Philosophy. The following keywords were used in the search, as well as combinations of them: causation, causality, intervention, interventionism, manipulability, agency, Woodward, Lewis, Menzies, Price, Pearl, as well as their French equivalent. The cutoff date for papers was 1990 (so as to include Menzies & Price (1993)).

Table 1: Type of criticism raised by each paper and their theory of causation

	Methodological	Metaphysical	Underlying Theory [i]
Baumgartner (2009)	+	-	Regularity[ii]
Cartwright (2007b)	+	-	Pluralism
Gijsbers & de Bruin (2014)	-	+	Agency
Kistler (2013)	-	+	Production [iii]
Licata (2019)	-	+	Agency
Price (2017)	-	+	Agency
Reutlinger (2012)	-	+	Counterfactual

[i] Theories marked by a footnote in this column are outside the scope of this paper and are included for accuracy [ii] Baumgartner is an adherent to an old theory of causation which maintains, simply, that all there is to causation is a regular succession of events [iii] Production accounts of causation maintain that a cause *produces* its efect, *properties of objects or systems have the power to do the 'causal work'* (Illari & Russo, 2014, p.152).

2.2 Results

The search resulted in both philosophical and scientific literature on the topic, as well as two PhD theses (with one being selected for further review, Strandin (2020)). The most common results for philosophical literature were by Woodward, Menzies, Price, Reutlinger and for the statistical were by Pearl, Halpern, Glymour and Rubin. There papers were selected based on the abstract of the article, with only articles written in the English or French[4] languages being selected. Results were filtered based on whether or

[4]However, no relevant articles written in French have been found

not the authors' main intent in the article was one or more of the following: to argue for a position regarding the role of intervention in causation, to present a position on the methodology of interventions in causal contexts, or to criticise the interventionist account. An overview of the type of criticism that each paper raises to Woodward is presented in Table 1. The further sections are the presentations of the results.

3 A brief overview of Woodward's interventionist theory

In recent philosophical literature, there are two main formulations of a *manipulability* theory of causation (Woodward (2016)). The first one (Price (1991), Menzies & Price (1993)) gives a distinctive place to the notion of *agency* (which I have discussed in section 4), while the other (Woodward (2003)) places the emphasis on *intervention*. In this section, I will give the reader a brief overview of the interventionist theory of causation.

The definition of a cause in the interventionist theory is the following (Woodward, 2003, p. 55):

> A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set 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 V besides X and Y are held fixed at some value by interventions.

To illustrate, consider the following example (adapted from Woodward (2016)). To concretize, and oversimplify, suppose we have three events, X, Y and Z. X is the event "Leaving out the trash", event Y is "The Kitchen has a foul smell", and an event Z being "Having stinking beetles in the kitchen". Causal structures are most commonly represented as graphs (Pearl (2009)). Then, graphically, the structure would look as the one in Figure 1[5]:

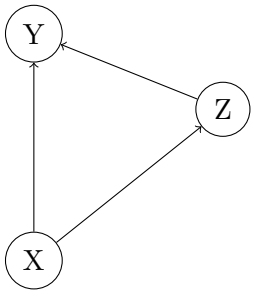

Figure 1: Causal Structure Example

[5]For completion, V would be the set {X, Y, Z}

This graph is saying that

> X is causing Y (Leaving out the trash makes the kitchen have a foul smell)
>
> X is causing Z (Leaving out the trash leads to having stinking beetles in the kitchen)
>
> Z is causing Y (Having stinking beetles in the kitchen makes the kitchen have a foul smell)

Moreover, this graph has an *indirect* path from X to Y that is going through Z. Having and indirect path means that the list of intermediate nodes along the path X-Y is not empty. So, in this example, X is both directly and *indirectly* [6] causing Y.

What would then an intervention look like in this case? Not *any* intervention will do. Woodward defines a list of conditions that must hold for an intervention, I, to be valid. They are ((Woodward, 2003, p. 98), restated in Woodward (2016)):

> 1. I must be the only cause of X; i.e., [...] the intervention must completely disrupt the causal relationship between X and its previous causes so that the value of X is set entirely by I,
>
> 2. I must not directly cause Y via a route that does not go through X as in the placebo example,
>
> 3. I should not itself be caused by any cause that affects Y via a route that does not go through X, and
>
> 4. I leaves the values taken by any causes of Y except those that are on the directed path from I to X to Y (should this exist) unchanged.

So, there are three elements that are part of an intervention. The intervention itself, I, and the two events between which we want to establish the causal relation, X and Y. An intervention I consists in *adding* a new node to the graph, and connecting it to the node of the event X. At the same time, any incoming edges of X besides the one from I need to be eliminated according to (1). For example, in the previous example Z is a direct cause of Y, and an intermediary node along the X-Y indirect cause path. If an intervention were to be done on Z, the causal link from X to Y would need to be eliminated, and be instead included as part of the intervention. In other words, the intervention needs to take into account the effect that X has on Z and counteract it. In this case, the value of X would remain unchanged, as condition (4) requires.

[6] For more on indirect causes, see (Woodward, 2003, p. 59). For the current presentation, working only with direct causes will suffice in giving the reader an idea what interventionism is about

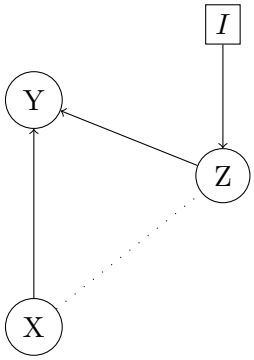

Figure 2: Example of a valid intervention on Z

A valid intervention on Z in the previous example would look like the one in Figure 2 [7]. This intervention, I [8], is valid because it is the only cause of Z (thus respecting (1)) [9], there is no direct link between I and Y (condition (2)), there is no direct cause of I (respecting condition (3)), and all other variables have not been modified (condition (4)).

4 Causation and Agency. Menzies-Price, Woodward and back to Price

The statement of the agency theory of causation is the following:

> an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B(Menzies & Price, 1993, p. 187).

Price (2017) claims that Woodward has misunderstood the aim of the agency theory. Price (2017) mentions that the project of the agency theory is not to reduce causation to agency, as Woodward has thought, but instead it is concerned with

> the anthropological project of explaining [the concept of causation's] genealogy and use (p. 77-78).

In other words, their aim (of Menzies & Price (1993)) is to provide an account of how humans acquire and utilise the concept of causation, as opposed to

[7] I have drawn I in a rectangular shape in order to make it visually distinctive from the regular variables of the causal structure

[8] In natural language, the intervention might sound like: *Putting stinking beetles in the kitchen*

[9] The previous link between X and Z has been dashed to show the change

offering an analysis of causation. Therefore, Price (2017) suggests, the claim of circularity advanced by Woodward (2003) is unfounded (p. 149).

Price (2017) also suggests that there are still traces of the agency theory in Woodward's own interventionist theory of causation (p. 86-87), thus subjecting it to the same counterarguments and criticisms. Woodward (2003) claims that

> [the agency theory of causation] leads us toward an undesirable kind of anthropomorphism or subjectivism regarding causation (p.123).

and that, unlike agency

> The notion of an intervention is an abstract representation of a human experimental manipulation that is stripped of its anthropocentric elements and characterized in terms that make no reference to human beings or their activities (p. 374).

Accordingly, interventions are the objective part of agent manipulations. Their change is *surgical* (Woodward, 2016, paras. 20), meaning that the intervention has to be well defined, and not any action can be considered an intervention [10]. This is how Woodward aims to solve the issue of anthropocentrism of the agency theory. However, Price (2017) points out that

> when Woodward says that 'our notion of causality developed in response to the fact that there are situations in which *we* could manipulate' (2003: 120, emphasis added), the indexical term 'we' is ineliminable. Agents with different epistemic 'situations' to our own will make different judgements about what could be manipulated by manipulating what, and there's no objective sense in which we are right and they are wrong (p. 86-87).

In other words, there is an inescapable degree of subjectivity in choosing *which* interventions to make, and choosing one over another is a matter completely up to an agent's will. For example, in Figure 2, from the previous section, the intervention could have been different were another agent to perform it. It is up to a particular agent how to make the intervention, and another, different, agent could have done an equally valid yet different intervention, with no objective criterion of *betterness* between the two. Thus, Woodward's interventionist theory, according to Price, is still vitiated by circularity.

In recent literature, there has been an attempt to show that agency is, in fact, what *can* solve the problem of circularity, contrary to Woodward's own view. Thus, the view of Price (2017) is shared by some recent works. For example, Gijsbers & de Bruin (2014), are showing how the interventionist

[10] see section 3

theory can be derived from the agency theory, while Licata (2019) is claiming that the presence of anthropomorphic elements is actually an advantage for agency theories, as the whole concept of cause can trace its origin in the human free will, resembling the argument advanced by Menzies & Price (1993) that other sentient beings might have a different concept of causation altogether.

On top of the traditional issue of circularity, there are other types of criticism that do not follow the traditional objections that have been raised to interventionism (i.e., circularity, anthropomorphism, non-manipulable causes and confusing *what is* with *what we know*). It is these kinds of criticism that the next section will present.

5 Non-agency related criticisms of interventionism

The aim of this section is to briefly state the main criticisims raised by the following four papers: Reutlinger (2012), Kistler (2013), Baumgartner (2009), and Cartwright (2007b).

5.1 Metaphysical criticism: Reutlinger and Kistler

The usefulness of a distinction dissipates when it can be simply reduced to the thing it wants to distinguish itself from. This is what Reutlinger (2012) claims is the case with the interventionist theory as well. He is showing how the interventionist theory is simply the counterfactual theory in disguise. Interventionist approaches depart from a counterfactual approach to causation, which claims that, according to Menzies & Beebee (2020):

> the meaning of causal claims can be explained in terms of counterfactual conditionals of the form "If A had not occurred, C would not have occurred".

The gist of Reutlinger's argument is based on the fact that Woodward requires that interventions be merely *logically* possible. As he puts it,

> interventions with this modal character fail to contribute non-trivially to the truth conditions of causal claims, i.e. interventions can be eliminated without loss (Reutlinger, 2012, p.791)

Reutlinger (2012) shows that the interventionist reasoning can be reduced to a counterfactual formulation without loss of generality, as a condition for logical possibility can be simply added to a counterfactual statement which replaces the notion of an intervention. For example, an intervention such as *setting X to value x, then ...* could very well translate to *If X were set to value x, then ..., and it is not logically contradictory to say that* $X = x$.

This is, in a simpler form, Reutlinger (2012)'s argument, with him giving a concrete example, choosing X to be *The Big Bang* (p. 792).

Kistler (2013) advances a different line of criticism, which relates to the sufficiency of Woodward (2003)'s conditions for a direct cause. What Kistler aims is to show that,

> If X(s, t)[11] and Y(s, t) are related by an association law, the interventionist analysis yields the result that X(s, t) is a direct cause of Y(s, t), and that Y(s, t) is a direct cause of X(s, t). (Kistler, 2013, p.68)

In other words, when there is already an underlying law of association between two variables, Woodward's interventionism cannot establish causation between them because of association's symmetry, which is in opposition to causation's asymmetry[12].

5.2 Methodological criticism: Baumgartner and Cartwright

Baumgartner (2009) goes beyond the metaphysical concern and claims that circularity in Woodward's theory is a methodological issue as well, and that the use of it leads to infinite regression and not, in fact, causal discovery as Woodward intended. Baumgartner (2009) points out that due to the fact that Woodard is interdefining causation and intervention. (i.e., X is causing Y if there is a possible intervention I which is *causing* X), there are epistemic problems that arise when *using* interventionism in causal discovery or causal inference. The merits of Baumgartner (2009), as (Strandin, 2020, p.132) has observed, is that it criticizes Woodward on solely methodological grounds, without appealing to any metaphysical concerns.

Another methodological criticism is being put forward by Cartwright (2007b). She takes issue with the assumption of *modularity* of a causal system under intervention. First of all, what is *modularity*? Simply put,

> causal relations are modular if the causal structure of the underlying system isn't altered when one makes interventions on the putative cause. (Illari & Russo, 2014, p.105)

In other words, modularity is the assumption (or property of a causal system) that intervening on a variable will not disturb the values of other variables in the system. It is, in a sense, what makes an intervention *surgical*. To give a general example of modularity, consider clothing. A person can change shoes without changing hats as well. If clothing were not modular,

[11] in Kistler's notation, s is the system to which the variable belongs and t is a time in the system

[12] In other words, causation goes only one way, from cause to effect, and not from effect to cause. Association goes both ways, from A to B and from B to A

the person changing shoes would have resulted in having to change hats as well.

Cartwright's point is that the assumption of modularity concerning causal systems is *epistemically convenient* (Cartwright, 2007b, p.81). In other words, this assumption is, in her view, not reasonable. Woodward (2003) takes modularity as part of the interventionist theory, which is precisely what Cartwright claims is unfounded. Modularity might or might not be part of a certain causal system, but it is not a *sine qua non* part of causation, and assuming it unjustifiably leads to hollow theories.

6 Why is this view popular in the sciences? A Discussion

While Woodward (2003)'s goal is to provide a theory of interventionist causality, that is only one side of the story. His second goal is to provide a theory of causal explanation that is founded on interventionism. Why would that be an important aim at all? The answer: Scientific claims aim to describe causal relations. For example, *The oxidation of iron leads to rust* aims to tell that oxygen, roughly speaking, is *causing* the iron to rust. The advantage of interventionism, as Woodward (2015) claims, is that it helps in clarifying the contents of causal claims. While scientists might claim that the methodology is at most what they want from Woodward's interventionism, the metaphysical underpinning comes along. By trying to take solely the methodology from Woodward's theory, scientists aim to have their cake and eat it as well. The metaphysical assumptions are inherent and precede the methodology, as one needs to know *what* something is before one can make full use of it.

Furthermore, one of the most important debates surrounding interventionism is regarding non-manipulable causes. Baumgartner (2009) observes that Woodward (2003)'s theory entails that what cannot be manipulated cannot be deemed a cause (p. 180). However, this might run against our intuitions. Variables such as race and gender, or the Moon, do seem like causes, and papers such as Glymour & Glymour (2014) and Pearl (2018) aim to show just that, contrary to Woodward's theory. Interventionism, in such situations, seems to fail our intuitions. The question is whether such a failure is metaphysical or methodological. The answer seems to lean towards the metaphysical side, as it fails to encompass valid cases of causation under the concept's definition, yet the reason for that is methodological. Further investigations need to be made in this regard.

7 Conclusion(s)

In this review, I have shown that, while interventionist approaches to causation might be popular in the scientific practice, their philosophical underpinning suffers from pressing criticism, in the sense that there are still unavoidable objections that make its theoretical foundation hollow. I have given the reader my method of literature collection, an overview of the interventionist approach to aid in their understanding, a presentation of the agency theorists' criticism of interventionism, as well as more general criticism in general, of both metaphysical and methodological nature. The main criticisms, in brief, are the following: Interventionism still retains anthropocentric features (Price (2017)), removing agency from interventions leads to circularity (Gijsbers & de Bruin (2014)), interventionism can be reduced to counterfactuals (Reutlinger (2012)), and interventionism fails when there is an underlying association between the variables (Kistler (2013)). Moreover, using interventionism leads to infinite regression (Baumgartner (2009)) and the assumption of modularity is unreasonable (Cartwright (2007b)).

References

Baumgartner, M. (2009). Interdefining causation and intervention. *Dialectica*, *63*(2), 175-194. doi: https://doi.org/10.1111/j.1746-8361.2009.01191.x

Cartwright, N. (2007a). Causation: one word, many things. In *Hunting causes and using them: Approaches in philosophy and economics* (p. 11–23). Cambridge University Press.

Cartwright, N. (2007b). Modularity: it can – and generally does – fail. In *Hunting causes and using them: Approaches in philosophy and economics* (p. 80–96). Cambridge University Press.

Gijsbers, V., & de Bruin, L. (2014). How agency can solve interventionism's problem of circularity. *Synthese*, *191*(8), 1775–1791. Retrieved from `http://www.jstor.org/stable/24020026`

Glymour, C., & Glymour, M. R. (2014). Commentary: Race and sex are causes. *Epidemiology*, *25*(4), 488–490. Retrieved from `http://www.jstor.org/stable/24759150`

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. Retrieved from `http://www.jstor.org/stable/2289064`

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. Oxford University Press.

Kistler, M. (2013). The interventionist account of causation and non-causal association laws. *Erkenntnis (1975-)*, *78*, 65–84. Retrieved from http://www.jstor.org/stable/24010951

Licata, G. (2019). Free action and interventionist theory of causality. *Rivista Internazionale di Filosofia e Psicologia*, *10*, 282–294.

Menzies, P., & Beebee, H. (2020). Counterfactual Theories of Causation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2020 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/.

Menzies, P., & Price, H. (1993). Causation as a secondary quality. *The British Journal for the Philosophy of Science*, *44*(2), 187–203. Retrieved from http://www.jstor.org/stable/687643

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press. doi: 10.1017/CBO9780511803161

Pearl, J. (2018). Does obesity shorten life? or is it the soda? on non-manipulable causes. *Journal of Causal Inference*, *6*(2). doi: doi:10.1515/jci-2018-2001

Price, H. (1991). Agency and probabilistic causality. *The British Journal for the Philosophy of Science*, *42*(2), 157–176. Retrieved from http://www.jstor.org/stable/688050

Price, H. (2017). Causation, Intervention, and Agency: Woodward on Menzies and Price. In H. Beebee, C. Hitchcock, & H. Price (Eds.), *Making a Difference: Essays on the Philosophy of Causation.* Oxford: Oxford University Press.

Reutlinger, A. (2012). Getting rid of interventions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(4), 787-795. (Causality in the Biomedical and Social Sciences) doi: https://doi.org/10.1016/j.shpsc.2012.05.006

Reutlinger, A. (2013). *A Theory of Causation in the Social and Biological Sciences*. Palgrave Macmillan UK.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*(396), 961–962. Retrieved from http://www.jstor.org/stable/2289065

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT press.

Strandin, H. (2020). *Taking Control. The role of manipulation in theories of causation* (Unpublished doctoral dissertation). Department of Philosophy, Stockholm University, Stockholm.

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

Woodward, J. (2015, Nov 01). Methodology, ontology, and interventionism. *Synthese*, *192*(11), 3577-3599. doi: 10.1007/s11229-014-0479-1

Woodward, J. (2016). Causation and Manipulability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/win2016/entries/causation-mani/`.