



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Mind Association

Making Things Happen: A Theory of Causal Explanation by James Woodward

Review by: Peter Menzies

Mind, Vol. 115, No. 459 (Jul., 2006), pp. 821-826

Published by: [Oxford University Press](#) on behalf of the [Mind Association](#)

Stable URL: <http://www.jstor.org/stable/3840611>

Accessed: 30/11/2014 12:35

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Oxford University Press and Mind Association are collaborating with JSTOR to digitize, preserve and extend access to *Mind*.

<http://www.jstor.org>

Making Things Happen: A Theory of Causal Explanation, by James Woodward. Oxford and New York: Oxford University Press, 2004. Pp. vi + 410. H/b £37.50.

A revolution has taken place in the study of causation over the last decade. Drawing on an older tradition of research in experimental design and econometrics, researchers in computer science, philosophy, and statistics have developed the theory of causal Bayes nets as a framework for the scientific study of causation. Two books stand out as landmarks in this field of research. One is Peter Spirtes, Clark Glymour and Richard Scheines' 1993 book (now in its second edition 2000) *Causation, Prediction and Search*. The other is Judea Pearl's 2000 book *Causality*. These books develop in different ways an extensive and complex theory of causal inference. They show how causal Bayes nets, which are essentially directed acyclic graphs with nodes representing random variables, can play a systematic role in tasks of causal discovery, inference and prediction. Very little of this technical knowledge has, however, percolated through to philosophers until now.

The book under review will help to make this forbiddingly technical literature accessible to philosophers, not because it is a philosophical primer to the theory of causal Bayes nets, but because it decomposes the theory, subjects its underlying concepts to careful scrutiny, and then reassembles them in an original theory of causation, all of this in a way that is especially sensitive to philosophical issues. Whereas Spirtes et al. and Pearl are concerned with practical problems of causal inference from statistical data, Woodward is concerned with the semantic or interpretative task of explaining the meaning of causal claims. For example, Spirtes et al. and Pearl assume a primitive concept of causal relevance or of causal mechanism and use this concept to define other concepts required to estimate quantitative magnitudes and to make causal inferences and predictions. In contrast, Woodward's goal is to try to explain the very concept of causal relevance or causal mechanism taken to be primitive by the theorists of causal Bayes nets by linking these concepts with that of intervention. On his account, a claim that a variable X is causally relevant to another variable Y means roughly that an ideal intervention that changes the value of X will change the value of Y . What is distinctive about causal relationships on this account is that they are potentially exploitable for manipulation and control. Woodward calls his theory a manipulability or interventionist account of causation. One of his goals is to show that this theory is free from many of the problems that have made earlier manipulability or agency theories seem objectionable to philosophers. Thus he argues that when the notion of an intervention is appropriately defined, his account of causation does not attempt to reduce causal concepts to concepts of agency or manipulation; nor is it anthropocentric in linking causation to actual human manipulative capacities.

I think that Woodward is magnificently successful in his goals in this book. The interventionist theories of causation and explanation developed in the book are substantial and important contributions to the literature. In my view,

this is one of a small handful of philosophical books on causation and explanation published in the last twenty years that makes a significant and lasting contribution to the study of these subjects. Because of the clarity and cogency of its argument and the richness of its theoretical framework, the book will inform and guide philosophical thinking about causation, explanation and related subjects for a long time to come.

Chapter one sets out the philosophical orientation of the work and lays down some desiderata for an acceptable theory of causation and causal explanation. Woodward emphasizes that our practices of making causal and explanatory claims have a practical purpose or point, which can be deployed to motivate descriptive theories of what we mean by these claims, but also to justify normative recommendations about what we should mean by them.

Chapter two provides the basic framework for the interventionist approach to causation and explanation. The chapter explains that the fundamental point of making causal claims, as opposed to simple claims about correlations, is to isolate those relationships that are invariant under intervention and hence potentially exploitable for the purposes of manipulation and control. To provide a formal framework for articulating his theory of causation, Woodward introduces the formal apparatus of directed graphs and structural equations of the causal Bayes nets tradition. These are useful structures for representing general and specific causal dependences between variables. Woodward's immediate goal is to spell out exactly what these causal dependences amount to. He distinguishes a number of different causal concepts—direct cause, total cause, and contributing cause—and gives a precise definition of each in terms of what would happen when there is an intervention on one or more variables. For example, X is defined to be a direct cause of Y with respect to some variable set V if and only if there is a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in V besides X and Y are held fixed at some value by interventions. The causal concepts that Woodward is mostly concerned with are type-causal concepts. But some space is devoted to developing an account of what Woodward calls actual causation, or what is more commonly called singular or token-causation. Woodward's account of actual causation is closely related to ones developed by Pearl and his collaborator Joseph Halpern, and by Christopher Hitchcock.

Chapter three is devoted to providing an account of the fine-grained causal structure of an intervention. The rough idea Woodward tries to capture is that an intervention on a variable X with respect to a second variable Y is a process that fixes the value of X in an exogenous way and changes the value of Y only by virtue of changing the value of X . Clearly, this notion of an intervention is causal in character, but this fact does not render the interventionist account of causation vacuous because in general we need not know whether X causes Y in order to determine whether something counts as an intervention on X . Woodward argues that earlier manipulability or agency theories of causation have gone astray in trying to provide reductive analyses of causal concepts in terms

of agency-related concepts. A virtue of his account is that it can make sense of causal relationships holding even in circumstances in which it is nomologically impossible for human beings to intervene and manipulate variables. He attempts to distance himself from the anthropocentrism of earlier manipulability theories by stressing the objective character of the causal relationship: it is built into the very concept of causation that it is a relationship that would hold irrespective of whether it is physically possible for us to carry out any manipulations.

Chapters four and five move the focus of discussion from causation to causal explanation. Chapter four develops a powerful case against the standard deductive-nomological model of explanation: it argues persuasively that the subsumption of explanans and explanandum under a law is neither necessary nor sufficient for causal explanation. Chapter five presents a positive account of causal explanation, according to which causal explanation consists in exhibiting patterns of counterfactual dependence. To explain some phenomenon is to locate it in space of alternative possibilities and to show how it would have been different if different initial conditions had been realized. Woodward's slogan is that causal explanation is successful to the extent that it answers a series of implicit *what-if-things-had-been-different* questions about the explanandum. Such information is explanatory precisely because it is relevant to the manipulation and control of the explanandum.

Chapter six explores the notion of invariance that is central to Woodward's framework. It argues that a necessary and sufficient condition for a generalization to represent a causal or explanatory relationship is that it be invariant under some interventions on variables occurring in the relationship. Generalizations can be invariant to different degrees depending on the number, kind and significance of the interventions under which they are stable. Woodward employs the notion of invariance to reframe familiar philosophical doctrines. For example, he notes that many of the generalizations that play a role in the special sciences do not satisfy most of the standard criteria for being a law. The usual move by philosophers intent on saving the doctrine that successful explanations appeal to laws is to treat these generalizations as second-grade, *ceteris paribus* laws. However, this conceptual doublethink is not necessary, Woodward argues, if we see that the distinctive feature of these generalizations is that they are relatively invariant under interventions on their variables, despite the fact that they are not exceptionless.

Chapter seven applies the ideas about interventions and invariance to illuminate the nature of structural equations, including the regression equations that are used extensively in the biomedical, behavioural and social sciences. It is argued that if a structural equation is to be interpreted causally it must be invariant under some interventions on its right-hand side variables. The chapter also explores the desirability of requiring that explanatory systems of equations be modular in the sense that intervening on one equation does not disrupt the other equations. The significance of imposing this requirement is that it ensures

that each equation in a set of equations represents a distinct causal mechanism that can be replaced or disrupted independently of the others.

Chapter eight compares the proposed account of explanation with the causal mechanical account developed by Wesley Salmon and the unificationist account developed by Philip Kitcher. It argues successfully, to my mind, that the causal-mechanical account cannot, without being completely recast, make the crucial discrimination between those properties that are causally relevant to some outcome and those that are causally irrelevant; and that the unificationist account exploits the wrong sort of generality—a generality of scope rather than invariance—in trying to explain how one explanation can be deeper than another.

As this brief summary indicates, Woodward's book casts new light on many familiar controversies about causation and explanation, but also raises a fair number of interesting new problems and issues, especially in connection with its reformulation of the manipulability approach to causation. I want to consider briefly just one issue concerning this reformulation. Woodward makes much of the subjectivism of early agency and manipulability theories of causation, targeting in particular a theory once defended by Huw Price and myself to the effect that our concept of token causation is grounded in the primitive concept we have as agents of intervening in the world to bring about one event by means of another. I think that many of Woodward's criticisms of this theory strike home. I am persuaded that it is much better to formulate a manipulability account that eschews reduction by using a notion of intervention that bears its causal character on its face; and that is better to avoid anthropocentrism by appealing to a notion of intervention that is general enough to apply even in cases where human manipulation is physically impossible. None the less, it seems to me that Woodward's theory is not as full-bloodedly realist as he suggests.

For instance, consider how Woodward's manipulability theory of causation would answer the Euthyphro question: Does an intervention on X change Y because X causes Y , or does X cause Y because an intervention on X changes Y ? Traditionally, realists choose to answer the Euthyphro question by accepting the first alternative. Woodward writes at times as though this is what he believes. For example: 'When an intervention changes C and in this way changes E , this exploits an independently existing causal link between C and E . One can perfectly well have the link without the physical possibility of an intervention on C ' (p. 132). And elsewhere he writes: '... quite independently of our experience or perspective as agents, there is a certain kind of relationship with intrinsic features that we exploit or make use of when we bring about B by bringing about A ' (p. 125). But I wonder whether he really wants to maintain this. To do so would be to suppose that the causal relation has some intrinsic nature that is only contingently associated with invariant relationships between variables. But this supposition surely runs counter to whole tenor of his interventionist theory, which might meaningfully be expressed as

the view that the whole nature of the causal relation between *X* and *Y* is exhausted by the fact that an actual or hypothetical experiment on *X* would change *Y*. It seems to me that the manipulability theorist of causation should answer the Euthyphro question by embracing the second alternative: *X* causes *Y* because an intervention on *X* would change *Y*. But embracing this is to take one step back from a full-blooded realism about causation.

Another respect in which Woodward overstates his theory's realist credentials, or rather understates its anti-realist implications, is by downplaying the representation-dependent features of the causal concept. It turns out that his definitions of important causal concepts such as those of a direct cause, a total cause, and a contributing cause are relative to a variable set: whether a variable is a direct, total or contributing cause of another is sensitive to which variables are used to represent these causal relationships. (Indeed, it turns out that these causal relations also depend on the range of allowable values the variables can take (p. 56).) As Woodward remarks, this fact is a surprising but inevitable consequence of using systems of structural equations and directed graphs to represent causal relationships. Within this framework, the choice of variables is crucial for the representation of causal relationships: the variables chosen reflect the possibilities one is willing to take seriously. Indeed Woodward exploits this feature of the framework for representing causal relations to explain a feature of the concept of actual or token causation that is puzzling from the perspective of his theory (pp. 86–91). Our application of the token causal concept tends to be highly selective. For example, we judge that the omission of a supervising doctor to administer a medicine to a patient caused the patient's subsequent illness, but we do not judge that a similar omission by a doctor who had no responsibility for the care of the patient caused the patient's illness, even though it is true in both cases that if the doctor had administered the medicine, the patient would not have suffered the illness. Woodward explains the difference in our causal judgements about these examples in terms of our willingness to take seriously certain possibilities. It was a serious possibility that the first doctor should have administered the medicine, but not a serious possibility that the second doctor should have done so. These judgements would be reflected in our representation of the causal structure of the situation by the inclusion of a variable representing the first doctor's action, but not a variable for the second doctor's action.

All of this seems plausible enough as an explanation of the selectivity of our actual or token causal judgements. None the less, the crucial role played by the choice of variables in the representation of causal relations raises some serious questions about the extent to which a full-blooded realism about causation can be sustained. The fact that the correct application of causal concepts depends on certain key representational choices seems to imply that the truth-conditions of causal concepts are not completely objective or mind-independent.

It is a pity that Woodward does not explore these matters at greater length since they impinge on foundational questions about his chosen framework for

explaining causal concepts. Other researchers in the causal modelling tradition are more up-front about the representation-dependent character of causal concepts in asserting that causal claims must always be understood relative to a model. On this surface this talk is puzzling from a philosophical point of view. What is a model? Is it an abstract mathematical object or a mental construction? How is causation relative to a model? Does this just mean that the truth-conditions of causal concepts are relativized to models in the same way that context-sensitive expressions are relativized to contextual parameters? Or does it have some more radically subjectivist meaning? Such questions that are raised by the recent philosophical borrowings from the modelling tradition deserve sustained philosophical discussion. Woodward does not address these questions partly because he is eager to downplay any departure from realism on the part of his theory. This is a pity because I am sure that Woodward would have given an extremely acute philosophical explanation of this talk.

In conclusion, these minor quibbles aside, this is a splendid book. It is beautifully and clearly written; and in countless ways sheds a flood of light on a range of topics to do with causation and explanation. It represents the most significant and substantial philosophical contribution to the study of these concepts in recent years. Reading and studying this book will be obligatory for everyone whose work bears directly or indirectly on the topics of causation and explanation.

Department of Philosophy
Macquarie University
North Ryde, NSW 2109
Australia
pmenzies@scmp.mq.edu.au
doi:10.1093/mind/fzl821

PETER MENZIES