University of Colorado, Boulder
CU Scholar

Philosophy Graduate Theses & Dissertations | Philosophy

Spring 1-1-2015

A Critical Assessment of Woodward's Theory of Causation and Explanation

Erlantz Etxeberria Altuna
University of Colorado at Boulder, erlantze@hotmail.com

Follow this and additional works at: https://scholar.colorado.edu/phil_gradetds

 Part of the Epistemology Commons, Metaphysics Commons, and the Philosophy of Science Commons

Recommended Citation
Etxeberria Altuna, Erlantz, "A Critical Assessment of Woodward's Theory of Causation and Explanation" (2015). *Philosophy Graduate Theses & Dissertations*. 47.
https://scholar.colorado.edu/phil_gradetds/47

This Thesis is brought to you for free and open access by Philosophy at CU Scholar. It has been accepted for inclusion in Philosophy Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.

A CRITICAL ASSESSMENT OF WOODWARD'S THEORY OF CAUSATION AND EXPLANATION

by

Erlantz Etxeberria Altuna

B.E., University of the Basque Country, 2008
B.A., Autonomous University of Madrid, 2012
B.A., University of the Basque Country, 2013

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Master of Arts
Department of Philosophy
2015

This thesis entitled:

A Critical Assessment of Woodward's Theory of Causation and Explanation

written by Erlantz Etxeberria Altuna

has been approved for the Department of Philosophy

Carol Cleland

Graeme Forbes

Raul Saucedo

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Etxeberria Altuna, Erlantz (M.A., Philosophy)

A Critical Assessment of Woodward's Theory of Causation and Explanation

Thesis directed by Professor Carol Cleland

Abstract

James Woodward develops in his book *Making Things Happen* (2003) an account of causation and explanation based on the concept of intervention. An intervention, in turn, parallels the concept of experimentation—widely used in the sciences—and offers an intuitive grasp of the more complex notions of causation and explanation, so Woodward's theory could provide a methodologically useful basis for scientific practice involving causation. In this essay I argue that both Woodward's theory of causation and of explanation suffer from important metaphysical problems as well as crucial practical difficulties, and they are ultimately unsuccessful. Metaphysical problems of his theory of causation derive from the fact that an intervention is essentially a counterfactual notion, but Woodward fails to determine with precision what are the truth conditions for counterfactuals. In addition, his theory lacks a clear notion possibility to make sense of those counterfactuals. The practical aspect of his theory of causation as a methodology for science also suffers from serious problems, since in order know a specific causal claim it is required to have a vast background knowledge about other causal relations, demanding too high epistemic standards. Finally I argue that Woodward's theory of explanation, which comprehends both causal and non-causal explanations, faces a decisive counterexample and is flawed too.

CONTENTS

INTRODUCTION

The aim of this thesis is to provide a critical assessment of Woodward's theory of causation and explanation. James Woodward develops in his book *Making Things Happen* (2003) an innovative and influential account of causation and explanation based on the concepts of manipulation and intervention. It is innovative because unlike previous manipulationist theories Woodward presents a non-reductive account that uses counterfactual interventions instead of human agency to elucidate causal claims. It is influential because its approach to causation resembles the actual scientific practice of experimentation and it has been widely used in the social sciences. These and other advantages make it a theory worth investigating.

However, I argue that Woodward's theory of causation and explanation suffers from important metaphysical problems as well as crucial practical difficulties, and it is ultimately unsuccessful. The problems fall in two categories: those that undermine his theory of causation and those that affect his theory of explanation.

The thesis is divided into two chapters. In the first chapter I summarize Woodward's theory of *causation*, and I highlight some of the virtues that make it an attractive alternative as well as some of its problems when it comes to difficult cases such as preemption or mere correlations. Then I argue that Woodward's project has several omissions and problems, including the absence of truth conditions for counterfactuals, the high epistemic standards required and the lack of a clear notion of possibility.

In the second chapter I present Woodward's theory of *explanation* and note that it is a general and comprehensive theory that accounts for both causal and non-causal explanations. While it is a promising theory it suffers from a fatal counterexample that forces us to abandon it.

CHAPTER 1: WOODWARD'S THEORY OF CAUSATION

James Woodward, in his book *Making Things Happen* (2003), presents a theory of causation and explanation based on a manipulationist understanding of causation. He first develops an account of causation as a dependence relation between variables such that if there were an intervention on one of the variables it would result in a change on the other variable. Woodward then exploits his characterization of causation to provide a theory of explanation that aims to cover a wide range of explanatory instances both in science and ordinary life, including problematic cases that other theories of explanation cannot deal with, thus making his account an attractive alternative approach to causation and explanation (I cover the topic of explanation on Chapter 2).

In this first chapter I focus on his theory of causation and I detail its central elements, including counterfactual dependence, interventions, as well as the motivations and advantages of his project. The last section of the chapter is devoted to problems and counterexamples to his theory, and I conclude that despite some virtues as a methodological project, it ultimately fails to provide a robust metaphysical and epistemic basis for his theory. This is aggravated by the fact that a vaguely articulated account of causation results in a less powerful tool as a methodological device.

1. A Counterfactual Notion of Causation

If we were to place Woodward's manipulationist theory of causation in one of the existing traditions, that would be the counterfactual one. Causation is understood roughly as a matter of counterfactual dependence between variables, such that given two variables X and Y,

> the claim that X causes Y means that for at least some individuals a possible manipulation of some value of X that they possess, which, given the appropriate conditions (perhaps including manipulations that fix other variables distinct from X at certain values), will change the value or probability distribution of Y for those individuals (2003:40).

The idea is that X is the cause of Y if some changes on the variable X result in some changes in the variable Y, that is, if X has an impact on what happens to Y. Two important elements make this account a prima facie attractive one.

First, the qualification introduced by "at least for some individuals" aims to capture causal relations that are only displayed within a certain range of values of the variables involved. For example, suppose our two variables are *x*: 'heating water to temperature *t*', which can take different values (in Celsius), and *y*: 'water boils,' which can take the values 'yes' or 'no.' Let us imagine we perform a series of experiments where we manipulate the temperature variable *x* and set it in different degrees, without exceeding 80°C. The result is that the variable *y* has the value 'no' or 'doesn't boil' for all these experiments, thus suggesting that there is no causal relation between heating the water and its boiling. The key in Woodward's formulation is that at least for some individuals there *is* a possible manipulation of the temperature, namely, heating the water over 100°C that changes the value of the variable of *y* (boils) from 'no' to 'yes.'

The second element is given by the qualification "given the appropriate conditions," which refers to how exactly a manipulation should be performed in order to capture all and only causal instances. For example, a given counterfactual relation (unless it's clearly explicated) will not straightforwardly distinguish between causes and simple correlations, since some counterfactual statements about correlations seem to be true. To use a stock example, suppose we

know that a low atmospheric pressure causes both a low reading in a barometer and the imminent formation of a storm. As it happens, a counterfactual like "if the barometer's reading hadn't been low there wouldn't have been a storm" may well be true depending on how we are to make sense of the meaning of counterfactual conditionals. However, since the barometer's reading and the storm are both effects of a common cause (the low atmospheric pressure), these two variables are merely correlated. The barometer's reading doesn't cause the storm, so either the counterfactual must turn out false or counterfactual dependence isn't sufficient for causation.

David Lewis's (1973) well-known theory of causation opts for the first option and relies on a semantics he developed in terms of possible worlds to render these counterfactual statements about correlations false. Woodward takes a different route. Because mere counterfactual dependence isn't sufficient for causation for him, Woodward isn't interested in giving a general account of the semantics for counterfactuals, although he is interested in their truth conditions as I will explain later. Instead, in order to capture causation, he introduces an additional element in counterfactual relations, namely, *interventions*, which are the right kind of manipulations performed on the variables to ensure that causal claims pick genuinely causal relations and not mere correlations. Roughly, a counterfactual relation that makes use of interventions is the right kind of counterfactual to describe a causal relation. This assumes that there are some counterfactuals that are causal (those which involve interventions) and some that are not. The resulting sufficient condition for causation for Woodward is the following: "if there is a possible intervention that changes the value of X such that carrying out this intervention (and no other) will change the value of Y, or the probability distribution of Y, then X causes Y" (2003:45). Even though the sufficient condition isn't stated using the grammar of counterfactual conditional the counterfactual dimension is embedded in the notion of possibility and

intervention employed by Woodward, which transcends actual, technological or physical possibility (see section 4 below). Let us see how Woodward characterizes the notion of intervention and how it aims to capture causation.

2. Interventions

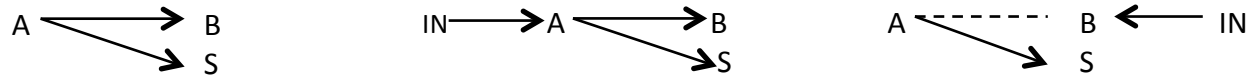
An intervention is the right sort of manipulation performed on one or more variables in a causal structure. That is, not all manipulations are interventions but all interventions involve some change in some variables; the difference between an intervention and an unqualified manipulation is that the former has precise requirements about how the change is carried out, including the requisite that the change is causally produced. The best way to understand interventions is to think of them as a set of causal manipulations that modify the value of some variables while they keep other variables fixed, so they constitute a surgical manipulation in the sense that they produce definite changes in the causal structure.

Woodward identifies four specific conditions that an intervention variable, *IN,* must meet. Suppose that there is a causal relation between variables *X* and *Y* and that according to the formulation of causation given in the previous section, if there is an intervention that changes the value of *X* and thereby *Y* also changes, then *X* causes *Y.* So, *IN* is an intervention variable of *X* with respect to *Y* if and only if (2003:98):

i. *IN* causes *X*
ii. *IN* must be the only cause of *X* (i.e. *IN* breaks the relation of other variables that cause *X*)
iii. *IN* must not cause *Y* via a route that doesn't go through *X*.
iv. *IN* is statistically independent of any variable *Z* that causes *Y* (unless *Z* is between *X* and *Y*)

The motivation behind these clauses is easy to see. The first (i) simply tells that the value of the intervention variable *IN* is the cause of the change in the variable *X*. This is a way to ensure that there is a possible change in variable *X* and that it is possible to carry it out causally. The importance of the intervention being causal is that it allows us to make sense of the counterfactual relation between *X* and *Y* in that we are able to understand the counterfactual situation where *X* has a different value due to *IN,* and thereby we are able to assess how the variable *Y* would respond to this change.

Clause (ii) tells that when *X* is manipulated, whatever other cause was previously producing *X*'s value ceases to influence *X*, such that the new intervention is the only relevant cause. As Woodward puts it the value of "the intervention variable thus 'breaks preexisting links between *X* and its cause' (2003:100). The requirement is important because it demands that the intervention is exogenous to the causal structure we are interested in, and this is crucial in distinguishing mere correlations from causal relations. Let us have a look again to the example of the barometer and the storm, where a low atmospheric pressure A is the common cause of both a low reading in the barometer B *and* a storm S. We expect that the right sort of manipulation on the barometer shouldn't result in a change on the storm. Let us represent it graphically:

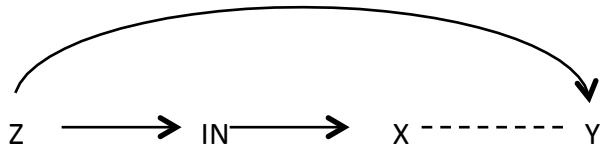


The structure on the left shows how things stand: A causes both B and S where arrows stand for causal relations. The structure on the center is the wrong kind of manipulation on the variable B with respect to S, since it doesn't comply with the clause (ii) given that we'd be altering the reading of the barometer by increasing or reducing the atmospheric pressure, which is precisely the preexisting cause of B. The result of this manipulation is that changes on B (via A) are

followed by change in S, which would mistakenly yield B-S as a causal relation. The picture on the right represents how the right sort of intervention should be performed in this causal structure, where the dashed line stands for a broken link. It is easy to imagine how this could be done. Suppose we make an intervention on the barometer manually moving the actual position of the hand to a different position. This intervention renders the atmospheric pressure irrelevant with respect to the variable B (assuming that the barometer doesn't have a strong enough mechanism to move the hands in accordance to the atmospheric pressure). As a result, the intervention won't have any impact on the storm variable, so the sufficient condition for causation isn't fulfilled in this case and it correctly describes the connection between B and S as non-causal.

Clause (iii) ensures that the intervention has effects only on the variable *X* and that it doesn't cause *Y* independently of *X*. That is, the intervention itself cannot be a common cause of both *X* and *Y*, so if *Y* changes after the intervention, it must change only due to a change in *X*. This clause is helpful to correctly capture some causal structures like that between a drug and the recovery from an illness for example: if the intervention in the variable '*drug'* involves administrating a pill to the patient and she is aware of it, then regardless of the effectiveness of the drug, there will be a change in the variable of recovery (or the probability distribution of it) caused by the placebo effect. That is, the manipulation to administer the pill to the patient independently (via placebo) causes a change in her recovery, which is precisely what clause (iii) forbids, and hence this kind of manipulation won't count as the right sort of intervention to elucidate the causal relation. The right kind of manipulation should account for the placebo effect to comply with (iii).

Finally, clause (iv) is similar to (iii) in that it guarantees that by making the intervention, *Y* isn't affected via another causal route except for the one that goes through the variable *X*. But condition (iv) applies to the causes of the intervention too, since it requires that *IN* is independent of *Z*-variables that cause *Y*. This prevents getting the wrong causal relations in more complex structures like the following one where the intervention *IN* on variable *X does* produce a change in *Y*, but only due to an independent causal relation existing between *Z* (the cause of the intervention) and *Y*:



The requirements (i)-(iv) and the associated examples of causal structures show that causation characterized in terms of interventions is a very precise formulation of what it is for *X* to be a cause of *Y*, and this precision is what allows to accommodate complex causal structures like the ones above in an interventionist way. In section 5 below I go over some other advantages of this account of causation followed by some crucial problems in section 6. However, before assessing Woodward's theory of causation I will point out additional relevant features in sections 3 and 4 to complete the overview of Woodward's view.

3. Goals and Motivations of Woodward's Account

Having sketched the main features of Woodward's account of causation, one might have noted some problems the theory faces. Before going over them it is helpful to mention some of the motivations and goals of Woodward's project, as well as its scope and possible limits.

As Woodward observes, manipulationist approaches to causation have been far more common among scientists than among philosophers. Economists (Hoover 1988), statisticians (Holland 1986) and natural scientists for example think of causal relations in terms of manipulation and control, and for good reason according to Woodward. He claims that "it is heuristically useful to think of explanatory and causal relationships as relationships that are potentially exploitable for purposes of manipulation and control" and that our causal claims are "informed by our interests as practical agents in changing the world" (2003:25). Moreover, alluding to Michael Dummett's (1964) example, he argues that if we were intelligent trees with observing and reasoning capacities but without the ability to manipulate nature, we would not have a robust notion of causation in the first place.

Causation and explanation are therefore practical activities of human beings, and Woodward insists that all cultures have made use of causal claims and inferences in a intuitive way (2003:20). From an evolutionary point of view this suggests that there must be some benefit in engaging in causal reasoning, and a theory of causation and explanation should also account for this manifest practical dimension. Moreover, it is also a common practice for scientists to perform an experiment in order to find out the relations existing between certain variables in a similar fashion to Woodward's interventions: holding certain variables fixed while manipulating the initial value of the interested variable and measuring the results. The guiding idea is that "causal relations are relations that can be used for manipulation and control" (2003:113).

These remarks exhibit, first, why manipulation and causation are two intimately related notions, and second, they determine both the motivation and the goal for his manipulationist version of causation, which seem to be that this intuitive approach to causation fits in with a robust philosophical account of causation and explanation. I say a robust philosophical account

because whereas some of Woodward's desiderata are explicitly pragmatic, his theory makes both metaphysical and epistemological commitments and assumptions too, and he engages in discussions with alternative theories of causation and explanation in the philosophical tradition, such as Lewis' counterfactual theory of causation or Hempel's deductive nomological theory of explanation. As a result, his account consists of an eminently practical motivation supplemented by an attempt to provide some metaphysical and epistemological basis, although as I shall argue later, he doesn't achieve neither of these goals.

Put differently, Woodward seems less concerned about giving a metaphysically complete account of causation than he is about making sense of causation in an intuitive way and in a way that we can use causal relations for the purposes of manipulation and control. However, this is not to say that he doesn't make any metaphysical commitments. He does some, and while some of them may be rather vague, he assumes that these are enough to get his theory off the ground. The following section overviews the most significant metaphysical commitments: realism about causation, the notion of possibility, invariance of causal relations, non-reductivism and causes as variables.

4. Metaphysical Commitments and Other Features of Causation

Realism. Woodward advocates a realist theory about causation. For him causation is a relation that exists in the objects independently of human beings, and he rejects subjectivist approaches that argue that causation is just an additional feature that our minds project into the world, and that it is up to our psychological theories to explain why we acquire the notion of causation beyond the perceived regularities. Woodward objects that such an explanation would fail to account for the evolutionary benefits of distinguishing, for example, between genuinely causal

processes and mere correlations, for if the difference between them is subjective and doesn't exist in nature it's hard to see why we're interested in them. Similarly, it would also imply that the truth conditions of causal claims are subjective or mind dependent, and that conducting experiments to determine causal relations would only give us information about "the experimenter's (or the scientific community's) projective activities" (2003:119) and not about the world. Thus, Woodward assumes realism about causation in developing his theory.

Notion of possibility. The formulation that 'C causes E' involves there is a possible intervention in C. We have already seen the details of how an intervention is characterized. Let us now look at what sense of possibility is required of interventions according to Woodward. A good way to understand the notion of possibility he has in mind is by setting the limiting cases. On the one extreme, the notion of possibility is not restricted in no way by human possibility. Following the objectivist stance Woodward takes, he also aims to overcome some problems that subjectivist manipulationist versions have faced in the past. In particular, agency theories developed in the last decades of the 20th century, such as von Wright's (1971) and Menzies and Price's (1993), gave human agency and the human ability to manipulate or act upon the world a central role in analyzing causation, such that the latter could be analyzed in terms of the more basic notions of human manipulation. The main problem with these strategies is that they are restricted to cases where a human agent is able to manipulate a variable, so they can't account for those causal relations. For example, it's not humanly possible to make a volcano erupt at will, but we do take the ash that follows an eruption as an effect of it. Theses strategies would also involve that in a world without humans there wouldn't be causal relations. Woodward takes a different route and doesn't give human agency a substantial role in his interventionist account. None the four

conditions described earlier for an intervention require that the changes in variables are possible for human beings to carry out, and they don't require that the interventions are actually or technologically possible either, since causation would still exist regardless of our capacities.

On the other extreme, an intervention has to be at least logically possible. Logical possibility sets the limiting case, although he also considers strong and weak physical possibilities even though he ultimately rejects them due to some counterexamples they face. For instance, suppose there is a given causal relation between C and E and suppose in addition that C has no causes. Since interventions are causal manipulations, there is no physically possible intervention that changes the value of C. So he rejects physical possibility too: "to make whether C causes E depend on whether interventions C are physically possible seems to be to make [the causal relation] depend on what sorts of causal histories are possible for C itself, and this consideration seems extrinsic and irrelevant to the nature of the relation between C and E" (2003:130). Therefore, the only strong notion of possibility required seems to be just the logical possibility that there is an intervention. This conclusion, however, leads to a crucial problem that I address later in section 6.3.

Invariance of causal relations. Another feature of causal relations is that they are invariant. Invariance is a property of the relation between causes and effects, and it ensures the relation remains stable when interventions are carried out and the values of the related variables change. It plays a role similar to the notion of a law of nature in that they both guarantee that the same relation will hold in future instances or counterfactual situations, but without requiring the degree of generality, comprehensiveness and precision that is expected of laws. In fact, Woodward claims that invariance "does not require exact or literal truth; I count a generalization as invariant

or stable across certain changes if it holds up to some appropriate level of approximation across those changes" (2003:239). That is, if a relation remains stable within a certain range of the values of variables, it suffices for invariance, while a law of nature seems to require a greater degree of generality. For example, the way physical objects behave near the surface of the Earth (e.g. accelerating at a rate of $9.8m/s^2$) is invariant under these circumstances, but will not hold in other planets, whereas Newton's gravitational laws do hold in different planets and are more comprehensive.

This apparent laxity of invariance with respect to laws is a virtue according to Woodard because, since invariance is sufficient to account for causal relationships, and there is no need to posit universal laws of nature to back every causal relationships. Woodward is skeptical about laws of nature for at least two reasons. First, it is unclear what a law of nature is and what are the requirements it should fulfill in terms of being universal and exceptionless, for even plausible candidates for a law such as 'water boils at 100°C' will have exceptions if the atmospheric pressure changes. A possible answer may be that while 'water boils at 100°C' isn't a law, it is backed by a true law of nature like the ideal gas law and others. But this brings him to the second reason: if a theory relies on laws of nature to account for causation it is again unclear to see how this is possible for a mundane causal relation such as 'my tipping the desk with the knee caused the ink to fall.' It isn't straightforward to find out what laws of nature support the causal claim, for it would be a bit of an exaggeration to claim that it is a law of nature the generalization that 'tipping desks with knees makes inks to fall.' The alternative solution, which Woodard refers to as the 'hidden structure strategy' involves relying on some genuine laws of nature (perhaps Newton's laws or other physical laws) backing up the causal claim that the tipping of the table with the knee made the ink fall, but he argues that believing in this hidden structure for every

causal claim may well require additional justification. In addition, he argues that from an epistemic point of view this strategy requires that someone must know the "laws or structures that underlie singular-causal (or other sorts of causal) claims" (2003:180), while intuitively (and according to his interventionist theory too), one doesn't need to have access to e.g. Newton's laws in order to know that tipping the desk caused the ink to fall in a given case.

Non-reductivism about causation. The notion of invariance just discussed is essentially a counterfactual one for Woodward: "a relationship would remain stable if, perhaps contrary to actual fact, certain changes or interventions were to occur" (2000:112) so it is not a surprise that counterfactuals cannot be reduced to invariance relations (unlike, for example, theories that attempt to account for counterfactuals in terms of laws of nature). The same is true of causation: causation is characterized by Woodward making use of the notions of counterfactuals and interventions but doesn't aim to reduce causation to these or any other more basic concepts. In fact, recall that interventions are themselves defined in terms of causal relations (see clause (i) above for instance), so in Woodward's account the notion of causation pervades other concepts in the vicinity that are used to elucidate the former.

He openly regards his project as an non-reductivist one and refuses the metaphysical aspirations of reducing the notion of causation to more basic notions. However, he argues that a theory that doesn't reduce causation to something else isn't automatically uninteresting, for it can be "illuminating without being reductionist" (2003:20) in that it shows how precisely these notions are non-reductively related. Against the objection that non-reductive theories are circular and uninformative he responds that his own is not "viciously circular in the way the explaining 'cause' in terms of a primitive notion of 'production' would be" (2003:22).

Causes as variables. Finally, Woodward doesn't make any specific ontological commitments and remains neutral regarding what kind of thing a cause is beyond its being a variable. Causal relata must be capable of having different values, but these values may well be numerical values (for variables like distance, temperature, etc.) or binary values such as yes/no or occurred/didn't occur. Events and processes for example, can be easily represented as variables with binary values thereby accommodating an event-ontology that is typically used in accounts of causation, but there is nothing in Woodward's account that requires all causes be events or to be translatable into events.

Despite this ontological neutrality some consequences do seem to follow from this characterization of variables. For example, a necessary event won't count as a variable in Woodward's theory, given that, of necessity, it will always have the same value. So for example, if god necessarily exists, then the existence of god is unmanipulable, because it's impossible to change the value from 'exists' to 'doesn't exist', so god's existence is not a variable proper. And if it's not a variable, it can't be neither a cause nor an effect according to Woodward's theory, which would produce interesting consequences for discussions in the philosophy of religion: e.g. the existence of god can't be the cause of the universe or the cause of herself. None of these are explored by Woodward, but it would be interesting to see what other metaphysical consequences follow from the metaphysical commitments of his theory.

This concludes the exposition of Woodward's theory of causation. In what follows I will first focus on some of its advantages, and I will then present some objections.

5. Advantages of Woodward's Theory of Causation

There are many important problems with Woodward's theory of causation as I shall argue in section 6, but it would be unfair to ignore some of the advantages that have contributed to make his theory an influential one in the social sciences. A few of these have already been mentioned in the previous sections, so I will briefly revisit those and highlight some additional virtues—although I also raise concerns that are associated with some of them, especially because they come at a certain cost.

In accordance with Woodward's motivations for his theory, the interventionist account is an intuitive way of thinking about what we mean when we say that C causes E. It also captures nicely how some scientists and economists think of causation in terms of relations between variables that can be tested in experiments, so Woodward's theory promises a higher degree of practicality than other philosophical approaches that focus on a metaphysical account.

Relatedly, from a methodological point of view, scientist are more likely to be familiar with concepts used in Woodward's theory such as 'experimentation,' 'manipulation,' 'variables' and 'invariance' rather than with the intricacies of the notion of causation. Accordingly, it could be argued that since his is not a reductionist theory, the strategy does seem helpful to illuminate the notion of causation if one is more familiar with the other notions and has a minimum understanding of the notion of causation (otherwise 'intervention' wouldn't be understood either, since it involves causation).

However, someone could be disappointed by this non-reductionist project and object that a metaphysical account of causation should do more than elucidating or illuminating the concept of causation by means of other notions that already presuppose causation. Nonetheless, the triangle formed by causation, laws of nature and counterfactuals are also notorious for being

difficult to define without falling in some sort of circularity, and Woodward's strategy consist in acknowledging this difficulty while aiming to clarify other aspects of causation. But as I shall argue later in section 6.1 and 6.2, even if such a circle were not entirely vicious, it presupposes a high degree of prior knowledge about causation and causal relations of the world, which casts doubt on the usefulness of Woodward's theory.

Another potential merit is its comprehensiveness to accommodate all sorts of causal claims without having to make too many ontological commitments. I argued earlier that Woodward's theory *does* make some ontological assumptions and tries to provide some metaphysical basis for his theory. However, there are many other metaphysical aspects that are left undecided and that could lead to some advantages. For example, there is nothing in Woodward's account that rejects spatially and temporally distant causes, or the temporal priority of causes over effects. Some people may find the notion of action at a spatial or temporal distance counterintuitive and some accounts of causation such as Salmon's (1984) causal mechanical model requires that causal processes are spatiotemporally contiguous. However, causal claims such as "the fish I ate yesterday caused my stomachache today" aren't straightforwardly true in Salmon's theory unless a full story of the spatiotemporally continuous interactions and processes is provided. In contrast, Woodward's account is compatible with there being such a story relating the fish and the stomachache but the causal claim may as well be true as it stands. Causal reasoning in the social sciences can especially benefit from this comprehensiveness in that it accommodates a vast number of causal claims that we ordinarily use in a simple and intuitive way. The clear drawback, however, is that usefulness and comprehensiveness come at the expense of a vaguer account of causation and a less complete

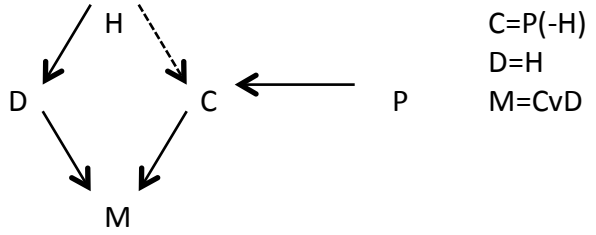
metaphysical theory. Whether or not Woodward provides sufficient metaphysical basis for his theory will be addressed more extensively in the next section.

A comparable virtue is not having to require laws of nature for every causal claim that intuitively seems to be true. This is not to say that there aren't laws of nature backing causal relations—again, there might or might not be. Woodward simply notes that if there are such laws of nature it isn't required to know or specify them in order to provide a true causal claim relating C and E. Disposing of laws of nature avoids the related problem of defining the standards and requisites for what it takes to a generalization to be a genuine law (e.g. whether or not allow exceptions to the generalization, whether the law has *ceteris paribus* conditions, etc.), although obviously, not all philosophers think of laws as problematic elements.

Finally, additional advantages result from the ability to accommodate challenging counterexamples that are used as test cases for a theory of causation. For example, I already explained how Woodward's account is able to correctly distinguish between causal relations and correlations by means of the very specific notion of intervention and clauses (i)-(iv). Woodard also offers solutions to cases of preemption and overdetermination. The solutions often involve recognizing that the notion of causation is far more complex than how it's generally understood, and Woodward introduces more precise notions of causation such as *actual cause* (the precise variable that caused an effect in a given instance, as opposed to type cause) and *direct cause* (X is a direct cause of Y if and only if there is a possible intervention on the value of X such that the value of Y also changes while all other variables are held fixed).

Preemption. The problem of preemption occurs when an effect Y is actually caused by X, but in the case where X was absent, another cause Z would have anyway caused Y. Therefore, the

counterfactual that "if X wouldn't happen, Y wouldn't have happened" is false, which apparently implies that X is not the cause of Y. The problem is called preemption because X's actual happening preempts Z's actual causing Y. Here's an often used example: traveler T goes to the dessert, whose water canteen has both a hole H and cyanide poison P. The water drains due to the hole, and the traveler dies-dehydrated D before even drinking the cyanide. So, H causes D. However, if H wasn't the case, yet the traveler would have died, so apparently, a change in H doesn't make any change in D, and therefore, H didn't cause D. One part of the solution in this case according to Woodward is to distinguish the traveler's death as "death-by-dehydration", a binary variable D, as opposed to "death-by-cyanide", also a binary variable C. In addition, we have the variable M for death *simpliciter*. The following graph represents the structure, and the equations on the right column represent the dependence relation between the values (T/F) of the variables:



The preempting causal relation from H to C is the dashed line. When we do that, and we distinguish between different causes we see that: P is a direct cause of C, but not an actual cause. H is a direct cause of C too, because when P is true, yet H will determine C. H is a direct cause of D, for the truth of D depends only on the truth of H. And both D and C are direct causes of M. But did H cause M? Here's the second part of the solution: since we are dealing with "actual causes" rather than type causes Woodward specifies some criteria for an actual cause: (2003:77)

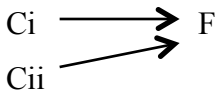
AC1: Actual value of X=x and Y=y, and

AC2: There's at least one route from X to Y, where intervening on X will change Y, given other direct causes of Y (Zi…) are held fixed at their actual values.

In our example, AC1 is satisfied because H is true and M true, and AC2: is satisfied, because there is at least the route H-M via D, where changing H (make it false) we get a change in M (from T to F), given that the alternative route, (via C) is held fixed (false), so that M=false because D=false and C=false. By applying again AC1 and AC2 to P, he concludes that P isn't the actual cause of M, roughly, because since T didn't actually drink any water, changing whether there was cyanide or not is irrelevant to the death M (i.e. no change in M).

The solution offered by Woodward manages to avoid the unwanted consequence that an irrelevant cause in the actual situation (the poison) turns out to be a cause of the death too. The apparatus to solve the problem, however, it is complex and convoluted. It requires defining new notions of cause and provide some specific rules to apply it. In addition, it is an ad hoc move when in the example above C is held fixed in 'false' for it requires to be done by changing the value of P, that is, it requires that for a variable to maintain fixed there have to be more than one manipulation that counteracts the effect of H in C. Again, there are some costs for the solution.

Overdetermination. Overdetermination occurs when two causes are independently sufficient for the effect. For instance, two campers Ci and Cii simultaneously throw a lighted cigarette each, and each cigarette is sufficient to start a fire F.



Woodward argues that, intuitively, both Ci and Cii equally are causes of F. However, since we are dealing with an actual cause rather than a type level cause, if we apply the conditions of actual cause (AC1 and AC2), we see that Ci didn't cause the fire because an intervention changing the actual value of Ci wouldn't result in a change in the actual value of F. And the same is true for Cii. So it seems *none* of them is an actual cause.

Woodward's first solution is to revise the AC2 criterion, such that it allows to freeze the variables in direct causes of Y (other than the X-Y route we're assessing) at some *non-actual* value instead of requiring to hold values fixed at actual values. Thus, we freeze Cii=false (non-actual) and go on to assess Ci, in the counterfactual "If Ci changed, F would change", which is true, thereby satisfying AC2 and so concluding that Ci is an actual cause of F. The same applies to Cii.

However, Woodward isn't satisfied with this solution because it is too permissible and renders non-causal relations as causal too. As a result, he proposes a new notion for the actual cause AC* whose condition AC2* is that "for each directed path from X to Y, fix by interventions all direct causes (Zi…) of Y [that aren't in X-Y route] at some combination of values within their redundancy range." (2003:84) This would allow us to determine then whether there is an intervention in X that will change the value of Y. If so, AC2* is satisfied.

Like in the case of preemption, the number of ad hoc requisites to deal with the problem of overdetermination make the solution extremely convoluted. The way in which variables are fixed at non-actual values is also unclear when these are not binary variables. At any rate, the solution involves abandoning the initial aspiration to provide an intuitive notion of causation that is in line with the actual scientific practice.

6. Problems with Woodward's Theory of Causation

I have characterized Woodward's theory of causation as a project that combines some practical motivations with some metaphysical claims that furnish it. There is a reasonable question as to how strong his theory is and whether it is enough to achieve the goals set, since there are important omissions with respect to other metaphysical aspects. Perhaps the eminently practical spirit of his theory justifies that less weight is placed on the metaphysics, but as I argue below, even the practical and methodological virtues of his theory are unsuccessful.

6.1 The Problem of Truth Conditions for Causal Counterfactuals

Counterfactuals are central to both Woodward's theory of causation and explanation.. However, Woodward is far from telling us in a clear way what are the truth conditions for counterfactuals or how are we supposed to evaluate them. Following the distinction I will make in chapter 2 between causal and non-causal explanations (depending on whether the counterfactual involves an intervention or not), there seems to be, respectively, causal as well as non-causal counterfactuals. The problem for the truth conditions of causal counterfactuals is more acute because unlike logical, geometrical or semantic relations, causal relations don't hold for all possible counterfactual situations. So, what exactly make these counterfactuals true?

First, note that causal counterfactuals are to be interpreted as there being a possible intervention that manipulates the variable that appears in the antecedent of a counterfactual. Now if we were able to actually carry out an experiment and make the appropriate intervention, then it would be easy to determine the truth conditions for the counterfactual just by looking at the effects. However, Psillos (2008) notes that this works only to determine the truth conditions for

actual experiments, but not for *counter*factuals, so it won't achieve the goal. In addition, although this strategy seems to fit well with the general practical spirit of his theory, Woodward claims that it is sufficient for an intervention to be just logically or conceptually possible to carry out, so in no way are counterfactuals to be restricted to actualized possibilities: "there must be a way of disentangling—perhaps merely conceptually or analytically rather than in actuality—the effect on E of changing just C from the effects on E of changes in other potential confounding variables, including direct effects from the intervention process itself" (2003:131).

In fact, by definition, counterfactuals will never be actualized, so their truth conditions cannot be dependent on actual experiments:

> We think instead of [this counterfactual] as having a determinate meaning and truth value whether or not the experiment is actually carried out – it is precisely because the experimenters want to discover whether [the counterfactual] is true or false that they conduct the experiment (2003:122).

A more promising alternative is to let causal claims support counterfactual conditionals, that is, causal relations that exist in the world account for the truth conditions of counterfactual situations. Woodward seems to favor this option when he writes:

> According to the manipulationist account, given that C causes E, which counterfactual claims involving C and E are true will always depend on which other causal claims involving other variables besides C and E are true in the situation under discussion. For example, it will depend on whether other causes of E besides C are present (2003:136).

When Psillos (2008) considers this possible solution he quickly notices that the account seems circular, because causal claims are elucidated by means of interventions and counterfactual conditionals while counterfactual conditionals seem to be determined by causal claims. So it seems that in analyzing a causal relation like C causes E and after considering what counterfactuals are involved in this relation we end up having to appeal to the causal claim itself to account for the truth conditions of counterfactual.

Nonetheless, Woodward seems to be aware of this problem and could respond to it pointing out that the circularity doesn't arise, since as the above quotation shows, the counterfactuals involving C and E will depend in *other* variables besides C and E. In another passage, Woodard also notes that the apparent circularity doesn't affect our epistemic possibilities, since the notion of causation "is not viciously circular in the sense that we already have to know whether there is a causal relationship between X and Y (or what its characteristics are) to apply them" (2003:105).

However, despite not being a reductive project, the apparent circularity where causal claims are specified appealing to other causal notions or relations still deeply problematic. The solution Woodward proposes is that given the characterization of what an intervention is, vicious circularity does not arise. He notes that the causal relationships that elucidate the claim that "X causes Y" are *other* causal relationships exogenous from the relation between X and Y, and therefore there is no vicious circularity so long as one causal relation depends on several other relations but not in itself. But another problem remains: this would generate a regress instead of a vicious circle. The regress, in turn, need not be a problem for truth conditions, but it does levy severe burdens for the pragmatic dimension of Woodward's project. I tackle this in the next section.

6.2 The Problem of the High Epistemic Standards

I have emphasized the that many of Woodward's theory's virtues come from the intuitive understanding of causation and explanation as performing hypothetical experiments on the variables we are interested. In fact, he conceives of causal explanation "as a practical activity" (2003:18) but since not all experiments are actually realized there must be some other way to assess the counterfactuals involved in those hypothetical experiments if we want to preserve the practicality of the theory. Woodard is not explicit in saying what requirements should be met to have the desired epistemic access to counterfactual claims, he just points that "all that is require is that we have some sort of basis of assessing the truth of claims about what would happen if an intervention were carried out" (2003:130).

But what is this basis? I take it that full epistemic access to causal and counterfactual claims would be possible if we had access to their truth conditions—this would ensure our knowledge to assess what would happen under hypothetical interventions. In the previous section we saw that the only possible understanding for truth conditions was one that generated a regress, and while that was not necessarily a problem for truth conditions, it does pose a threat to the possibility of knowledge of causal and counterfactual claims—it certainly seems to exceed reasonable epistemic standards due to the vast number (perhaps infinite) of other causal claims that one must be familiar with.

The characterization of the notion of an intervention epitomizes the extremely high epistemic standards that his theory of causation and explanation would require. Recall that interventions are central to Woodward's account and that he carefully characterizes them in order to fulfill a precise role in his causal theory claims as shown in the clauses i-iv laid in the

first chapter, section two. In turn, each of these clauses presupposes several epistemic preconditions if we want to use the interventionist counterfactual to assess its truth value.

Clause (i) would require first that we must be aware of a *possible* cause of X, namely the intervention that causes the change in the value of X. Clause (ii) tells that *IN* must disrupt previous causal relations that cause X's so it follows that we must know what the *actual* causes of X are. From (iii) it follows that we must know the effects (even distant ones) of our intervention performed on X that are causally related to Y. Finally, since the intervention must be itself brought about somehow, (iv) requires that we must know at least one cause of the intervention. Let's call this *IN'*. We therefore must know effects of *IN'* (even distant ones) and their causal connection or lack of with respect to Y. Now since *IN'* has to be somehow brought about too, say by *IN''* we would also need to take into account whether or not *IN''* have effects on Y, and so on for an apparently infinite series of causes.

So a single intervention presupposes the knowledge of vast number of other causal relations and structures: it is crucial that these causal structures are characterized correctly if we want to use them as epistemic tools for causal claims. Let us illustrate this point considering once again the example of the barometer and the storm. Suppose we don't know whether the barometer's reading B causes the storm S, but we do see typically a correlation between a falling reading in B and the occurrence of a storm S. We wonder whether B causes S. If it is the case that B causes S then there is an intervention in B such that it results in a change in S. So we could start first manipulating B's reading by lowering the atmospheric pressure P, which involves that we know at least some causes of B. However, it wouldn't count as an intervention because the atmospheric pressure also causes S. This implies that we already know what causes S and what are the effects of A. So, assuming we already know all of this, the right sort of

intervention on B would be M: manually moving the reading to a different number since it breaks the connection between A and B. Again this involves we know that M causes B and that M overrides A. Finally, since M needs to be carried out too, we would need to know what causes it, for example, my finger's pushing P will suffice to move the barometer's hand, and again, since moving my finger must be caused by something else (perhaps electric currents in my muscles) we would need to know those as well.

The point is that in order to find out whether B causes S we must already posses an vast knowledge of other relevant causal claims. In addition, if at some point we do not posses the required causal knowledge do determine whether B causes S, we may want apply the same strategy for these unknown causal claims: e.g. we can ask for example whether A causes S or not. But again this requires other causal knowledge. So in order to avoid infinite regress, at some point there must be some brute causal facts that we learn. I concede that this is not to say that such causal knowledge is impossible, in particularly because as a matter of fact it seems that we are aware of some causal relations without further considerations of their causes and effects. For example, when the intervention is performed on the barometer by moving the reading with my fingers, I just know how to make my finger move, without further identifying causes and effects of my finger's movement.

The latter suggestion is a more accurate representation of our epistemic access to causal claims, and Woodward probably agrees with the claim we don't really need to know all causal relations to assess a hypothetical experiment—he thinks that "it is realistic to suppose that our background knowledge often enables us to do this" (2003: 99). Background knowledge would therefore suffice for some cases, and we would certainly be able to use the notion of an intervention to gain knowledge of further causal claims: "we do seem to sometimes find out

whether a causal relationship exists between X and Y by manipulating X in the appropriate way and determining whether there is a correlated change in Y" (2003:105).

The worry, however, remains intact for more complex causal structures where our background knowledge of causal claims is much more reduced. In fact, unless we know what the underlying causal structure looks like we just can't be sure whether a manipulation performed on X is a genuine intervention or whether it violates one or more of the clauses i-iv. Another example considered in chapter 2, section 4 about whether my being an uncle is causally dependent on by my sister's giving birth clearly shows that the assessment of certain causal relations depends on how we devise the existing causal structure.

6.3 The Problem of the Notion of Possibility

Finally, Woodward fails to give a proper notion of possibility when he considers in what sense an intervention should be possible. He first seems to be satisfied with a notion of physical possibility, but due to some counterexamples he ends up extending the notion of possibility to that of logical possibility. One counterexample is the case where the moon causes the tides in the ocean, so physically possible manipulation that would change the moon's position (e,g, by a huge mass nearby that alters its orbit) would also need to change the tides in the ocean due to the attraction of the huge mass. The second counterexample Woodward notices (and already mentioned earlier) is the case of an uncaused event, which doesn't admit physically possible interventions either.

On the face of these counterexamples Woodward makes two moves that are fatal for his theory. First, it admits that a notion of logical possibility the only remaining candidate, but this

entirely undermines the rationale of using interventions to elucidate the meaning of counterfactuals. Recall that the advantage of introducing the notion of an intervention was the power to make sense of counterfactual conditionals. They could be used to understand and evaluate what would happen if an intervention of the right sort were carried out, and it is easy to see how this helps to make sense of counterfactuals. For example, counterfactuals such as "if the sun hadn't risen it would be dark" are difficult to evaluate because it is unclear what they mean. We can inquire for example whether the sun didn't rise because the counterfactual is located in a planet without a sun, whether it didn't rise because it never set, or because it just exploded during the night. The truth value of the counterfactual is different in each case, and the great benefit of interventions is that they provide some restrictions on these possibilities, and hence some guidance in evaluating them: they enable us think of this or that event happening in the form of an intervention, and we evaluate the outcome accordingly. However, if the notion of possibility for interventions is merely logical, then, there will always seem to be a logically possible intervention, even on non-causal relations, and they won't be helpful anymore to understand and evaluate the counterfactuals they are involved in.

The second problematic move is the response Woodward provides to the situation where the only available notion of possibility is the logical one. Perhaps he is aware of the problems that follow from it that I just described in the paragraph above, so he quickly notes that:

> the notion of an intervention characterized by IN represents a regulative ideal. Its function is to characterize the notion of an ideal experimental manipulation and in this way to give us a purchase on what we mean or are trying to establish when we claim that X causes Y. We have already noted that for this purpose it isn't necessary that an intervention actually be carried out on X. All that is required is that we have some sort of

> basis for assessing the truth of claims about what would happen if an intervention were carried out (2003:130).

This is a rather unsatisfying answer in that Woodard doesn't engage in a philosophical discussion nor tries to answer the counterexample by giving a solution to it, but rather opts for qualifying the IN as a mere regulative ideal. It may be noted that it is not entirely clear what exactly is meant here by a regulative ideal, but the most worrying consequence of this move is that it notably diminishes the importance of the notion of an intervention in his theory. Recall in the description given in section 1.2 how this notion is precisely defined and how its role turns out to be indispensable to give us the right sort of counterfactuals that capture causation. Without them, counterfactuals may just give us correlations, but not causal relations. But now with interventions being just a regulative ideal, it seems that Woodward is saying that we shouldn't really care about how the conditions 1-4 are strictly met, which is at stark contrast with the spirit of his project where interventions are key for causation. If their role is to be a regulative ideal it seems unnecessary that the counterfactuals interpreted in an interventionist fashion are able to cover the most difficult cases or the most complicated causal structures. Nor seems necessary to provide all sorts of details about them.

CHAPTER 2: WOODWARD'S THEORY OF EXPLANATION

Although Woodward's project is mainly focused on causal explanations he is aware of explanations that do not allow a causal interpretation—in his manipulationist sense—and yet seem explanatory. That is, there are both causal and non-causal explanations.

For Woodward an explanation provides information that tells us what would happen to the explanandum if the variables in the explanans were different in some way. This, in short, involves spelling out the counterfactual dependence holding between certain variables:

> an explanation ought to be such that it can be used to answer what I call a *what-if-things-had-been-different question* [w-questions]: the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways. We can also think of this as information about a pattern of counterfactual dependence between explanans and explanandum, provided the counterfactuals in question are understood appropriately (2003:11).

In the course of this chapter I aim to substantiate what it takes to understand counterfactuals appropriately for Woodward. The requirement is that the counterfactuals at stake can be interpreted in an interventionist fashion, but this will only apply to *causal* explanations and not other kinds of non-causal explanations.

Since Woodward's theory of causation is largely motivated by the problems that Hempel's Deductive Nomological model faces, it will be appropriate to present first this view. Next, I describe Woodward's theory of causal explanation in section 2 and explain how it overcomes the counterexamples to the DN model. Then, I focus on non-causal explanations as a way to motivate a general theory of explanation that comprehends both causal and non-causal

explanations. In section 4 I present Woodward's general theory of explanation and highlight its advantages. Finally, the section 5 is devoted to two crucial objections that pose an irreconcilable dilemma to Woodward's general theory of explanation.

1. The Deductive Nomological Model of Explanation

Carl Hempel and Paul Oppenheim (1955) provided a very influential account of explanation that dominated the discussion about explanation for some decades in the second half of the 20th century. Their basic thesis is that explanations are logical arguments consisting of two main elements. The conclusion of the argument comprising an explanation contains the *explanandum*, i.e. a statement of the phenomenon that needs to be explained and of which we ask *why*. The statements that comprise the premises of the argument are called the *explanans*. The explanans is the part that gives the relevant explanatory information, just like the premises of an argument give the relevant information to warrant the conclusion.

This approach is known as the 'covering law theory' because a necessary condition for an explanation is that one of the premises in the argument is a law of nature, that is, that the explanans contains a universal statement that subsumes the explanandum under a generalization. This statement can be either a general law of nature or a statistical law, which results in two different versions of the covering law model: the deductive-nomological model (DN), and the inductive-statistical model (IS) respectively (Hempel 1965). For the sake of illustration, let us focus just on the DN model. An argument is an explanation according to the DN model if the explanans contains a law of nature (i.e. nomological), and the conclusion logically follows from the premises (i.e. deductive). If these requirements are fulfilled and all the premises are true, they would successfully explain the explanandum.

From a formal point of view, an explanation is just an argument, but Hempel (1965) argues that from a pragmatic point of view there is something unique about DN explanations—when seeking an explanation we already know the conclusion of the argument (the phenomenon to be explained) and we look for premises that lead to it. Hempel defends this idea in the thesis of structural identity, which is the view that a DN explanation is a potential prediction, the only difference being the epistemic position we are in: when the known element of the DN argument are the premises (laws and the initial conditions) and we infer a conclusion, it constitutes a prediction. The thesis of structural identity is crucial to making clear what exactly confers upon an explanation its explanatory status: it is in virtue of the predictability from laws that an explanation explains.

Despite its initial success, the DN model of explanation suffers from critical counterexamples that show that the requirement of predictability from laws and the conformity to a deductive argument is neither necessary nor sufficient for an explanation. One decisive difficulty is that it does not account for the asymmetries that intuitively exist in the explanations of some phenomena where causes play a central role. Let us suppose that we want to explain why the length l of the shadow of a given flagpole is 10 meters long. We can construct a DN explanation by citing the height h of the flagpole, the angle α of incidence of the sunrays and using laws of optics about the trajectory of light through air. The information can be expressed in the following equation:

$$l = h / \tan \alpha = 5.77 / (\sqrt{3}/3) = 10$$

This results in an argument, which arguably meets the condition of predictability from laws of nature required by the DN model. Hence, it explains the length of the shadow. However, in this example, adapted from Bromberger (1966), it is obvious that we could equally provide an

argument deriving the height of the flagpole from the length of the shadow, using the same laws of nature and other relevant initial conditions. If we isolate *h* and give the following equation:

$$h = l \bullet \tan \alpha = 10 (\sqrt{3}/3) = 5.77$$

we could use the information to predict the height. By meeting the criterion of predictability from laws we have explained the phenomenon according to the DN model, or have we? There is something odd in saying that the length of the shadow *explains* the height of the flagpole, for the relation between the two variables seems to be explanatorily asymmetrical. Intuitively at least, it is the flagpole that explains the shadow and not vice versa. While the DN model does not discriminate between arguments that can go in both directions, an explanation seems to be satisfactory only in the first case, which suggests it corresponds to an asymmetry between causes and effects. While causes explain their effects, effects do not explain their causes, and if the length of the shadow is an effect of the height of the flagpole, it cannot be the case that an effect explains its cause.

The second main problem for the DN model is that nomic predictability is not sufficient for a satisfactory explanation: one can give an argument that meets the conditions for a DN explanation and yet it fails to explain. Woodward uses the following simple example as an illustration that conforms to a DN argument (2003:187)

> All ravens are black [law]
> a is a raven
> a is black

According to him, the fact that all ravens are black—assuming it is a law of nature—fails to give us answers about the relevant counterfactual questions or as Woodward puts it, what-if-things-had-been-different-questions (w-questions for short), such as how the color of the raven would change (explanandum) if the initial conditions (explanans) were different. As he puts it, a more

satisfying explanation of the color of a raven would need to "locate its explanation within a range of alternatives and would answer a range of w-questions about the conditions under which the explanandum and these alternatives would have ensued. Put differently, such an explanation would identify mechanism that could be used to explain why a nonblack raven (or a raven-like bird) has the color it has" (2003:204). Note that he is not saying that the DN explanation fails to explain because it doesn't cite causes, but rather because this explanation lacks the capacity to support important counterfactual situations or answering w-questions.

2. Causal Explanations

The two counterexamples to the DN model of explanation, the asymmetry problem and the failure to answer w-questions, have something in common. Both of them could be avoided with a causal theory of explanation that requires all explanans to cite causes instead of requiring to produce a valid argument. Thus in the flagpole example, the length of the shadow would explain the height of the flagpole only if the former causes the latter, which intuitively, doesn't seem to be the case. The second counterexample could also be avoided if we note that the generalization used in the argument is not a causal law, so even if it were a genuine law of nature it wouldn't do the job of citing causes. Such strategies depend, of course, on having a proper theory of causation.

Woodward's strategy, however, doesn't explicitly require that a causal explanation cites causes. For him, a causal explanation doesn't simply explain *because* it cites causes, instead, an "explanation is a matter of exhibiting counterfactual dependence" which is capable of showing us how the explananda would change if the initial conditions were different, that is, explanations "locate their explananda within a space of alternative possibilities" (2003:191). In other words,

explanations are able to give us information about counterfactual situations by answering what-if-things-has-been-different-questions.

Now, Woodard's theory of causation, as has been characterized in the previous chapter, sounds very similar to this idea of exhibiting patterns of counterfactual dependence relations. We could say that the information embedded in causal claims that relates variables in a counterfactual way is information that would be perfectly acceptable for a satisfactory explanation, so long as they are able to answer w-questions. But then, is giving a causal relation between C and E the same thing as giving an explanation of E in terms of C? Not necessarily. Woodward is careful enough not to simply equate the citation of causes with giving explanations. If an explanation just is a causal relation, there wouldn't be room for non-causal explanations in that they don't cite causes. Hence, this move opens up the possibility of having a theory of explanation that consist of the capacity to answer w-questions in both causal and non-causal contexts. Let us substantiate the relevance of this possibility in the next section.

3. Non-Causal Explanations

Non-causal explanations play only a peripheral role in Woodward's project. There are, however, non-causal explanations. For example, mathematical and philosophical explanations typically appeal to logical and conceptual relations, e.g. if I say that 'my not having siblings explains my not having a brother or a sister.' While it looks like a poor explanation it does contain some information that could be used to answer at least some w-questions.

Mathematical and logical relations may be rather controversial candidates for providing non-causal explanations, but other interesting varieties of physical explanations have a greater intuitive pull. Some geometrical and equilibrium explanations are of this sort, as the following

example given by Leptin (2007). Suppose we throw a bunch of sticks in the air and we want to explain the fact that at any time there are more sticks falling in a horizontal position than in a vertical one. The explanation appeals to the geometrical fact that there are more possibilities for a stick to be in a horizontal position than in a vertical one, because while there are only two positions on the vertical plane (at 90 or 270 degrees), the horizontal plane allows positions from 0 to 360 degrees. This example may as well admit a causal explanation if we are able to cite the causes of each stick's position, and thus provide a different explanation, but it is not enough reason to discard the geometrical explanation as a genuine explanation. In fact, some of these non-causal explanations are indispensable for some scientific theories because they provide a unique understanding of the phenomena such as in the case of equilibrium explanations of populations (Sober 1983).

Woodward explicitly concedes that at least some explanations do not consist in exploiting causal relations—albeit they do exploit counterfactual dependence relations—especially those cases where there are no causes in the first place. For instance, regarding the stability of planetary orbits, he admits that citing as an explanans the four-dimensionality of space-time "fits well with the idea that explanations provide answers to what-if-things-had-been-different-questions on one natural interpretation: we may think of the derivation as telling us what would happen if space-time were five dimensional" (2003:220). Not only does it seem clear that there are interesting non-causal explanations but also that their explanatory power derives from the ability to answer w-questions just as in causal explanations. The next section explores in greater detail how this is accomplished in Woodward's theory of explanation.

4. The General Account of Explanation and its Advantages

The central question any account of explanation needs to answer is what makes an explanation explanatory, that is, in virtue of what feature is an explanation explanatory. For Woodward the candidate seems to be the ability of an explanation to provide information about counterfactual situations involving the explanans and explanandum. Since this capacity to provide counterfactual information is not something that exclusively belongs to causal relations, Woodward's theory of explanation is a general account of explanation that accommodates both causal and non-causal explanations. Woodward writes: "the common element in many forms of explanation, both causal and non-causal, is that they must answer what-if things-had-been-different questions" (2003:221).

This is the only passage in the book that he explicitly characterizes non-causal explanation and doesn't fully develop such an account. Nonetheless, I think that he is right in pointing out that there is common element in both cases: if we want to admit that there are non-causal explanations (and there seems to be some very interesting non-causal explanations, like argued in section 3) as well as causal explanations, a theory that captures both with a single criterion in an intuitive manner is a preferable than having to deal with two separate theories that tells us different stories about each kind of explanation. Simplicity advises that the same criterion or criteria accounts for all sorts of explanations.

Another advantage is that a general theory would allow, at least in principle, some sort of comparison between causal and non-causal explanations. In the event that scientists come up with two competing explanations for the same phenomenon, a decision could be made given that causal and non-causal explanations are commensurable. Let us take the example mentioned earlier where there are a bunch of sticks falling in the air and we want to explain why at any

given time there are more sticks in a horizontal position. The explanation given by Leptin (2007) cites geometrical facts that have to do with the configuration of three dimensional space as well as probabilistic facts about the distribution of the position of the sticks in space. Following Woodward's theory, we see that in this explanation we are told that if space had not been three-dimensional (or if the probabilistic distribution had not been even) the position of sticks at any given time would be different too. But suppose that we come up with a different explanation for the same explanandum instead. Suppose that at any given time we are able to trace back each of the stick's trajectories and interactions according to gravitational and other laws, such that there is a causal relation for each individual stick that explains its actual position. The conjunction of all these claims seems to be equally used to explain our initial explanandum of why there are more sticks in a horizontal position at any given time.

Obviously, causal and non-causal explanations don't have all their elements in common. There must be something that distinguishes them and that allows us to capture the intuitive differences between say, a geometrical explanation and a causal one, so let us explore these in greater detail.

4.1 How to Distinguish and Hierarchize Causal and Non-Causal Explanations

Once there is room for both causal explanations and non-causal explanations it makes sense to ask how are we to distinguish between them. Woodward does in fact provide a sort of demarcating criterion for distinguishing causal from non-causal explanation. It depends on whether or not an explanation can be interpreted in terms of interventions:

> When a theory tells us how Y would change under interventions on X, we have (or have material for constructing) a *causal* explanation. When a theory or derivation answers a

> what-if-things-had-been-different question but we cannot interpret this as an answer to a question about what would happen under an intervention, we may have noncausal explanation of some sort (2003:221).

Let us call 'the manipulability criterion of explanation' the possibility to conceive or interpret a dependence relation between variables such that the value of one variable is adjustable in virtue of an intervention. The manipulability criterion involves that for a given explanation, if it is possible to interpret or conceive that an intervention changed the explanans (holding other variables fixed) and the value of the explanandum changed accordingly, then the explanation is causal.

Non-causal relations do not meet the criterion, for we cannot even conceive an intervention (i.e. a causally produced change) that would alter the explanans. To use the example Woodward mentions, if the four-dimensionality of space-time is the explanation of the stability of planetary orbits it won't count as a causal one because we are at a loss as to what kind of intervention could we carry out on the dimensions of space-time to switch it from four to five. Hence, it is not a causal explanation.

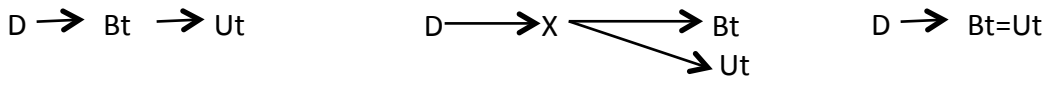
The manipulability criterion is also powerful enough to exclude troublesome non-causal explanations that were at the origin of the asymmetry problem. The derivation of the height of the flagpole in virtue of the length of its shadow and the position of the sun is an excellent illustration of the kinds of explanations that should not count as causal. The impossibility to conceive an intervention exogenous to the system that would only alter the shadow (part of the explanans) while maintaining fixed the position of the sun makes the alleged explanation clearly non-causal.

Possible Problems with the Manipulability Criterion

There are some possible counterexamples to Woodward's manipulability criterion that, I think, could be accommodated, although at a certain cost. Jaegwon Kim (1993) presents a series of examples where a dependence relation is expressed in the form of a counterfactual but which fail to be causal. For example, 'if yesterday had not been Monday, today would not be Tuesday' is a true counterfactual but Monday isn't the cause of Tuesday. Woodward could answer that this is a non-causal instance because it seems impossible to interpret it as there being a possible causal intervention on today's being a different day of the week. That is, although the situation seems conceivable, it cannot be interpreted causally wouldn't be an intervention proper.

Another of Kim's examples is the following: 'if my sister had not given birth at t, I would not have become an uncle at t,' which he takes to be a case where one event determines the other but non-causally. Suppose we use the relation to provide an explanation that my being an uncle is explained by my sister's giving birth, and that intuitively, this should be a non-causal explanation. However, following Woodward's manipulability criterion, it seems that we can interpret it as an answer to what would happen under an intervention. There is an apparently possible intervention on my sister's giving birth at time t: for instance an intervention of administrating a drug that accelerates or delays the delivery for some minutes.

I think there is one way to avoid the counterexample for Woodward. The possibility of there being an intervention of the right sort depends on how we conceive the underlying causal structure that relate the variables in this example. Let *Bt* be the variable of giving birth at time t, and *Ut* becoming an uncle at t. D is a manipulation by administrating a drug:



Suppose we go with the structure on the left, where Woodward's account does mistakenly devise the relation between *Bt* and *Ut* as causal. However, for all we know, both *Bt* and *Ut* could be the effects of a common cause X: for example becoming pregnant. Becoming pregnant, X, could be a cause of both my sister's giving birth at t and my becoming an uncle at t. So the intervention manipulation represented in the central graph above will not count as an intervention (for it violates clause ii) and hence the relation between *Bt* and *Ut* wouldn't be causal, as we expected. Finally, maybe we are just confused about what exactly is for me to become an uncle and for my sister to give birth. Maybe they are just two descriptions of the same event. This is represented on the right, where the same variable has two names, *Bt* and *Ut*. Again, it would be impossible to carry out an intervention on one of them (via D) without altering the other—they are identical. So it wouldn't give us the result that *Bt* causes *Ut*.

Hence, depending what the causal structure is there are ways to avoid the counterexample. However, upshot is that we cannot determine whether the counterfactual conditional gives us a causal explanation unless we know some additional features of those variables and causal relations with other variables, and in this case it is far from clear to determine which structure represents best the actual causal relations. More importantly, we notice once again that in order to determine whether a counterfactual relation is causal or not we need to know additional causal relations. This places some important constraints to our epistemic accessibility of these cases, and together with the epistemic problems mentioned chapter 1, section 6.2. it seems too demanding for a theory to be practically useful.

Hierarchy of Explanations: Explanatory Depth

Since already mentioned, a phenomenon may have different competing explanations both causal and non-causal. It is intuitive to think that while providing at least some relevant answers to w-questions is enough to explain, one explanation could be more powerful or more explanatory than an alternative one. In a paper written by Hitchcock and Woodward (2003), they provide an intuitive answer to this worry with the notion of *explanatory depth*. Usually deep explanations include comprehensive and general laws of physics such as Maxwell's electromagnetic theory to explain the propagation of light with other rather shallow cases like the claim that salt was put in water to explain why it dissolved. This notion of depth is related to the generality of the explanans, that is, the number of cases that theory in the explanans covers. For instance, Newton's gravitational laws are more general than Galileo's law of fall because the former covers more phenomena than the latter. Hence, roughly put, the more general the theory or information used in the explanans is the more explanatory it will be.

However, Woodward and Hitchcock note that generality should not be taken merely as including more information about *other* objects and systems unrelated to the particular explanandum at case, as is usually understood. They argue that generality is typically understood as subsuming a great number of phenomena, like for instance, the explanation of the tides in the sea in virtue of Newton's laws. Newton's laws do not only cover the motion of tides, but also a wide range of other gravitational phenomena of which tides are only one instance. Instead, the relevant generality for explanatory depth is "with respect to other possible properties of the very object or system is the focus of explanation" because it "has to do with the exhibition of patterns of counterfactual dependence describing how the systems whose behavior we wish to explain would change under various conditions" (2003:182). That is, deeper explanations are able to

answer more what-if-things-had-been-different questions about the relation between the explanandum and the explanans. This criterion could be very handy to decide between alternative explanations.

5. Problems with Woodward's Theory of Explanation

In this chapter I argue that Woodward's general theory of explanation suffers from important problems and is unsuccessful as it stands. Since Woodward mainly focuses on causal explanations in his book and given that he doesn't develop the details of non-causal explanations, many of the objections I raise here are not even considered by him. In particular, the objection of the asymmetry problem that I present in 5.2 places an unsolvable dilemma to his theory, which requires him either to abandon a general account of explanation (and hence lose all its advantages) and restrict to a theory of causal explanation, or to deal with a crucial counterexample.

It should be noted that given his causal theory of explanation relies on his theory of causation discussed in the first chapter, it will inherit all the difficulties of the causal theory. However, in the present section I will only focus on new objections targeted at explanations.

5.1 The Problem of Truth Conditions for Non-Causal Counterfactuals

This objection is really the counterpart of section 5.1 in the first chapter where I argued that truth conditions for causal counterfactuals are not fully specified by Woodward. A similar worry arises here for non-causal counterfactuals, and more specifically, for counterfactuals involved in non-causal explanations. If we want correct explanations we need true counterfactuals that associate the explanandum and the explanans, but since Woodward doesn't devote much

attention to this sort of explanation it's unsurprising that he never considers what their truth conditions are or how we are supposed to assess them.

One thing seems obvious: since the distinctive feature of non-causal explanations in his account is that they cannot be interpreted as involving interventions, causal relations won't be part of the truth conditions. A possible solution is to appeal to certain logical, geometrical or semantic relations that hold in counterfactual situations could be used. Interventions would be unhelpful here, for these non-causal counterfactuals don't allow a causal interpretation of a manipulation in the variables. For example, 'if this polygon hadn't had three angles it wouldn't have had three sides" is true due to the geometrical properties of triangles, and the relation between sides and angles. This relation holds for whatever counterfactual situation we conceive about triangles, so it can be used to derive the truth of the counterfactual claim. Woodward doesn't entertain this strategy, and while it may be applicable to some non-causal counterfactuals it is not clear how are we supposed to evaluate other non-causal situations such as 'if space-time hadn't been four dimensional planetary orbits wouldn't have been stable'. This strategy would also require to elucidate what are the truth conditions of these general relations about geometrical properties (and perhaps a story about what properties are for him).

Only at one point in the book he entertains a similar idea, that "there must be a way of disentangling—perhaps merely conceptually or analytically rather than in actuality—the effect on E of changing just C" (2003:131). Woodward is convinced that counterfactuals, both causal and non-causal, must have truth conditions, and that we should be able to assess them, but he repeatedly fails to give a proper way to carry out the most crucial step putting into practice his of explanation.

5.2 The Problem of Asymmetry is Back

In a recent paper, Saatsi and Pexton (2013) rightly note that an immediate tension follows if we interpret Woodward's account as positing a general notion of explanation as the one I argued for in chapter 2. Recall that an explanation explains in virtue of providing answers to w-questions, i.e. to what would happen to the explanans and explanandum under counterfactual situations. Explanations, then, do not depend on the manipulability criterion, that is, they do not need to be interpreted causally in order to explain. So, if causation is not a necessary element for an explanation (although it is for *causal* explanation), it follows that the problems that the DN model of explanation faced are back.

For example, the explanation that a raven is black in the DN model would appeal to the generalization that all ravens are black but Woodward rejects it because it is not a genuine (causal) explanation. But if the only requirement for an explanation is that it answer some w-questions, then this generalization will do the job—poorly, but it will do it. For example, it supports the counterfactual that 'if X hadn't been a raven it wouldn't have been black' and therefore answers at least one, albeit uninteresting, w-question.

What can we say about this objection? Well, perhaps Woodward could make use of the notion of explanatory depth introduced earlier, which allows us to discern good explanations in virtue of the amount of w-questions they can answer about the variables involved. The generalization about ravens will turn out a terrible explanation if we compare it with a causal explanation that relates genetic information of the phenotypic characteristics for instance and which not only tells us about the counterfactual where the individual is not a raven but also what color would this creature be had it had a different genetic makeup (which we could manipulate).

The adoption of the explanatory depth will help overcome problems similar to this one, but won't work for the asymmetry problem. Recall the case of the flagpole, where the problem arises from the intuition that while the height of the pole (plus the position of the sun etc.) can explain the length of the shadow, the opposite doesn't hold. That is, the length of the shadow won't explain the height of the pole. This constituted a problem for the DN model because it produces derivations both from the flagpole to the shadow as well as from the shadow to the flagpole by means of the equation that represents the relation holding between these variables:

$$h = l \bullet \tan \alpha$$

But again, for the comprehensive version of explanation what is required of a relation to be explanatory is that it answers w-questions for counterfactual values of the variables. So if we put forth an explanation of the length of the shadow in terms of the height of the flagpole (and other conditions, e.g. $\alpha=30°$) it will be able to answer several w-questions and counterfactual situations. For example, 'if the length of the shadow had been 7m the height of the pole would have been 4.04m' is a contrary to fact situation answered by the suggested explanation. We cannot conceive a direct intervention on the length of the shadow, and yet, the equation specified above ($h = l \bullet \tan \alpha$) together with other initial conditions perfectly answers many w-questions that an explanation requires and can support related counterfactuals.

Here it won't work to introduce the notion of explanatory depth, because for the two explanations at stake here, namely that the height of the pole explains the length of the shadow and that the length of the shadow explains the height of pole, the amount of w-questions that are able to answer is exactly the same:

if α=30°

	shadow	pole
actual	10	5,77
Counterfactual 1	1	0,58
Counterfactual 2	2	1,15
Counterfactual 3	3	1,73
Counterfactual 4	4	2,31
…	…	…

The aim of comparing the two explanations running in opposite directions is not to decide which one is better about the same phenomenon, since the each of the explanations explains a different phenomenon indeed. The point is that intuitively one of them is a terrible explanation, but by Woodward's standards they are both equally explanatory. A final strategy could be to favor causal explanations over non-causal explanations, such that the asymmetry is captured given that the explanation in one direction is causal and therefore a good one, while the explanation in the opposite direction is a poor one for not being a causal one. However, this would require an extra argument that established the fact that causal explanations are better than non-causal ones by default. Moreover, this move also seems to imply that *any* causal explanation of any phenomenon (however mundane or uninformative) is better than a non-causal explanation that helps scientists understand black holes.

As a consequence, Woodward's theory of explanation faces a crucial dilemma. Either we accept these sort of explanations as proper explanations or we give up the pretension to have a general theory of explanation. If we take the first horn and want to stick to a general notion of explanation that covers both causal and non-causal instances, then we end up with the asymmetry problem, which absurdly suggests that explaining the height of the flagpole in terms of the length of the shadow is not only an explanation but a pretty good one given the criterion of explanatory depth. The second horn of the dilemma requires us to abandon the general account of explanation—and hence lose all its advantages—and restrict to a theory of causal explanation.

CONCLUSION

There are two ways we can assess Woodward's theory of causation. First, as a robust metaphysical project that aims to be a rival theory to current philosophical accounts of causation such as Salmon's mechanical model or Lewis' counterfactual theory. I have argued that as a metaphysical project it is incomplete and has omissions with respect to other theories of causation. More precisely, it lacks the necessary tools to account for truth conditions for counterfactuals, it's circular and it doesn't have a clear notion of possibility that allows us to make sense of counterfactuals. In addition, Woodward's responses to these objections are unsatisfying, mainly because they don't try to solve the problems raised (e.g. when he demands that there must be some way to assess counterfactuals but fails to give explicit details) or he tries to show that these were not genuine problems (e.g. when the problems regarding the notion of possibility are resolved by claiming IN requirements were just regulative ideals).

A second way is to be more permissive about these omissions and assess the value of his theory mainly focusing on its practical aspect as a methodology for science instead. This route, however, also encounters serious problems. First, the extremely high epistemic standards restrict it to cases where we already have a vast knowledge about causal relations, and only then can we assess and test a given causal claim. For other cases, we are left in the dark as to how we're supposed to evaluate a counterfactual statement that may or may not relate causes and effects. The same problem arises when we wonder whether a certain counterfactual dependence relation is causal or non-causal: we need to know with precision the way these elements are related.

Finally, regarding Woodward's theory of explanation, I described a general notion of explanation that includes at least two kinds of explanations, namely causal and non-causal. The theory of causal explanation is promising because it overcomes important problems that the

Nomological Deductive model of explanation can't deal with. However, given that it relies on his theory of causation, it also suffers from the difficulties mentioned in the paragraph above. But there is a more important objection to his causal theory of explanation that derives from the general account of explanation: since explaining consist in answering what-if-things-had-been-different-questions, effects could explain causes, and as a consequence, the counterexamples targeted at the Deductive Nomological model are also applicable to Woodward's causal explanation. Regarding non-causal explanations, a counterfactual theory could be a potentially good theory but Woodward doesn't develop one.

As a result, the initially promising advantages of Woodward's theory of causation and explanation are overshadowed by the many difficulties that remain unsolved.

REFERENCES

Bromberger, S. 1966. "Why Questions," *Mind and Cosmos: Essays in Contemporary Science and Philosophy.* Pittsburgh: University of Pittsburgh Press. 86-111.

Dummett, M.1964. "Bringing About the Past," *Philosophical Review* 73:338-59.

Holland, P. 1986. "Statistics and Causal Inference," *Journal of the American Statistical Association* 81:945-60.

Hoover, K. 1988. *The New Classical Macroeconomics.* Oxford: Basil Blackwell.

Hempel, C. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hempel, C. and P. Oppenheim. 1948 (1965) "Studies in the Logic of Explanation of Science", reprinted in C. Hempel, 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hitchcock, C. and J. Woodward. 2003. "Explanatory Generalizations Part 2: Plumbing Explanatory Depth." *Noûs* 37: 181-199.

Kim, J. 1993. "Causes and Counterfactuals", in Sosa, E. and M. Tooley, *Causation.* Oxford: Oxford University Press

Lewis, D. 1973. "Causation," *Journal of Philosophy,* 70/71: 556-567.

Menzies, P and H. Price. 1993."Causation as a Secondary Order Quality," *British Journal of Philosophy of Science* 44: 187-203.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Sober, E. 1983. "Equilibrium Explanation," *Philosophical Studies* 43 (2): 201-210.

Popper, K. 1963. "Science: Conjecture and Refutations," *Conjecture and Refutations.* London: Routledge and Kegan Paul. 33-39.

Saatsi, J. and M. Pexton. 2013. "Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations." *Philosophy of Science* 80: 613-624.

von Wright G. 1977. *Explanation and Understanding*. Ithaca, NY: Cornell University Press.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press.

Woodward, J. and C. Hitchcock. 2003. "Explanatory Generalizations, Part 1: A Counterfactual Account." *Nous* 37(2):181-199.