Comments on Woodward's article

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

As much as I appreciate Woodward's drive to attribute deeper meaning to the notion of "intervention," I cannot convince myself that this simple concept should require complex definitions such as WIN and SIN. As a devout interventionist, I have found inner peace and unchallenged comfort and clarity in the standard "equation wipe-out" conception, where "intervening on X" means removing the equation that determines X and replacing it with some other equation (often a constant). In *Causality* (2000), I have shown that this ideal conception, dubbed "atomic," can serve as a basis for a wide variety of causal concepts and tasks, including the analysis of complex types of interventions, (for example, those that change several mechanisms at once, or those that merely modify, not wipe-out the mechanism controlling X) and including interpretational projects such as the one discussed by Woodward.

In Section 2, Woodward argues convincingly that the notion of "hypothetical idealized experimental manipulation" is needed as a theoretical tool for understanding the meaning of structural equations, as well as the meaning of causal claims in general. Indeed, the meaning of structural equations enjoys a simple characterization in the atomic framework (*Causality*, (2000, pp. 137)). For example, the meaning of Eq. (2.3) in Woodward's example reads: If we were to intervene and set X_2 to x_2 and X_3 to x_3, then the value of X_4 would be $cx_2 + dx_3$, regardless of the value we set for X_1.[1] Likewise, the meaning of the presence (or absence) of direct (or indirect) causal relationships between two variables is simply and rigorously characterized in the atomic framework. For example, the absence of a direct causal relationship between X_1 and X_4 in Woodward's example (Eq. (2.3)) means: If we were to intervene and hold X_2 and X_3 fixed, than X_4 would be independent of the value at which we may set X_1.

[1] Formally, this reads: $Y(set(x_1, x_2, x_3)) = cx_2 + dx_3$, where $Y(set(x))$ stands for the counterfactual "Y, if X were x." Note that this interpretation is stronger than the one advanced by Woodward, for it insists explicitly on X_4's invariance to interventions on X_1 and, in general, to any interventions on variables not mentioned in the given equation.

In Section 3, on the other hand, Woodward seems to abandon the *idealized* and universal character of our target concept - intervention - and advocates instead a strategy whereby the basic characterization of intervention should vary with the purpose of one's analysis.

Oddly, Woodward dismisses the atomic framework as good for prediction (or "calculation") only, and argues that a whole new conception of intervention is needed for dealing with a project he calls "interpretive": "to provide an account of the meaning or content of causal claims in terms of claims about the response of certain variable to intervention on others."

I will argue that it is counterproductive to invent new conceptions of intervention with each new usage. Not only is the atomic framework sufficient for handling interpretive projects in general, but it is also necessary for clarifying the peculiar interpretive project addressed by Woodward, as well as for substantiating the solution that he conjectures in SIN and WIN.

First, note that the distinction between the calculational and interpretive projects is merely one of quantification – the former subsumes the latter. This means that any conception adequate for calculating the magnitude of a causal relationship is also adequate for ascertaining whether such relationship exists.

Second, Woodward's strategy of re-conceptualizing interventions with each new project stands contrary to common scientific practice. Normally, we define the meaning of scientific terms (e.g., basic physical quantities like mass or velocity) using ideal experiments and we expect them to serve a wide variety of projects and circumstances, including the analysis of non-ideal experiments as well as projects aimed at new purposes. Woodward himself recognizes the wisdom of this practice, stating (Section 6): "On my view, the content of causal claims often may be clarified by invoking what would happen under hypothetical interventions even if those interventions cannot or will not be carried out." Thus, I fail to see why Woodward does not take the "atomic intervention" as a primitive and, when faced with problems involving non-ideal manipulation,[2] derive (not merely conjecture) conditions such as S/WIN that would be sufficient for answering the questions he poses, say whether X has a causal influence on Y. In other words, conditions such as those conjectures in S/WIN should be construed, not as a new characterization of intervention but, rather, as experimental conditions that would render certain type of inferences valid within the atomic characterization of intervention.

In *Causality* (2000) I demonstrate the ease with which this approach can be executed. For example, Section 3.4.4 (p. 88) gives conditions under which the causal effect of X on Y can be deduced from an experiment in which another variable, Z, is randomized instead of X. These conditions include SIN as a very special case. To tackle the more general experiment where a manipulated variable, say I, is judged to modify several mechanisms, we simply add node I

[2] In the salt-chamber example, building a new pipe from I to X_2 is not ideal (in the theoretical sense) since it does not block the flow from X_1 to X_2 as required by the standard "equation wipe-out" definition.

to the graph and draw arrows from I to all variables that are determined by the modified mechanisms. Subsequently, if our aim is find the causal effect of X on Y, we ask whether the target quantity $P(y|set(x))$ can be deduced from $P(y, x, z, w, \ldots|set(I))$, which is the joint distribution obtained under manipulations of I.[3] With this aim in mind, the conditions stated in SIN and WIN are much too restrictive; I_3, I_4 and I_5 are not necessary for inferring $P(y|set(x))$ from $P(y, x, z, w, \ldots|set(I))$.

Woodward interpretive project, which underlies the conditions in WIN and SIN, aims in fact at a much more ambitious goal than just finding whether X causally influences Y. What Woodward must have had in mind (and left implicit in motivating SIN) are conditions under which one is at liberty to assume that, if intervention I produces changes DY and DX in X and Y, respectively, then an atomic intervention on X, of magnitude DX, would produce a change DY in Y. This assumption, formally written

$$P(y|set(x)) = P(y|x, set(I))$$

is convenient, for it requires no measurement of any other variable except X, Y and I. Sufficient conditions for this formula to hold can, again, be proved correct using the set(x) calculus, as shown in *Causality* (pp. 89); they need not be posited without proof, as in SIN.

It should also be noted that Woodward is mistaken in asserting (Section 10) that "W/SIN does not impose the reservation requirement," namely, that W/SIN "says nothing about whether I alters the connection between X and Y." Condition I_2, which rules out a direct link from I to Y amounts precisely to saying that I does not alter the connection between X and Y.

Likewise, it is not true that the "atomic" conception of intervention requires that "we already have some information about the causal relationship, if any, between X and Y" (section 10). Knowing whether an intervention (say I) affects the causal relationship between X and Y does not require any prior information about that relationship. It requires merely knowledge of whether I may change Y if we hold X fixed, that is, whether an arrow should be drawn between I and Y. This knowledge is invoked in WIN and SIN as well (through condition I_2). Thus, Woodward's concerns with a "vicious circularity" infecting the atomic framework are unwarranted. More generally, there are no real differences between the interpretive problem addressed by Woodward and the standard problems addressed in the literature, dubbed "calculational."

In a series of papers (see [Cartwright, 2002]), James Woodward, Daniel Hausman and Nancy Cartwright have expressed difficulties defining, communicating, and agreeing on basic notions of intervention, modularity, invariance and causal correctness. A striking example is the confusion surrounding equations (9.1)–(9.5). Here, I agree with Woodward that (9.5) is NOT causally

[3] Formal machinery for deducing such quantities is provided in chapter 3 of *Causality* (2000)

correct, and that Cartwright's account of causal correctness fails to represent interventions on endogenous variables.

In my opinion, these confusions and difficulties stem primarily from a reluctance to communicate in a formal language, such as the one offered by the atomic framework. In my experience, I have found that, invariably, questions about interventions and experimentation, ideal as well as non-ideal, practical as well as epistemological, calculational as well as interpretive, can be formulated precisely and managed systematically using the atomic intervention as a primitive notion. The same applies to questions about "correctness" of causal models.[4] I will thus end this commentary with a conjecture (or a challenge) that any philosophical question, disagreement, or difficulty concerning causal objects or causal relationships can be resolved if expressed formally in the language of atomic interventions and reduced to a mathematical exercise in the calculus of $P(y|set(x))$.

References

[Cartwright, 2002] N. Cartwright. Against modularity, the causal Markov condition, and any link between the two: Comments on Housman and Woodward. *Brit. J. Phil. Sci.*, 53:411–453, 2002.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York, 2000.

[Tian and Pearl, 2002] J. Tian and J. Pearl. A new characterization of the experimental implications of causal Bayesian networks. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 574–579. AAAI Press/The MIT Press, Menlo Park, CA, 2002.

[4]Causality (2000, p. 24) gives three properties that characterize precisely what is meant for a Markovian causal model to be "correct." Extensions to semi-Markovian models are given in Tian and Pearl (2002).