THESIS TITLE:	Making Things Happen – A Critical Philosophical Reflection
CHAIR GROUP:	Theoretical Philosophy
AUTHOR:	Anne Albert van der Galiën
GRADUATION VARIANT:	Philosophy of a Specific Programme (Economics)
NUMBER OF CREDITS:	10 ECTS
DATE OF COMPLETION:	27/02/2018
SUPERVISOR:	Prof. Dr. J.J. Vromen
ADVISOR:	Prof. Dr. L. van Bunge
WORD COUNT:	15913

Table of Contents

MAKING THINGS HAPPEN

A Critical Philosophical Reflection

I. INTRODUCTION

The notion of causation has been discussed extensively in philosophy, from the time of the ancient Greeks until now (Ruben, 1990). Causation also plays a central role in our daily lives. Properly studying causes good exam grades, drinking coffee causes one to wake up, oversleeping causes one to be late, to mention a few examples. Philosophers have developed various kinds of theories aiming to clarify what we mean when we talk about causation, ranging from causal-process theories to probabilistic theories and from regularity theories to manipulationist theories.

However, each of those approaches has its difficulties and counterintuitive judgments in certain cases. For instance, an example due to Hall (2004) concerns Billy and Suzy, both throwing a rock at a bottle, with Suzy's rock reaching the bottle first and hence shattering it. Theories of causation that are based on counterfactuals, such as counterfactual and manipulationist theories, often cannot explain why Suzy's throw, not Billy's, caused the bottle to shatter—there is no counterfactual dependence between the shattering of the bottle and Suzy's throw, for without Suzy's throw the bottle would still have shattered, be it due to Billy's throw. Due to the absence of counterfactual dependence, theories of causation based on counterfactuals often cannot conform to the intuitive judgment that Suzy's throw caused the bottle to shatter. Causal-process theories, on the other hand, can usually conform to the intuitive judgment, given that there is a clear process between Suzy's throw and the shattering of the bottle. At the same time, there are examples that theories based on counterfactuals can handle satisfactorily where causal-process theories fail. For instance, a counterfactual theory can judge that a doctor failing to administer appropriate medicine for some patient caused his death, given the counterfactual dependence between administering medicine and the patients' survival. A causal-process theory would not be able to deliver this judgment, for there is no process between the event 'not administering medicine' and the death of the patient.[1]

Given those and other problems, the literature on causation is still developing new ways along which to understand the concept of causation. One recent influential treatise on causation is *Making Things Happen* (also referred to as *MTH*), developed by James Woodward (2005). The manipulability theory presented in *Making Things Happen* aims to deliver a conceptual analysis of our usage of the concept of causation, both in scientific and non-scientific contexts. Furthermore, Woodward is concerned with what we *should* mean by our causal claims.

This thesis is a critical philosophical reflection on Woodward's manipulationist theory of causation presented in *Making Things Happen*. Besides offering a theory of causation, Woodward also offers a theory of causal explanation in this book. This thesis restricts itself to the parts of the book relevant for the theory of causation. To this end, the focus will be on the first three chapters of *Making Things Happen*, combined with some sections from the sixth chapter. The first chapter serves as an intuitive introduction to Woodward's theory, after which Woodward introduces his formal machinery and formally develops his theory in chapter two. The third chapter elaborates on the core notion of an intervention, and the sixth chapter deals with the core notion of invariance. The left-out chapters deal with causal explanation (chapter four and five), causal interpretation in structural models (chapter seven), and the eighth chapter compares the theory developed by Woodward with Salmon's causal-mechanical model and Kitcher's unificationist model. A number of sections of the included chapters is left out as well, either because the section

[1] It should be noted that not everyone agrees that so-called cases of omission-causation, i.e. where the absence of some event causes another event, are in fact cases of causation. Hence, some may want to argue that the causal-process theory also delivers the correct verdict in this case.

is irrelevant for the assessment that this thesis aims to perform or because it is not necessary for a proper understanding of the relevant parts of Woodward's theory.[2]

After summarizing the relevant parts of Woodward's theory, this thesis first presents what it judges to be strong aspects of *Making Things Happen* and subsequently develops some criticisms on some of Woodward's arguments.

The structure of this thesis will largely follow the structure of Woodward's book. Sections II, III, IV, and V summarize Woodward's first, second, third, and sixth chapter, respectively. Those sections thus represent Woodward's theory of causation and his arguments; the contribution of the present author in those sections is limited to a concise analysis of Woodward's argumentative strategy that is used to guide the reader through the summary. The main contribution of the present author can be found in section VI, where I discuss strong aspects of Woodward's theory and develop some criticisms. Section VII concludes.

II. MTH CHAPTER 1: INTRODUCTION AND PREVIEW

The sections devoted to a summary of Woodward's *Making Things Happen*—sections II, III, IV, and V—all start with a description of Woodward's argumentative strategy in the relevant chapter. The general strategy employed by Woodward, both across and within chapters, consists of first introducing the topic at hand in a simple and intuitive way, after which concepts and arguments are refined based on difficulties that the simple and intuitive concepts and arguments encounter. Thus, chapter one serves (inter alia) as an intuitive introduction to various concepts that Woodward uses or develops throughout *Making Things Happen*, which are then further developed and refined in subsequent chapters. The main advantage of this strategy is that having an intuitive understanding of the topics discussed helps the reader understand the formal machinery developed later on.

The strategy of Woodward's first chapter can be described as follows. It starts by describing the aims and focus of Woodward's proposed theory, after which Woodward presents an intuitive introduction to his theory by introducing intuitive definitions of the core concepts of manipulability, interventions, and invariance. Subsequently, he motivates his manipulability approach by arguing that our interest in causation must be due to the practical payoffs that it yields—it gives us the possibility to manipulate some factor by manipulating its cause. Also, *desiderata* of causal theories are listed, things that any theory of causation should be able to do.

MTH §1.2: What Should an Account of Causal Explanation Aim to Do?

In the first chapter of 'Making Things Happen', Woodward sketches his approach to the issue of causation and causal explanation, and outlines some of the major themes that are explored later in the book. Woodward focusses on the concept of causation and the practices of causal inference and causal explanation in both scientific and non-scientific contexts. An important part of Woodward's work is devoted to conceptual analysis, i.e. describing and clarifying both ordinary and scientific usages of the concept of causation and related notions. However, Woodward wants to go well beyond merely offering a conceptual analysis of the concept of causation. For one, he not only focusses on verbal usage of the concept of causation, but also on non-verbal usages. For another, Woodward aims to distinguish between different sorts of causal and explanatory claims. Lastly, Woodward's project is also concerned with what the point is of our causal claims, what purpose(s) we seek to achieve by using the concept of causation or by formulating causal claims. Related to this last point, Woodward does not just want to give a description of our usage of the

[2] More precisely, §1.5, §1.8, and §6.3 are left out for being irrelevant for understanding Woodward's theory sufficiently properly. Connections with probabilistic ideas about causation and a short discussion on pluralistic conceptions of causation, §2.4 and §2.9 respectively, are not necessary for the assessment performed in this thesis. The same holds for comparisons with other theories of causation in §3.4 and §3.6. The left-out sections §6.5-§6.15 elaborately discuss the notion of invariance versus the notion of lawfulness; this is also not relevant for the assessment in this thesis.

concept of causation, but also wants to provide a normative component; he wants to make suggestions or recommendations about what we *ought* to mean by our causal claims. Given that Woodward thinks that the positive and normative sides of his project are very intertwined, he is of the opinion that both aspects should be investigated together.

The normative aspect of Woodward's project stems from several sources. A first source is that we are not only interested in the meaning of causation or what causation is, we also want to *do* things based on our causal claims. For example, we may want to give a patient some kind of medication because we think this medication will cause the patient to get better. A good account of causation should fit with those practical purposes. Another source for this normative aspect concerns epistemic limitations and limitations regarding clarity and connections with other concepts or purposes. Concepts of causation can be designed properly or improperly for those purposes. One motivation for Woodward's project is thus to provide a properly designed concept of causation that can be used for the purposes mentioned above.

MTH §1.3: The Manipulability Conception of Causal Explanation

Woodward aims to provide a manipulationist account of causation and causal explanation. To causally explain, Woodward contends, is to provide information that is relevant to manipulation or control. Merely providing a description is not a causal explanation. This illustrates why we are often so interested in causation and causal explanation: it provides information that can be used practically, to manipulate or control something. It should be noted that the notion of a manipulation is not to be understood as an actual notion, but as a modal or counterfactual notion. Hence, in order for Woodward's manipulationist account of causation to underwrite the claim that c causes e, it is not necessary that c can be manipulated in practice. This is an important departure from previous manipulationist accounts of causation, which often employed an anthropocentric concept of manipulability. This anthropocentrism is often seen as one of the main weaknesses of manipulability accounts of causation, a weakness that Woodward thus avoids.

MTH §1.4: Causal Explanation, Invariance, and Intervention Illustrated

Woodward goes on to illustrate some important concepts that he uses throughout the book by discussing two examples. I will limit myself to one of the examples, which involves the link between atmospheric pressure, the occurrence of storms, and barometer readings. The relation between any two of those three elements is one of counterfactual dependence: the occurrence of storms depends counterfactually on the atmospheric pressure and vice versa, the occurrence of storms also depends counterfactually on the barometer reading and vice versa, and the barometer reading depends counterfactually on the atmospheric pressure and vice versa. However, not all those relations of counterfactual dependence are causal relations. The barometer reading does not cause storms to occur; the only causal relations are from atmospheric pressure to the barometer reading and from atmospheric pressure to the occurrence of storms. In order to distinguish between relations of mere counterfactual dependence and genuine causal relationships, Woodward introduces the concept of an *intervention* and the concept of *invariance*. The definition of those concepts here is still intuitive, the formal definition will be provided in section IV and V for interventions and invariance, respectively. An intervention is always defined in terms of an intervention on one factor with respect to some other factor. An intervention on X with respect to Y is then defined as a manipulation of X that results in changes in Y, with the condition that the changes in Y may only be due to this manipulation of X. Having established an intuitive definition of an intervention, Woodward can now proceed to (intuitively) define the concept of invariance: some generalization G is invariant if it continues to hold after some intervention on X. Generalizations can be more or less invariant, and at least some degree of invariance is required for, and sufficient for, a generalization to count as a causal generalization.

Both concepts can be used to illustrate why certain counterfactual dependences in the barometer example are genuine causal relationships while others are not. Using the concept of an intervention, it is clear that there cannot be an intervention on the barometer reading with respect to either the occurrence of storms or the atmospheric pressure: merely changing what the barometer tells us does not change whether or not a storm occurs, neither does it influence the atmospheric pressure. In contrast, intervening on the atmospheric pressure would change the occurrence of storms and the barometer reading. Because such interventions on atmospheric pressure with respect to the occurrence of storms and the barometer reading are possible (again, not in an anthropocentric sense), one can draw the conclusion that atmospheric pressure causes the occurrence of storms and the barometer reading. Similarly, because interventions on the barometer reading with respect to the occurrence of storms or the atmospheric pressure are not possible, the barometer reading does not cause the occurrence of storms or the atmospheric pressure. Using the concept of invariance, one can show that the generalization 'If the barometer reading would fall, a storm would occur' is not a causal generalization: if one would intervene on the barometer reading, this generalization would cease to hold because the occurrence of storms will not be affected by an intervention on only the barometer reading. Hence, this generalization is not invariant and thus cannot describe a causal relationship.

One more remark on invariance has to be made. With the concept of invariance, Woodward intends to replace the notion of *laws of nature* that philosophers often refer to when discussing issues pertaining to causality and (causal) explanation. For example, in the Deductive-Nomological model explanation is seen as a combination of initial conditions, a law or laws of nature, and the phenomenon to be explained. Woodward's position is that most causal relations are clearly not exceptionless laws of nature (for instance, it would be hard to argue that the causal relation between atmospheric pressure and barometer readings is a law of nature), which is why he proposes the more flexible concept of invariance[3].

MTH §1.6: Causal Explanation as a Practical Activity

As noted before, Woodward identifies practical payoffs as the reason, or at least as one of the most important reasons, why we are interested in causality. The widespread interest in causal relations, also by cultures in the distant past, suggests that knowledge of causal relationships at least sometimes have a practical payoff. Theories of causation, therefore, should be able to identify this practical payoff. Woodward's theory of causation identifies this practical payoff of causal knowledge as providing opportunities for manipulation and control. Furthermore, Woodward takes this widespread interest in causality to suggest that we should expect continuity between everyday use of the notion of causality and how causality is used in the sciences.

MTH §1.7: Accounts of Causation and Explanation Can Be Illuminating without Being Reductionist

Woodward then continues by discussing the non-reductive character of his approach. Many philosophers claim that theories of causation should be reductive, which roughly means that causation should be analysed in non-causal terms[4]. Hume's regularity theory of causation is such a reductive theory, since it analyses causation in terms of regularities. The reason behind the idea that theories of causation should be reductive is that it is often thought that non-reductive theories of causation will be circular and hence unilluminating or trivial. Woodward disagrees with this, and claims that non-reductive theories are not necessarily (viciously) circular, and hence can be illuminating and non-trivial. The manipulationist approach proposed in his book is not (viciously) circular, Woodward contends, because even though assessing whether there is a causal relationship between two variables can in this approach not be done without reference to causal relations, those causal relations that are referred to are always other causal relations than

[3] We will return to the concept of invariance in section V. Given that Woodward mainly uses the concept of invariance for his discussion of causal explanation, and given that this thesis focusses on causation as such, invariance will be discussed relatively concisely. For a full discussion, especially on the advantages offered by the concept of invariance compared to the concept of laws of nature, the reader is advised to turn to Woodward's book where he dedicates a full chapter on the concept of invariance.

[4] More precisely, Woodward uses the term 'circle of concepts' to denote closely interrelated concepts. A reductive theory aims to explain concepts in this circle by referring to concepts outside this circle, whereas a non-reductive theory allows itself to also use the concepts in this circle.

the putative causal relationship whose existence is being verified. Hence, to know whether X causes Y does not require knowledge of a causal relationship between X and Y, and circularity is avoided. Despite this, some may claim that non-reductive accounts, whether circular or not, are likely to be trivial. Against this, Woodward makes two points. First, it is in itself interesting to compare the judgments from different accounts of causation, including non-reductive and reductive theories. Second, also within a non-reductive account of causation, non-trivial choices have to be made concerning, for example, how to connect different elements or variables in the causal system, or which counterfactuals should and should not be used in counterfactual theories of causation. Those non-trivial choices have to be made by both non-reductive and reductive theories of causation, and can be made on grounds independent from whether the theory is reductive or not.

MTH §1.9: Desiderata

Woodward concludes his introductory chapter by listing some *desiderata* of theories of causation, based on the discussion so far[5]. The first *desideratum* is that the account should be descriptively accurate: it should capture relevant features of the ways ordinary people as well as scientists talk about causation. Second, a theory of causation should solve problems of older theories of causation. Third, a theory of causation should have plausible epistemological underpinnings; it should be able to make sense of practices of causal inference and of practices of testing causal relationships. This is not to say that a conceptual theory of causation needs to have a component describing how causal inference or the testing causal relationships should be done, for that part of the work should be done by theories of causal inference. Nevertheless, a conceptual theory of causation should fit together with theories of causal inference; together they should be able to form a plausible story of causation and causal inference.

III. MTH Chapter 2: Causation and Manipulation[6]

Having introduced the core concepts of his manipulability approach in an intuitive way in chapter one, Woodward now turns to developing his theory in a more formal and refined way. The argumentative strategy of this second chapter is as follows. First, Woodward introduces the formal framework that he uses to develop his theory (i.e. directed graphs and structural equations). Having done this, he turns to defining several notions of causation, starting off with a necessary and a sufficient condition for causation in a manipulability framework. Woodward's next step is to recognize that the formulated necessary condition runs into trouble when the causal effect of one variable is offset by the causal effect of some other variable. Based on this, he can argue that we should distinguish between total causes and contributing causes. Defining contributing causes, in turn, requires the notion of a direct cause. Having defined what a direct cause is, Woodward can subsequently define a necessary condition for some variable to be a contributing cause. Putting this necessary condition together with the definition of a direct cause yields the definition of his manipulability theory. Woodward then elaborates on some aspects of his manipulability theory, most notably on the distinction between patterns of counterfactual dependence and judgments of causation, where he argues that the latter combines the objective patterns of counterfactual dependence with the subjective notion of serious possibilities.

MTH Chapter 2—Introduction

Woodward starts the second chapter by observing a difference between philosophers and scientists with regards to a manipulability conception of causation. Whereas the philosophical literature has largely been unappreciative of manipulability conceptions, looking at statistics or economics textbooks, for instance, shows that a manipulability notion of causation is common in

[5] The second *desideratum* that Woodward mentions concerns the evaluation of causal explanations. Given the focus of this thesis on causation itself, this *desideratum* is left out.

[6] Chapter section §2.6 is incorporated in §2.2 in this thesis.

practice. Dismissing scientists' conceptions of causality by claiming that they are not aware of the complexities surrounding the concept of causation would be uncharitable, given that causation has been and still is a central and much-discussed concept in various sciences.

MTH §2.1—Motivation

Before laying out his theory, Woodward pays some attention to the question of what the point of having a notion of causation actually is. A standard philosopher's answer would roughly say that the point simply is disinterested intellectual curiosity. This answer, however, is unsatisfying for two reasons. First, it does not explain why we are more interested in causation instead of, for example, mere correlation. Second, it does not explain the fact that animals and children, presumably being less or not intellectually driven, recognize causal relations. In contrast, Woodward's answer to this question is that having a notion of causation has some practical payoff. This explains why causal knowledge is esteemed more highly than knowledge about correlations, for correlations yield less practical payoff. It also explains why animals and children recognize causal relationships, despite their presumed lack of disinterested intellectual curiosity. If one accepts this, the question that remains to be answered is what this practical payoff may consist of. The answer suggested by Woodward's manipulability theory is that having a notion of causality yields opportunities for manipulation and control. For example, eating certain kinds of food enables us to manipulate or control our health (at least to some extent), the productivity of a field can be manipulated by throwing manure on it, and an animal can control its feeling of hunger by eating.

Besides the advantage of explaining what the point of having a notion of causation is, Woodward advances some additional motivations for supporting a manipulability concept of causation. One of those is that it explains why experiments are widely used and believed to be critical in discovering causal relations: properly manipulating one factor in one group but abstaining from doing so in a second group can establish the existence of a causal relationship between the manipulated factor and its putative effect by ascertaining in which group(s) the effect took place, if at all. Another motivation is that it helps to answer the worries of a considerable number of philosophers that the notion of causation necessarily involves excessive or problematic metaphysics in one way or another, and that having causal knowledge does in fact not add more to our knowledge than knowledge of correlations. It does so by demonstrating that interest in causal relations has legitimate and important purposes, namely the practical utility that, Woodward contends, is indispensable for daily life, for survival, and other areas of life.

MTH §2.2: Graphs and Equations as Devices for Representing Causal Relationships

Having explained his motivation for developing a manipulability account of causation, Woodward turns to introducing the basic framework that he uses to develop his theory. In this framework, causal relations are represented as relations between variables. Furthermore, the principal causal relations that Woodward discusses are type-level causal relations, i.e. causal relations in the abstract, that do not refer to particular situations in which the causal relation manifested itself (as opposed to token-level or actual causal relations). To illustrate, the causal relation 'The aspirin that John took caused his headache relief' is a token-level causal relation, whereas 'Aspirins causes headache relieve' is a type-level causal relation. How the type-level causal framework that Woodward develops could be extended to include actual causal relations is discussed in *MTH* §2.7.

Related to this, the type-level causal relations in this framework should be reproducible, meaning roughly that a posited causal relation should continue to hold when the same changes in the value of the putative cause are repeated. Putting it differently, the response of Y to an intervention on X should be general or stable enough to use manipulating X as a strategy for manipulating Y. This reproducibility requirement is intended as a relatively undemanding requirement; it is supposed to rule out causal relationships that are very unstable or hold only once (e.g., buying a lottery ticket causes me to win the jackpot). For deterministic cases, where Woodward focusses on, this reproducibility requirement is stricter than for indeterministic cases.

In the framework that Woodward uses, there are two ways of representing causal relationships: systems of structural equations and directed graphs. Starting with the latter, a directed graph consists of an ordered pair $\langle \boldsymbol{V}, \boldsymbol{E} \rangle$, where $\boldsymbol{V}$ and $\boldsymbol{E}$ are sets representing the vertices (or variables) and the directed edges of the graph, respectively. A directed graph may look as follows:

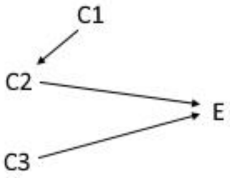


Figure III-1

Some terminology can now be introduced (some of which will be made more precise in subsequent sections). A *direct cause* is a cause that is connected to its effect via one directed edge. For example, in figure III-1, $C1$ is a direct cause of $C2$, but not of E. A *directed path*, or *route*, is a sequence of variables in which every variable is the direct cause of the subsequent variable. For instance, the sequence $\{C1, C2, E\}$ is a directed path from $C1$ to E in figure III-1. A variable Y is a *descendant* of variable X if and only if there is a directed path from X to Y. If Y is a descendant of X, then X is an *ancestor* of Y. Direct causes are also called *parents*. Applying this to figure III-1, we can for example state that E is a descendant of $C1$, $C2$, and $C3$, implying that $C1$, $C2$, and $C3$ are ancestors of E. The parents of E are $C2$ and $C3$, but not $C1$.

Causal relationships can also be represented by means of systems of structural equations. Representing the situation depicted in figure III-1 using structural equations may result in the following system of equations:

$$C1 = 2$$
$$C2 = 2(C1)$$
$$C3 = 0$$
$$E = C2 + 0.5(C3)$$

The variables $C1$ and $C3$ are exogenous variables and hence do not refer to other variables. Structural equations of endogenous variables such as $C2$ and E encode counterfactual information; they tell us what would happen if we would change the value of some variable figuring in the structural equation.

Note that representing causal relationships by structural equations conveys more information than representations by directed graphs. Whereas directed graphs only tell us that there is some relationship between two variables (or, more precisely, that the variable where the directed edge starts figures in the structural equation of the variable that the directed edge points towards), structural equations also show us the specific relationship between two variables. For instance, above we represented the variable E in figure III-1 as being determined by the equation $E = C2 + 0.5(C3)$, but we could also have chosen the equation $E = C2 * C3$ since the graphical representation of both structural equations would be the same.

MTH §2.3: Direct, Total, and Contributing Causes

Based on this framework, Woodward can now develop his manipulability account of causation more precisely. He starts by asking whether manipulability is a necessary or a sufficient condition for causation. He provides a definition of both:

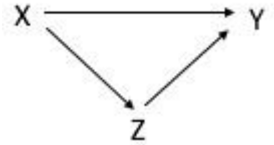
> <u>Sufficient Condition ***SC***</u>
> If (i) there is a possible intervention that changes the value of X such that (ii) carrying out this intervention (and no other intervention) will change the value of Y, or the probability distribution of Y, then X causes Y. *(MTH, p. 45)*

Necessary Condition $\boldsymbol{NC}$
If X causes Y then (i) there is a possible intervention that changes the value of X such that (ii) if this intervention (and no other interventions) were carried out, the value of Y (or the probability of some value of Y) would change. *(MTH, p. 45)*

Some clarifications are in order. Why do both definitions incorporate the restriction of one intervention and no other interventions? The reason is straightforward: carrying out multiple interventions may result in not being able to distinguish between the various causal relationships that play a role. For example, suppose that X does *not* cause Y, but that every intervention on X also brings about an intervention on some other variable Z that does cause Y. Then Y will change systematically with interventions on X (via the resulting interventions on Z) even though there is no causal relationship between X and Y.

Another important clarification concerns the meaning of 'possible intervention'. Woodward emphasizes that this is no anthropocentric notion of 'possible', in contrast to many previous manipulability theories of causation. What 'possible' does mean is made more precise in chapter 3, and is discussed here in section IV. The notion of 'interventions' is also discussed more precisely in chapter 3 (discussed in section IV); intuitively something counts as an intervention if it meets ideal experimental conditions, i.e. conditions that ensure that if some effect takes place, it only takes place due to the putative cause.

How plausible are the conditions defined above? Woodward claims that $\boldsymbol{SC}$ is very plausible: if it is possible to manipulate Y by (only) manipulating X, then there must be some causal relation between the two variables. In contrast, Woodward shows that $\boldsymbol{NC}$ does not necessarily hold. Suppose we have the following causal structure:



$$X = 1$$
$$Y = X + Z$$
$$Z = -X$$

Figure III-2

In this causal structure, the effect of X on Y is cancelled out by the effect that Z has on Y. Hence, $\boldsymbol{NC}$ is not met in this example since Y does not change in response to an intervention on X. Nevertheless, there is a causal relation between X and Y. This example thus shows that we must distinguish between what Woodward labels *total cause* and *contributing cause*. To illustrate, applying this distinction to figure III-2 would result in classifying X as a contributing cause of Y (because there is some causal connection between X and Y), but not a total cause because intervening on X does not result in changes in the value of Y.

The concept of total cause can be characterised more precisely:

Total Cause $\boldsymbol{TC}$
X is a total cause of Y if and only if there is a possible intervention on X that will change Y or the probability distribution of Y. *(MTH, p. 51)*

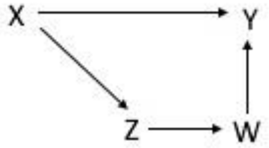
In order to characterise the concept of contributing cause more precisely, we first need to formalize the notion of a *direct cause*[7]:

Direct Cause $\boldsymbol{DC}$
A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set $\boldsymbol{V}$ is that there be a possible intervention on X that will change Y (or the

[7] Besides for the notion of a contributing cause, the notion of a direct cause is also necessary for figuring out causal relations under multiple interventions, formulating connections between probability and causality, defining the notion of an intervention, and to capture the notion of distinct causal mechanisms.

> probability distribution of Y) when all other variables in $\boldsymbol{V}$ besides X and Y are held fixed at some value by interventions. *(MTH, p. 55)*

By means of illustration, consider the following causal structure:



$X = 1$
$Z = 0.5X$
W = 2Z
Y = X + W

Figure III-3

Assessing whether X is a direct cause of Y should thus be done by fixing the values of all other variables in $\boldsymbol{V}$ (hence we fix Z and W at some value, say 0.5 and 1, respectively), intervening on X (for instance by setting its value to 2), and subsequently verifying whether the value of Y changed or not. It can readily be verified that fixing Z and W at the said values and intervening on X to set its value to 2 would change the value of Y from 2 to 3, hence X is a direct cause of Y.

An important and somewhat problematic characteristic of $\boldsymbol{DC}$ is that whether a variable is a direct cause of another variable is relative to a variable set $\boldsymbol{V}$. In other words, whether a variable is a direct cause is representation dependent. For example, suppose that we would want to represent the situation of figure III-3 in a slightly different way by excluding the variable W. The structural equation for Y would thus become $Y = X + 2Z$. In such a situation, variable Z would be a direct cause of Y, in contrast to the original situation of figure III-3 where Z is only a contributing cause. This is not necessarily problematic or undesirable, for a variable that is a direct cause of some other variable in one variable set will at least remain a contributing cause (unless the variable itself is excluded, of course). However, there are some more problematic examples, notably examples where cancellation plays a role. Consider, for instance, the causal structure sketched in figure III-2, where the effect of X on Y is cancelled out by the effect that X *via* Z has on Y. Removing Z from this causal structure would result in X ceasing to be a cause—direct or contributing—of Y (for the equation for Y would become $Y = X - X = 0$). According to Woodward, this problem illustrates a more general problem for theories of causation that make use of directed graphs and structural equations: any such theory of causation needs to have some account telling us which representations are suitable or possible, and what (kind of) representations are not suitable or possible. Woodward discusses this more extensively in *MTH §2.8*, a section on so-called serious possibilities which I discuss later in this thesis. Also, Woodward claims that this problem of representation-dependency suggests that “there is a sense in which facts about patterns of counterfactual dependence are more basic than facts about what causes what in *either* the total cause or direct cause sense” (*MTH*, p. 57).

Suspending those issues for now, Woodward turns to defining contributing causes:

> Necessary Condition for X to be a contributing (type-level) cause of Y $\boldsymbol{NC^*}$
> If X is a contributing type-level cause of Y with respect to the variable set $\boldsymbol{V}$, then there is a directed path from X to Y such that each link in this path is a direct causal relationship; that is, there are intermediate variables along this path, $Z_1 \ldots Z_n$, such that X is a direct cause of Z_1, which is a direct cause of … Z_n, which is a direct cause of Y. Put differently, if X causes Y, then X must either be a direct cause of Y or there must be a causal chain, each link of which involves a relationship of direct causation, extending from X to Y. *(MTH, p. 57)*

Note that transitivity is not assumed here. According to Woodward, if there is a role at all for transitivity or related requirements in causation it is in defining sufficient conditions for causation, not necessary conditions like $\boldsymbol{NC^*}$. Note the ‘if there is a role at all’, for a number of examples in the literature exhibit failures of transitivity. One example is the Dog-Bite example due to McDermott (1995, p. 531) in which assuming transitivity would result in the claim that a dog biting of the right forefinger of a terrorist causes the terrorist to detonate a bomb. This claim is based on the idea that the dog biting off the terrorists’ right forefinger causes the terrorist to use his left finger to push a button, which in turn causes the bomb to detonate. Hence, Woodward does not want to assume transitivity of causation, and instead requires, in addition to the directed path from X to Y required in $\boldsymbol{NC*}$, that Y is sensitive to changes in X if X is to be a contributing cause of Y. This handles the Dog-Bite example satisfactorily, for the detonation of the bomb is not sensitive to whether the dog bites off the right forefinger of the terrorist.

Putting all of this together suggests the following definition of Woodward's manipulability theory:

> Manipulability Theory $\boldsymbol{M}$
> A necessary and sufficient condition for X to be a (type-level) *direct cause* of Y with respect to a variable set $\boldsymbol{V}$ is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in $\boldsymbol{V}$.
> A necessary and sufficient condition for X to be a (type-level) *contributing cause* of Y with respect to variable set $\boldsymbol{V}$ 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 $Z_1, \dots, Z_n$ such that X is a direct cause of Z_1, which in turn is a direct cause of Z_2, which is a direct cause of $\dots Z_n$, 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 $\boldsymbol{V}$ 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 $\boldsymbol{V}$. *(MTH, p. 59)*

It seems that $\boldsymbol{M}$ requires quite some work if one wants to know whether X causes Y, especially in more complex causal structures. This is because $\boldsymbol{M}$ requires one to try out various combinations of values for the variables in **V** until one has established that X indeed causes Y (or until all possible values have been tried and X apparently does not cause Y). Therefore, Woodward proposes two shortcuts to verifying whether $\boldsymbol{M}$ is satisfied. The first shortcut is the following:

1. Draw a directed graph using $\boldsymbol{DC}$
2. Check all routes P_i by
    a. freezing at least one intermediate variable at each of its possible values along all other routes containing intermediate variables between X and Y, and
    b. freezing all direct causes of Y that are not on a route from X to Y (again at each possible value)
3. If for some combination of those frozen values an intervention on X changes Y, then X is a contributing cause of Y.

And the second shortcut:

1. Draw a directed graph using $\boldsymbol{DC}$
2. Check all routes P_i by freezing all direct causes of Y that are not on path P_i
3. If for some route P_i, this freezing allows an intervention on X to change Y, then X is a contributing cause of Y.

It may seem like $\boldsymbol{M}$ is a rather complex theory of causation, but the underlying idea is in fact quite simple: X is a contributing cause of Y with respect to $\boldsymbol{V}$ if an intervention on X changes Y when the right variables are held fixed at the right value.

MTH §2.5: Causal Claims as Telling Us What Happens under Some (not all) Interventions

One may wonder why $\boldsymbol{M}$ only requires one possible intervention on X with respect to Y that changes the value of Y—given that the values of other variables are held fixed—in order for X to count as a cause, direct or contributing, of Y. Should we not require that all interventions on X with respect to Y result in changes in Y? Woodward gives two reasons why we should not.

The first reason is that there may be interventions that do not change the value of X sufficiently to obtain a change in Y. For example, suppose that there is a 180° switch that causes a lamp to be off whenever the position of the switch is less than 90°, and causes a light to be on whenever the position of the switch is more than 90°. In that case, intervening on the position of the switch by changing it from 30° to 60° will not change the state of the lamp. Requiring that all possible interventions should meet $\boldsymbol{M}$ would thus result in mistakenly classifying such and similar causal relationships as non-causal. This brings us to a more general point: stating that X causes Y is not very informative. A more informative statement about the causal relation between X and Y would include the structural equations in the causal structure under consideration.

A second reason is that causal relationships will almost certainly have exceptions. Taking the example of Hooke's law describing the relationship between the extension of a spring and the force the spring will exert ($F = -K_S X$), though this may usually be an accurate description of the causal relation between the extension of a spring and the force the spring will exert, extreme interventions that break the spring, for instance, will not result in the spring exerting the force predicted by Hooke's law. This does not mean that Hooke's law does not describe a causal relationship; instead it means that it does not always hold. This relates to Woodward's notion of invariance, which roughly refers to the idea that a causal relationship may hold under some interventions or background conditions, but not all. Section V discusses this notion more extensively.

MTH §2.7: Actual Causation

So far, Woodward's discussion focussed only on type-level causation, not on actual causation (or token causation). In this section, Woodward sketches how his manipulability approach could be extended to include an account of actual causation.

As noted before, the difference between type-level and actual causation is that type-level causation abstracts away from particulars, whereas actual causation focusses on particulars. In his discussion of actual causation, Woodward assumes knowledge about type-level causal relationships. The question is what this knowledge (in combination with background knowledge) implies for actual causation.[8]

Woodward first provides an initial, rather restrictive definition of actual causation:

Actual Causation $\boldsymbol{AC}$

(AC1) The actual value of $X = x$ and the actual value of $Y = y$.

(AC2) There is at least on route R from X to Y for which an intervention on X will change the value of Y, given that other direct causes Z_i of Y that are not on this route have been fixed at their actual values. (It is assumed that all direct causes of Y that are not on any route from X to Y remain at their actual values under the intervention on X.) *(MTH, p. 77)*

If those conditions are met, $X = x$ is an actual cause of $Y = y$.

The two conditions are illustrated using the well-known case of the desert traveller. In this case, person A poisoned the water in the traveller's canteen with cyanide, and person B punctures a hole in the canteen due to which the water drains out. Drinking the poisoned water would have killed the traveller, but instead he dies of dehydration because of the hole that person B punctured. Figure III-4 illustrates the situation more precisely.

[8] Woodward admits that there are cases where we can have knowledge of actual causation without knowing something non-trivial about type-level causal relationships. Those cases are not discussed in this section.

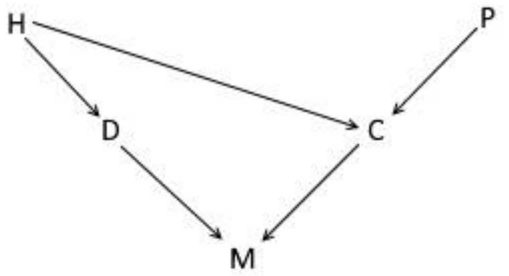


Figure III-4

<u>*Structural Equations:*</u>
$C = P * \neg H$
$D = H$
$M = C \vee D$

<u>*Actual values:*</u>

$H = 1$	$C = 0$
$P = 1$	$D = 1$
	$M = 1$

Where H indicates whether a hole is punctured in the travellers' canteen (1=true; 0=false), D whether the traveller is dehydrated, P whether the water in the canteen is poisoned, C whether the traveller ingested cyanide, and M whether the traveller dies. The structural equation for C says that the traveller ingests cyanide unless a hole is punctured in the canteen. Whether the traveller is dehydrated depends on whether a hole is punctured in the canteen, and whether the traveller survives depends on whether he is dehydrated and / or poisoned.

By using $\boldsymbol{AC}$, we can now verify whether the poisoning—P—or the puncturing of the hole—H—was an actual cause of the traveller dying. It can be readily verified that condition (AC1) is met for both H and P. Starting with P, verifying the second condition amounts to fixing all variables not on the route from P to M at their actual values except for P, which is set by an intervention to 0. This intervention on P does not change the value of M: M remains true, or 1, because C remains true. It follows that $P = 1$ is not an actual cause of $M = 1$. Turning to H, the necessary variables should again be fixed at their actual values, and an intervention on H sets its value to 0. This results in a change in M: given that in the actual situation, C is false, and given that the intervention on H results in D taking on the value false as well, M takes on the value false. Hence, puncturing the hole in the canteen caused the traveller to die.

Note that whether a variable is an actual cause does not depend on whether it is a type-level cause. The poisoning of the water is a type-level cause of the desert travellers' death, because it is possible to set the relevant variables at certain non-actual values—specifically, fixing $H = 0$—, which enables one to change whether or not the desert traveller dies by changing whether the water is poisoned. Actual causes are hence not necessarily type-level causes (and vice versa).

There are, however, cases that $\boldsymbol{AC}$ cannot handle. Those are cases of symmetric overdetermination: two putative causes that simultaneously cause the same effect. A stock example is that of two marksmen who execute a prisoner by shooting him simultaneously. Given that it does not matter whether either or both of the marksmen shoot, $\boldsymbol{AC}$ will judge none of the marksmen to be a cause of the prisoners' death: intervening so that the first marksman does not shoot will still result in the prisoners' death due to the second marksman, and similarly for an intervention on whether the second marksman shoots.

The solution for this problem comes from Hitchcock (2001) and Halpern and Pearl (2000), and consists of introducing the notion of a *redundancy range*. Given some actual causal situation, variables that are not on the directed route under consideration may be set to non-actual values on the condition that this does not change the value of the putative effect. The range of values of the variable(s) meeting this condition is called the redundancy range. The intuition behind this idea is that there is some causal factor that 'masks' the presence of some other causal factor of causal effect. Allowing variables to be set to non-actual values amounts to unmasking such 'hidden' causal effects.

Applying this to the symmetric overdetermination case above, if we are considering the route from the variable indicating whether marksman one shot to the variable indicating whether the prisoner dies, we may set the value of the variable indicating whether the second marksman shoots to a non-actual value as long as this does not change whether the prisoner dies. This is the case here: if we stipulate that the second marksman does not shoot, the prisoner still dies because of marksman one. After fixing that the second marksman does not shoot, we may verify whether the shot of the first marksman is a cause of the prisoners' death by intervening on this shot. Given that there is now a clear pattern of counterfactual dependence between the shot of the first marksman and the death of the prisoner, we may conclude that the shot of marksman one caused the death of the prisoner. Note that we can repeat this exercise with the second marksman as putative cause, hence both shots count as causes of the prisoners' death.

Incorporating the notion of a redundancy range in the formal definition of actual causation results in the following two conditions:

Actual Causation $\boldsymbol{AC^*}$

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

(AC*2) For each directed path P from X to Y, fix by interventions all direct causes Z_i of Y that do not lie along P at some combination of values within their redundancy range. Then determine whether, for each path from X to Y and for each possible combination of values for the direct causes Z_i that are not on this route and that are in the redundancy range of Z_i, whether there is an intervention on X that will change the value of Y. (AC*2) is satisfied if the answer to this question is "yes" for at least one route and possible combinations of values within the redundancy range of Z_i. *(MTH, p. 84)*

One thing that Woodward notes here is that not all people will share the same intuitions about whether $\boldsymbol{AC^*}$ handles the above case satisfactorily. Some people may feel that only one shot can be the cause, or may have some other intuition about the case. Woodward suggests that in such ambivalent cases it is better to rely on patterns of counterfactual dependence than on intuitive common-sense.

MTH §2.8: Causation, Omissions, and Serious Possibilities

Consider the following example due to McDermott:

> Suppose that I reach out and catch a cricket ball. The next thing along in the ball's direction of motion was a solid brick wall. Beyond that was a window. Did my action prevent the ball hitting the window? (Did it cause the ball to not hit the window?) Nearly everyone's initial intuition is, "No, because it wouldn't have hit the window irrespective of whether you had acted or not." To this I say, "If the wall had not been there, and I had not acted, the ball would have hit the window. So between us—me and the wall—we prevented the ball hitting the window. Which one of us prevented the ball hitting the window—me or the wall (or both together)?" And nearly everyone then retracts his initial intuition and says, "Well, it must have been your action that did it—the wall clearly contributed nothing." *(McDermott 1995, p. 525)*

Why do most people initially judge that catching the ball did not prevent the window from breaking? Woodward follows Collins (2000) in arguing that the reason for this causal judgment lies in the possibilities people are prepared to take seriously. For instance, consider the counterfactual "If the fielder had not caught the ball and the wall had not stopped the ball (e.g. by disappearing), then the window would have been shattered". This is a counterfactual constructed along the lines of $\boldsymbol{M}$ that shows that for some value of the intermediate variable 'Wall stops the ball' in this situation (specifically, a value indicating that the wall did not stop the ball), intervening on whether or not the fielder catches the ball changes whether the window shatters. Hence, accepting this counterfactual would yield the conclusion that the fielder indeed caused the window to not shatter. However, is it a serious possibility that the wall does not stop the ball because it disappeared? Woodward argues that it is false to claim that the fielder prevented the window from shattering because we do not (or should not) want to take seriously the possibility of the wall disappearing (or the ball simply passing through the wall and similar unlikely possibilities). Therefore, we should only take into account the counterfactual where the wall does stop the ball—"If the fielder had not caught the ball and the wall had stopped the ball, then the window would not have been shattered"—or where the fielder catches the ball and the wall does not disappear—"If the fielder had caught the ball and the wall had not disappeared, then the window would not have been shattered". According to those counterfactuals, whether or not the fielder caught the ball would not have made a difference to whether the window shattered or not. Hence, the fielder catching the ball is not a cause of the window not shattering.

It should be noted that whether a possibility is a serious possibility or not is not a dichotomous issue; seriousness of possibilities comes in degrees. A number of things influence whether or not we are prepared to take a possibility seriously. This includes the probability of some event happening given certain background conditions, expectations, moral requirements, customs, and perhaps technology. Woodward deems it unlikely that one can be very precise about when we will or ought to take seriously a possibility, hence some vagueness will likely remain.

What, then, are the implications for $\boldsymbol{M}$, given that there are clear patterns of counterfactual dependence in this example? Apparently, only the notion of counterfactual dependence is not sufficient to capture our ordinary causal judgments; we need some notion of serious possibilities. This, however, raises a worry: whether a possibility is to be taken seriously remains a subjective and somewhat vague matter. Countering this worry, Woodward argues that though some subjectivity may play a role, objectivity is also involved: walls do not suddenly disappear, and balls do not pass through walls. Furthermore, Woodward claims that if this need to employ some notion of serious possibilities is an important flaw in his theory, it is a flaw that many theories of causation share or will share.

Summarizing the discussion above, one can say that patterns of counterfactual dependence exist and are objective. Causal judgments, on the other hand, combine patterns of objective counterfactual dependence with our (partially subjective) judgments of serious possibilities, and are hence not completely objective.

IV. MTH Chapter 3: Interventions, Agency, and Counterfactuals

In chapter two, Woodward formally developed his manipulability theory of causation. However, a number of core concepts are still in need of further development, and some additional consequences of Woodward's manipulability theory require further discussion. This third chapter focusses on making precise the notion of an intervention, which is a central concept in Woodward's manipulability theory. Following his general argumentative strategy, Woodward starts with an intuitive definition of an intervention, after which he turns to providing a formal definition. The remainder of the third chapter is used for discussing some additional features and implications both of the notion of an intervention and of Woodward's manipulability theory.

MTH §3.1: Interventions Characterized

Before giving a precise and formal definition of an intervention, Woodward discusses what idea he wants to capture by the notion of an intervention. This idea is that "an intervention on some variable X with respect to some second variable Y is a causal process that changes the value of X in an appropriately exogenous way, so that if a change in the value of Y occurs, it only occurs in virtue of the change in the value of X and not through some other causal route" (*MTH*, p. 94).

Taking the example of a drug experiment, an intervention on the variable 'subject received treatment' with respect to the variable 'the subject recovers' should be a causal process (e.g., the subject gets the drug administered) that is appropriately exogenous (i.e., administering the drug to a subject should not go together with some other form of treatment that influences recovery).

An intervention in Woodward's framework, say on variable X with respect to Y, takes place via an intervention variable I. This intervention variable must be the single cause of the variable that is intervened on (that is, variable X); all other causal relationships between the variable intervened on and its causes are removed. Furthermore, the intervention variable should not be correlated with any other causes of Y, and if the intervention variable I is to be a cause of Y, this causal relationship must go through X.

Woodward formalizes those ideas in two steps. The first step is to define an intervention variable $\boldsymbol{IV}$, which can be used in the second step to give a definition of an intervention $\boldsymbol{IN}$[9]:

> Intervention variable $\boldsymbol{IV}$
>
> I1. I causes X
>
> I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
>
> I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the I-X-Y connection itself; that is, except for (a) any causes of Y that are effects of X (i.e., variables that are causally between X and Y), and (b) any causes of Y that are between I and X and have no effect on Y independently of X.
>
> I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X. *(MHT, p. 98)*

'Cause' in this definition should always be interpreted as a contributing cause. Continuing with the second step, an intervention is defined as follows:

> Intervention $\boldsymbol{IN}$
>
> I's assuming some value $I = z_i$, is an intervention on X with respect to Y if and only if I is an intervention variable for X with respect to Y and $I = z_i$ is an actual cause of the value taken by X. *(MHT, p. 98)*

Requiring all this from intervention variables and interventions ensures that if an intervention on some variable X with respect to Y changes the value of Y, X is indeed a cause of Y.

Note that the notion of an intervention on X is only defined relative to some other variable Y. Hence, I_1 may be an intervention variable for X with respect to Y_1 but not with respect to Y_2. Another important feature of an intervention is that it is nonanthropomorphic. That is, no reference to human action is made (though it is also not excluded). This is a major difference with earlier manipulationist accounts, where the anthropomorphism was usually criticized as a weak point of manipulationist theories of causation. A third important feature of both $\boldsymbol{IV}$ and $\boldsymbol{IN}$ is that there is some circularity in the definitions, given that the concept of causation is used in the definitions. It is also non-reductionist, for it does not aim to analyse the concept of causation by using non-causal terms. According to Woodward, this is to be expected from theories of causation. For example, analysing causal relations in terms of correlations does not work because a given set of correlations is compatible with multiple causal systems. Woodward continues by claiming that the circularity in his account is no vicious circularity. The fundamental idea is to "explain what it is for a relationship between X and Y to be causal by appealing to facts about other causal relationships" (*MTH*, p. 105). Verifying whether there is a causal relationship between X and Y thus does not assume anything about the causal relationship between X and Y, it only appeals to other causal relationships. This circularity, Woodward claims, does not pose problems.

Given that a central element of Woodward's theory is changing the values of variables, it follows that all causal claims must be change-relating. That is, there should be a well-defined notion of what it is like to change the value of the variables figuring in the system. For example, the generalization claiming that physical objects cannot go faster than the speed of light is not a valid causal claim, for the object being physical does not cause it to move slower than the speed of light. The reason for this is that it is not well-defined what it means to change a physical object to a non-physical object. This also means that any generalization involving the physicality of an object does not qualify as a valid causal claim. In contrast, the variable defining whether or not an

[9] Note that $\boldsymbol{IV}$ is a type-level causal notion, whereas $\boldsymbol{IN}$ is a token-causal notion.

experimental subject received treatment is change-relating: there is a well-defined notion of what it is to administer a drug to an experimental subject. Though he would like to, Woodward cannot give a general and precise characterization of what it is for a change in the value of a variable to be well-defined. Often, Woodward claims, this will be based on an intuitive understanding.

Woodward emphasizes that the role of $\boldsymbol{IN}$ is best thought of as a regulative ideal: it tells us what should be true of a relationship between two variables if it is to count as causal, and, in that way, helps us to think about what we should establish if we want to verify the truth or falsity of a causal claim. If it is not possible to actually construct an experiment with an intervention meeting the conditions specified in $\boldsymbol{IN}$, but one still wants to engage in causal inference, one should—according to Woodward—think of oneself as constructing a hypothetical experiment in which the conditions in $\boldsymbol{IN}$ are satisfied.

MTH §3.2: Causal Claims and Hypothetical Experiments

A manipulability theory of causation, Woodward writes, suggests that the meaning of a causal claim can be clarified by thinking about what hypothetical experiment would be associated with that claim. As long as it is unclear what hypothetical experiment can be associated with some causal claim, that claim itself remains unclear. Such unclarity will usually be due to either the non-existence of a well-defined notion of what it is to change the value of some variable, or because it is not made clear what would happen under some intervention. Thinking about a causal claim as a hypothetical experiment helps to identify the issues that makes the causal claim unclear.

MTH §3.3: Realism about Causation

Woodward's manipulationist theory of causation is realist about causation. In contrast to antirealist or subjectivist theories of causation, which claim that causation is something that is projected on the world (similar to how colour is projected on the world), Woodward's theory says that causation is out there in nature. As discussed before, despite his realism about causation, Woodward explicitly incorporates one subjective element his theory: the possibilities that we want to take seriously or not. Once we fix which possibilities we are prepared to take seriously, however, the rest is realist and objective. This is how it should be, argues Woodward, for the counterfactuals that form the basis for causal claims are mind-independent in that they have truth-values independent of what humans may believe about them. For example, whether or not an increase in atmospheric pressure causes a storm is independent of my beliefs or desires about the relationship between atmospheric pressure and the occurrence of storms. Counterfactuals are thus entirely objective; causal claims may incorporate some subjectivity by considering what counterfactuals are serious possibilities.

MTH §3.5: In What Sense Must Interventions Be Possible?

Recall that Woodward's manipulability theory is non-anthropomorphic, i.e. does not refer to human abilities and capacities. Hence, the notion of possible interventions in $\boldsymbol{M}$ does not refer to what is humanly possible. One constraint on what possible interventions are is the requirement that it should be well-defined what it is like to change the value of a variable figuring in the situation at hand (discussed in this section under *MTH §3.1*). This implies that interventions must at least be logically and conceptually possible.

Do interventions need to be physically possible? Such a requirement, Woodward argues, would rule out some counterfactuals what we would like to be able to use. This is because some physical processes that are necessary to realize some counterfactual may not be fine-grained enough to meet the conditions specified in $\boldsymbol{IV}$. Hence, Woodward does not want to incorporate such a requirement in his theory. Not including a 'physically-possible' requirement is, moreover, in line with the characterization of the notion of an intervention $\boldsymbol{IN}$ as a regulative ideal (discussed in this section under *MTH §3.1*): its function is to clarify what we mean with causal claims and tell us what we are trying to establish when we want to verify the truth or falsity of some causal claim. As long as there is some basis that we can use to establish the truth or falsity of the relevant counterfactual claims, the associated interventions need not be physically possible.

MTH §3.7: Some Additional Consequences of a Manipulability Theory

Three additional consequences, or features, of Woodward's manipulability theory are worth noting. First, Woodward's theory naturally has a contrastive focus. This is because it focusses on changing the values of variables from one value to the other, implying a contrastive focus between the situation where a variable had its original value and the situation in which an intervention sets the value of this variable.

A second feature is that Woodward's theory does not rely on some notion of lawhood. No specific interpretation of what counts as a law is required to hold; it is only required that generalizations are reproducible.

A third consequence is that spatiotemporal continuity or connectedness is not necessary for causation: all that counts—at least for Woodward's theory—are counterfactual dependencies (and serious possibilities). Hence, Woodward's theory can handle claims involving omission, double prevention, and other cases involving lack of spatiotemporal continuity. This, Woodward claims, is a virtue of his theory.

V. MTH Chapter 6: Invariance

The chapters summarized in the previous sections develop Woodward's manipulability theory of causation and his notion of an intervention, and discuss some additional implications and features of his approach. However, one important concept still needs to be developed in more detail—the notion of invariance. This sixth chapter is devoted to developing and discussing this notion in detail. Following his general argumentative strategy, an intuitive description is presented first, after which the notion is characterised more precisely, and some aspects are elaborated upon. The summary presented here concerns only the sections of the sixth chapter insofar relevant for Woodward's theory of causation; sections pertaining to his theory of causal explanation are left out.

MTH §6.1: Introduction

The last chapter discussed (partially) in this thesis elaborates on the notion of invariance. Intuitively, a relationship is invariant when it is (approximately) stable under changes. This is a key feature that relationships must possess in order to count as causal relationships. If a relationship or generalization lacks invariance, it is not possible to use such a generalization for the purposes of manipulation and control, and hence does not qualify as a causal generalization.

MTH §6.2: Invariance Characterized More Precisely

The focus of this chapter is on invariance of change-relating generalizations, i.e. generalizations that describe how the value of some variable changes in response to changes in the value of some other variable. Two types of changes can be distinguished, namely changes in background conditions and changes in variables explicitly figuring in the system under consideration. The latter type of change can be subdivided in interventions and so-called non-I-changes, that is, changes in variables that do not qualify as interventions. For the question of whether some generalization counts a causal, invariance under interventions is the most important type of invariance. This is because many non-causal generalizations are invariant under changes in background conditions or under non-I-changes. Invariance under such changes is thus not helpful in distinguishing causal from non-causal generalizations.

In order to test for invariance, not all interventions are useful. For instance, some interventions on X may be too small to have an effect on Y and are hence uninformative. Recall the 180° switch example discussed in section III, where a lamp is off as long as the switch is on an angle of less than 90°, and on as long as the switch is on an angle of more than 90°. Turning the switch from 40° to 60° will not affect the state of the lamp, and given that the generalization claims that this is the case, one may be tempted to conclude that this generalization is invariant under interventions. The problem here is that it may be the case that the switch is broken, and hence the generalization pertaining to the state of the switch and the state of the lamp may in fact not be invariant under interventions. In order to properly test this, an intervention that according to the generalization will change the value of the putative cause and putative effect variable is necessary. Such interventions are called *testing interventions*. As with regular interventions, it is not necessary that the intervention could in fact be carried out: as long as it is well-defined what it means to change the value of the variable that is intervened on, it is irrelevant whether carrying out some testing intervention is physically possible[10]. Hence, invariance is not an actual but a modal or counterfactual notion. That is, it is not (solely) based on an actual situation but predominantly on the counterfactuals belonging to the situation at hand: in the 180° switch example the invariance of the generalization relating the angle of the switch to the state of the lamp is based on what would happen *if the angle of the switch were changed sufficiently*, not on the actually obtaining angle of the switch or state of the lamp.

To illustrate this notion of invariance, consider an example involving the ideal gas law. This law describes—or at least approximates—the behaviour of gases under various conditions and changes in the values of the variables figuring in the generalization. Hence, this generalization is invariant. However, the ideal gas law does not always hold. For instance, when the temperature of a gas becomes sufficiently high, intermolecular forces become important and the ideal gas law breaks down. This implies that generalizations are not necessarily invariant under all interventions. In fact, usually this will not be the case. Another aspect of the notion of invariance that this example illustrates is that it is relative to particular systems: the ideal gas law will hold for gases, but not for, say, liquids.

MTH §6.4: Degrees of Invariance

Invariance is a concept that comes in degrees; it is not an all-or-nothing matter. While there is a threshold that divides generalizations in invariant and non-invariant generalizations, different invariant generalizations may exhibit more or less degrees of invariance. To what extent a generalization is invariant depends on (1) the range of interventions under which it is invariant, and (2) the importance of the interventions under which it is invariant. Whether or not an intervention is important depends on the subject or domain that the generalization under consideration belongs to. For instance, some interventions that may be considered important in neurobiology may be much less important in the domain of economics or history. One such instance could be the manipulation of an individual's brain structure that is important for generalizations in neurobiology (assuming for the sake of the argument that brain structure is an important concept in neurobiology) but not for generalizations in the field of history (given that brain structure is not a central concept in history).

An example may illustrate the idea of degrees of invariance most clearly. Recall that the ideal gas law (approximately) describes the behaviour of gases under various circumstances, though it ceases to hold under some other circumstances. The ideal gas law is thus invariant under some interventions, but not under all. Another generalization, namely van der Waal's force law, also describes the behaviour of gases under the circumstances in which the ideal gas law works, but in addition to that holds under circumstances where the ideal gas law breaks down as well (e.g. when intermolecular forces are important). The range of interventions under which van der Waal's force law holds is thus greater than the range of interventions under which the ideal gas law holds. Hence, van der Waal's force law has a greater degree of invariance than the ideal gas law (on the plausible assumption that the importance of those circumstances—or, more precisely, those interventions—is more or less equal).

[10] See section IV (*MTH §3.5*) for a discussion on this.

VI. ASSESSING *MAKING THINGS HAPPEN*

The previous sections summarized the chapters and sections relevant for Woodward's manipulability theory of causation. This section sets out to assess *Making Things Happen*, which it does by first describing what are judged to be the strong aspects of Woodward's approach, after which a number of criticisms are developed.

VI-A. Making Things Happen: Strong Aspects

Overall, the manipulability theory that Woodward proposes in *Making Things Happen* is a very interesting approach to understanding causation. His theory sheds much light on our use of the concept of causation, and the (renewed) concepts of interventions and invariance help to clarify what we (should) mean by our causal claims. A number of aspects that, I think, are particularly worth emphasizing as strong aspects of Woodward's theory are discussed here.

I. The motivation for the manipulability approach

One motivation for his manipulability approach that Woodward repeatedly emphasizes (e.g. in *MTH* §1.6 and §2.1) is that it explains why we are interested in causation by appealing to the practical payoffs that it yields. Those practical payoffs consist of being able to manipulate or control something by manipulating or controlling one or more causes of it. This motivation contrasts with the disinterested-intellectual-curiosity motivation that many philosophers would advance. Woodward's argument that this disinterested-intellectual-curiosity motivation is unsatisfactory is compelling: animals and children indeed seem to be interested in causal relations, as opposed to, for instance, correlations, which cannot be explained by appealing to disinterested intellectual curiosity. This argument against the disinterested-intellectual-curiosity motivation combined with the practical-payoffs argument in favour of Woodward's practical motivation make a convincing case for Woodward's approach[11].

A second argument that Woodward advances is that his manipulability approach adequately explains why we use experiments to discover causal relationships. His notion of interventions, for example, is closely related to how experiments are ideally conducted (recall that Woodward intuitively characterized interventions as ideal experimental manipulations). This is another persuasive argument for adopting a manipulability approach.

Besides those two motivations, Woodward also motivates his theory by noting that it claims that there should be continuity between the use of the concept of causation in everyday life, in science, and in applied science. A fourth motivation is that manipulability conditions deliver clearer verdicts in distinguishing causal from non-causal generalizations than intuitions do. A fifth motivation is that his approach shows that no odd or excessive metaphysical baggage is attached to the notion of causation. The two motivations discussed above, however, are in my opinion most convincing for adopting a manipulability approach.

II. Alignment with scientists' ideas about causation

Woodward's manipulability approach to causation aligns well with the conception of causation that many scientists have. Many textbooks in statistics, econometrics, and other social and natural sciences employ a manipulability conception of causation (Woodward, 2005, chapter 2). A number of philosophers may not be impressed with this alignment, and may argue that scientists do not have a sufficiently sophisticated understanding of the issues surrounding the concept of causation to be able to contribute to conceptual and philosophical discussion. This, however, would be uncharitable according to Woodward. Moreover, it is unlikely to be true, for given that many scientists encounter the same or similar issues with causation as philosophers do and that they use the notion of causation throughout their work, a plausible case can be made that

[11] It should be noted, though, that the motivation from practical payoffs is not a motivation that Woodward uniquely developed; it has already been used by other manipulability theorists such as Gasking (1955), Collingwood (1940), von Wright (1971) and Menzies and Price (1993).

scientists' understanding of causation will or at least can be sophisticated enough to make valuable contributions to conceptual and philosophical discussion.

Given this, it seems evident to me that alignment with scientists' conceptions of causation is a virtue of any theory of causation; hence this is another strong point of Woodward's approach.

III. Non-anthropomorphism and the notion of an intervention

A third strong aspect of Woodward's manipulability theory is the non-anthropomorphism, which is related to the notion of an intervention. Previous manipulability often referred to human agency instead of to interventions, in an attempt to provide a non-reductionist account of causation. Given that the anthropomorphic focus is usually seen as one of the weak aspects of manipulability approaches, the non-anthropomorphism of Woodward's theory is a significant improvement over earlier manipulability theories. Moreover, the notion of an intervention that Woodward develops, which roughly does the work that agency notions did in earlier manipulability theories, aligns well with our most important scientific practice of causal inference: experiments. This provides a strong motivation for using this notion of an intervention. That the notion of an intervention seems to do its job properly provides another strong motivation for using this notion.

IV. Woodward's strategy

At the start of sections II, III, IV, and V, I describe Woodward's strategy with regards to how he develops his arguments and manipulability theory of causation in *Making Things Happen*. The core idea behind Woodward's strategy appears to be first introducing a topic or notion intuitively, and subsequently refining and formalizing this topic or notion. Despite this not being a substantive point, I am convinced that this strategy helps readers to understand the formal machinery more easily and makes Woodward's proposal transparent with regards to the underlying intuitions and motivations. Those evidently desirable points achieved by Woodward's strategy warrants giving his strategy a place among the strong aspects of his *Making Things Happen*.

VI-B. Making Things Happen: Criticisms

Despite Woodward's approach being interesting and illuminating, a number of weaker aspects can be identified and criticized. Most criticisms developed here pertain to the notion of serious possibilities, the exception being the last criticism dealing with intuitions, type-level causation, and actual causation.

I. Serious possibilities: strange counterfactuals in MTH

In *MTH §2.8*, Woodward argues that we should not take into account strange counterfactuals, i.e. counterfactuals that do not seem to be serious possibilities to us. This idea was developed further by discussing the Catch-the-ball example due to McDermott (1995) (see chapter III on *MTH §2.8*). This example is represented structurally in figure VI-1.

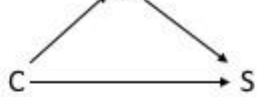


Figure VI-1

Structural Equations:
$W = \neg C$
$S = \neg C \vee \neg W$

Variable C represents whether the first fielder catches the ball, W represents whether the wall stops the ball, and S indicates whether the window shatters. The strange counterfactual in this example arises when one wants to verify whether C causes S. Following $\boldsymbol{M}$, when verifying the path $\{C, S\}$, one should set W to various values and verify whether given this value of W, manipulating C results in a change in S. This results in the following combination of variable values: $C = 0$, $W = 0$, and $S = 1$—fielder does not catch the ball, wall does not stop the ball, and the window shatters in ordinary language. Woodward thinks that if $C = 0$, then $W = 0$ is not a serious possibility given that walls do not magically disappear or allow balls to fly through them.

The problem identified here is that Woodward seems to use strange counterfactuals in other examples in *Making Things Happen* himself. For example, recall the case of the desert traveller discussed in chapter III (*MTH §2.7*), represented structurally in figure VI-2.

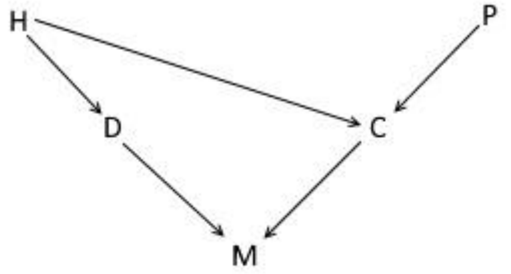


Figure VI-2

<u>*Structural Equations:*</u>
$C = P * \neg H$
$D = H$
$M = C \vee D$

<u>*Actual values:*</u>
$H = 1$ $C = 0$
$P = 1$ $D = 1$
$M = 1$

Again, H indicates whether a hole is punctured in the travellers' canteen, D whether the traveller is dehydrated, P whether the water in the canteen is poisoned, C whether the traveller ingested cyanide, and M whether the traveller dies. Verifying whether H causes M requires one to construct counterfactuals along the line of $\boldsymbol{AC^*}$[12]. Doing so for the path $\{H, D, M\}$ requires fixing C at its actual value—$C = 0$—, and subsequently varying the value of H and verifying whether this results in changes in the value of M. This inter alia results in the counterfactual 'If no hole is punctured in the canteen and hence no dehydration occurs, and if the water is poisoned but the traveller does not ingest cyanide, the traveller would not die'. This is a strange counterfactual, for it does not seem to be a serious possibility that the desert traveller drinks the poisoned water, but nevertheless does not ingest cyanide.

What should we make of this? Apart from noting the inconsistency, there are two options: either Woodward should not use this counterfactual in verifying whether H causes M, or we must conclude that it is inevitable to use strange counterfactuals. The next criticism argues that the second option is the way to go.

II. Serious possibilities: inevitability of strange counterfactuals

Recall that the strangeness in the Catch-the-ball counterfactual discussed above involved the fielder not catching the ball and the window shattering because the wall somehow also did not stop the ball. It is the combination of variable values that makes this counterfactual strange: if the fielder did not catch the ball, it is strange that the wall did not stop the ball. If, in contrast, the fielder had caught the ball, it would not be strange that the wall did not stop the ball, for the ball did not reach the wall in the first place. In more precise terms, the strangeness is due to *both* C and W being equal to zero. As a result, we cannot fix $W = 0$—given that $\boldsymbol{AC^*}$ tells us to vary the value of C, fixing $W = 0$ would inevitably lead to both C and W being equal to zero.

If this is correct, then a serious problem arises. If the combination $C = 0$ and $W = 0$ leads to strangeness in counterfactuals, then so does the combination $C = 1$ and $W = 1$. This is the case because it is not a serious possibility that the fielder catches the ball *and* the wall stops the ball. Either the fielder catches the ball or the wall stops the ball, it cannot be both. Given that fixing $W = 1$ would lead to this combination of variable values, the notion of serious possibilities tells us that we cannot set $W = 1$. But given that setting $W = 0$ is ruled out as well (as discussed above), it appears that the notion of serious possibilities does not let us fix W at either 1 or 0 and

[12] Not along the lines of $\boldsymbol{M}$, given that this is a token-causation example. The issues of serious possibilities pertains both to token and type-level causal relationships.

hence deprives us of the possibility to verify whether C causes S. If this assessment is correct, then it turns out that (at least for this example) it is inevitable to use counterfactuals involving non-serious possibilities.

Is this inevitability to use strange counterfactuals limited to a few examples with a specific structure, or is it a general issue? I believe there is a good reason to think that it is a general issue, which originates from the definitions of $\boldsymbol{M}$ and $\boldsymbol{AC^*}$. Both definitions require one to set some variables to some certain value or values, which amounts to breaking the connection between the variable that is set to some value and its parent. For example, in the Catch-the-ball example, W ceases to depend on its parent C, leading to the strange combinations of neither the fielder nor the wall catching or stopping the ball, and both the fielder and the wall catching or stopping the ball. Similarly, in the desert traveller example the link between P and C is broken, leading to strange combinations of values between P and C (e.g. drinking poisoned water but not ingesting cyanide). Of course, this is no definite proof of strange counterfactuals generally being inevitable for Woodward's manipulability theory, but it does suggest, quite strongly in my opinion, that strange counterfactuals are inevitable in a considerable range of cases.

III. Serious possibilities: mundane cases of omission-causation

Whereas the former criticisms suggest that the requirement of serious possibilities is too strong, this criticism will argue that the serious possibilities-requirement is simultaneously too weak. Recall that Woodward's proposal allows omissions to count as causes. For instance, an assassin not shooting down its target victim counts as a cause for the survival of the target victim. While I agree that allowing omissions to be causes is a virtue of any theory of causation, allowing this also results in judging many what I call mundane cases of omission-causation that cannot be ruled out by the serious possibilities requirement. Examples of mundane cases of omission-causation would include the causal claims 'John not deciding to have coffee causes the coffee cup to remain where it is', 'Ann not waking up causes the bedsheets to remain in place', and 'Ali not using his pen causes the pen to remain full'. While in some situations those causal claims may be ruled out by the serious possibilities requirement (for example, Ann waking up at 03.00 may be too unlikely as a possibility to use for constructing some counterfactual), there will be situations in which those causal claims and corresponding counterfactuals do meet the requirement of serious possibilities. Nevertheless, it seems in the spirit of Woodward's serious possibilities to not count such mundane cases of omission-causation. If this is correct, then it seems a failure of the requirement of serious possibilities that such cases are not ruled out.

A counterargument may be that the three examples of mundane cases of omission-causation provided above are in fact properly classified as causal claims. I must admit that intuitions on the cases may differ. However, given that during any given day many actions are *not* undertaken, it is unlikely that all those actions that were not done will be said to have caused all the effects that resulted from those omissions (in so far not ruled out by the serious possibility requirement). I take this to establish that there are mundane cases of omission-causation that most intuitions will judge not to involve causation.

One may also argue against this that the requirement of serious possibilities was never intended to rule out such mundane cases of omission-causation. While this may be true, this would at most save the notion of serious possibilities of one criticism; the substance of the criticism, however, would still apply to Woodward's proposed manipulability theory of causation.

IV. Serious possibilities: some tensions

Some more tensions surround the notion of serious possibilities. First, it seems that some inconsistency results from this notion. Suppose that we do not want to take seriously the possibility that a meteorite strikes the White House and kills Donald Trump (perhaps because this is a very low probability event). Then suppose that in fact, a meteorite strikes the White House and kills Donald Trump. Clearly everyone would say that the meteorite caused Donald Trump's death. This is odd: how can the absence of a factor not be the cause of survival, even though the presence of that factor is a cause of non-survival? Though this does not seem to be a fundamental problem for Woodward's theory or for the notion of serious possibilities, there seems to be a tension that needs to be explained away by Woodward.

A second tension is the following. According to Woodward, it is obvious that the notion of serious possibilities comes in degrees (2005, p. 88). However, it seems that whether or not a counterfactual is taken into account is a dichotomous issue. Woodward writes that "to the extent that the possibilities that figure in them [counterfactuals] are nonserious, they do not guide our causal judgments" (2005, p. 90), but leaves it unclear to what extent possibilities figuring in counterfactuals have to be nonserious in order to be allowed to guide our causal judgments. Given that one cannot simply convert an issue of degrees into a dichotomous choice, Woodward should have given some guidance there on how to deal with this tension.

The last tension pertaining to the notion of serious possibilities concerns whether common sense should be prioritized over patterns of counterfactual dependence, or the other way around. On page 85, Woodward (2005) writes that "the suggestion I want to make is that to the extent that commonsense causal judgments are unclear, equivocal, or disputed, it is better to focus directly on the patterns of counterfactual dependence that lie behind them—the patterns of counterfactual dependence are, as it were, the "objective core" that lies behind our particular causal judgments, and it is such patterns that are the real objects of scientific and practical interest". The tension is this: the notion of serious possibilities involves commonsense judgments about whether a possibility is to be taken seriously. Hence, on the one hand Woodward wants to take patterns of counterfactual dependence as primary (be it in cases where commonsense judgments are unclear, equivocal, or disputed), but on the other hand commonsense judgments decide which counterfactuals we take into account. In cases where commonsense judgments are clear, this will not be a problem. However, in cases where commonsense judgments about the correct causal claims are unclear, which, given that what counts as a correct causal claim depends on what possibilities are taken seriously, implies disagreement or unclarity about what possibilities should be taken seriously, it is unclear whether commonsense judgments or patterns of counterfactual dependence should be decisive in judging the correctness of causal claims. This is a tension that Woodward leaves unaddressed.

V. Actual causation, type-level causation, and intuitions

On the difference between actual and type-level causation, Woodward writes on page 40 of *Making Things Happen* that "a claim such as "X is causally relevant to Y" [type-level] is a claim to the effect that changing the value of X instantiated in particular spatiotemporally located individuals will change the value of Y located in particular individuals". This is not limited to the putative cause X and putative effect Y, but also holds for intermediary variables. This becomes clear from a discussion in *MTH §2.7* on whether short-circuits cause fires, represented in figure VI-3.

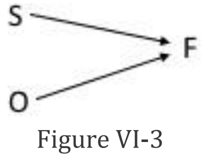


Figure VI-3

Structural Equations:
$F = S * O$

Whether a short-circuit occurs is represented by S, whether oxygen is present is indicated by O, and F tells us whether a fire occurs. On a type-level, verifying whether S causes F involves varying the value of O and verifying whether at some value of O there is an intervention on S that changes the value of F. Given that when oxygen is present, changing whether or not a short-circuit occurs results in changing whether a fire occurs or not, we can say that short-circuits type-level causes fire. Whether in some situation a short-circuit actually caused a fire requires us to specify whether in that particular situation oxygen was present, after which we can verify whether short-circuits caused a fire in this situation by varying whether a short-circuit occurs and checking whether varying this results in changes in F. Hence, whether a short-circuit actually causes fire depends on the presence of oxygen.

Two issues can be identified from this, the first being that the relation between type-level causal claims and actual causal claims as described by Woodward in the quote above is incomplete. Though on a type-level it can be said that short-circuits cause fires, this does not mean that a short-circuit happening in some particular situation will cause a fire, for this depends on the circumstances (or more precisely the values of the other variables in the situation): only when oxygen is present, a short-circuit will actually cause a fire. Hence, Woodward's description of the relation between type-level and actual causation should be adapted as follows: a claim such as "X is causally relevant to Y" [type-level] is a claim to the effect that changing the value of X instantiated in particular spatiotemporally located individuals will change the value of Y located in particular individuals, if the right conditions (i.e., values of other variables in the situation under consideration) obtain. Putting this in simpler words, the claim "X type-level causes Y" means that X *can* cause Y if the right conditions obtain. In contrast, the claim "X actually caused Y" means exactly what it says: X caused Y.

Clearly, then, it is possible that type-level causal claims differ from actual causal claims, even though both are based on the same causal structure. As a result, the second issue is that it is important to be clear about whether a situation concerns type-level causation or actual causation, especially when trying to elicit intuitions: it seems likely that people will have different intuitions when asked whether X *can* cause Y compared to when asked whether X causes Y in some particular situations. This clarity, however, is not always present in *Making Things Happen*. Returning to Catch-the-ball, the way in which the situation is described (see section III in this thesis, on *MTH §2.8*) strongly suggests that Woodward discusses a situation of actual causation. However, the discussion that follows this example—the discussion on serious possibilities—concerns type-level causation.

The effect of this unclarity can be made clear by reconsidering Catch-the-ball in light of this discussion. The type-level causal claim "Catching the ball prevented the window from breaking" means that catching the ball *can* prevent the window from breaking if the right conditions obtain. Given that 'the ball' and 'the window' still suggest a particular situation, it is most accurate to specify the type-level causal claim as "Catching a ball can prevent a window from breaking if the right conditions obtain"[13]. I take it that everyone would agree with this causal claim. But this is not the claim that Woodward discusses, for he (following McDermott) already specifies what conditions obtain (specifically, a wall that is present on the trajectory of the ball). Therefore, the intuitions elicited by this example will be intuitions concerning the actual causal claim "Catching the ball prevented the window from breaking (in this particular situation with a wall present on the trajectory of the ball)". It is evident that intuitions for this actual causal claim will differ from the type-level causal claim "Catching a ball can prevent a window from breaking if the right conditions obtain", which, I argued, is the proper type-level causal claim associated with this example. Given that intuitions are usually considered an important benchmark for theories of causation, it is important that intuitions for the right (type of) causal claim are elicited; intuitions on type-level causal claims may differ strongly from intuitions on actual causal claims, even if it concerns the same causal structure.

VII. DISCUSSION & CONCLUSION

Undoubtedly, the last word about the concept of causation and theories of causation has not been said yet. Despite the manipulationist theory of causation developed in *Making Things Happen* being a very interesting and insightful treatment of causation, the criticisms raised in this thesis show that the conceptualization that Woodward provides is, in some respects, still unsatisfactory. Strong aspects of Woodward's theory include his motivation for employing a manipulability approach (especially the focus on practical pay-offs), the alignment with scientist's ideas about causation, and the non-anthropomorphism that is a considerable point of improvement over previous manipulability theories of causation. Weak aspects of Woodward's theory include the

[13] Compare: a short-circuit can cause a fire if the right conditions obtain.

notion of serious possibilities, against which several criticisms were raised, and the unclarity resulting from discussions on type-level causation illustrated by examples that suggest actual causation.

Despite the manipulationist theory developed in *Making Things Happen* being unsatisfactory in some respects, the strong aspects and the literature that it inspired show that Woodward's work has not been in vain: his useful insights may be used to develop theories of causation that are satisfactory (or at least more satisfactory), and the literature it inspired undoubtedly provide more insights as to how a (more) satisfactory theory of causation may be developed.

VIII. BIBLIOGRAPHY

Collingwood, R. (1940). *An Essay on Metaphysics.* Oxford: Clarendon Press.

Collins, J. (2000). Preemptive Prevention. *The Journal of Philosophy*, 223-234.

Gaskin, D. (1955). Causation and Recipes. *Mind*, 479-487.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and counterfactuals* (pp. 225-276). Cambridge, MA: MIT Press.

Halpern, J., & Pearl, J. (2000). *Causes and Explanations: A Structural Model Approach.* Technical report R-266, Cognitive Systems Laboratory. Los Angelos: University of California.

Hitchcock, C. (2001). The Intransitivity of Causation Revealed in Equations and Graphs. *The Journal of Philosophy*, 273-299.

McDermott, M. (1995). Redundant Causation. *British Journal for the Philosophy of Science*, 523-544.

Menzies, P., & Price, H. (1993). Causation as a Secondary Quality. *British Journal for the Philosophy of Science*, 187-203.

Ruben, D. (1990). *Explaining Explanation.* London: Routledge.

von Wright, G. (1971). *Explanation and Understanding.* Ithaca, NY: Cornell University Press.

Woodward, J. (2005). *Making Things Happen: A theory of causal explanation.* New York: Oxford University Press.