Economics and Philosophy, **19** (2003) 341–344
DOI: 10.1017/S0266267103001196

REPLY TO WOODWARD

JUDEA PEARL
University of California, Los Angeles

I thank Dr. Woodward for his illuminating review of my book *Causality*, for explicating so clearly the basic contributions of the book, and for giving me the opportunity to further clarify some aspects of the *do*-calculus, specifically those that pertain to the notion of intervention.

Woodward's concerns regarding the notion of atomic intervention (dubbed **PI**) fall into two main categories. First, we rarely find such delicate interventions in experimental practice. Second, quoting Woodward, the notion 'provides no basis for (or at least misidentifies the real basis for) distinguishing between those experimental manipulations that . . . are defective from the point of view of assessing the causal relationships between D and R and those that are acceptable' (332). I will show that the first concern is not a restriction on the use of the $P(y \mid do(x))$ formalism, while the second is mistaken; the **PI** formalism provides in fact the formal basis for making distinctions between acceptable and defective manipulations meaningful and precise.

Basic to any discussion of interventions and manipulations is the understanding that (*i*) a manipulation may potentially modify several causal mechanisms in the world (or in our theory about the world) and (*ii*) before we can say anything meaningful about any specific manipulation we must make assumptions about which mechanisms are potentially modifiable by the manipulation in question and which remain intact. Woodward seems to share this understanding, for assumption (*ii*) is used implicitly throughout his discussion and shines most clearly in his definition of **EI**.[1]

[1] In section 3, Woodward argues for a new formulation of intervention in order to manage cases where one has limited information about the true causal graph. Paradoxically, **EI** makes repeated references to the paths in that causal graph. Recall, an arrow $X \rightarrow Y$ in a causal graph signifies merely suspicion, not certainty, in the existence of causal relationship between X and Y.

Given this assumption, let us characterize a given manipulation I by the set $W(I) = \{W_1, \ldots, W_k\}$ of mechanisms that I is suspected of modifying. Recalling further (*Causality*, p. 226) that, in a causal model, there is a one-to-one correspondence between mechanisms and variables (i.e., the value of each variable is determined by one and only one mechanism), we might as well characterize I by a set $Z(I) = \{Z_1, \ldots, Z_k\}$ of variables, where each Z_i stands for the variable that is determined by mechanism W_i. Given this characterization, one can represent manipulation I graphically by adding to the causal graph G a new node, labeled I, and drawing arrows from I to each variable in $Z(I)$. Furthermore, the effect of intervention I on any variable Y would be characterized by the expression $P(Y = y \mid do(I = i))$.[2]

We see that any manipulation whatsoever, including of course those used in Woodward's examples, can be encoded conveniently using the $do(x)$ notation, coupled with an augmented graph $G(I)$ in which the manipulation itself is represented by a distinct variable I. Moreover, all questions related to the impact of I can be handled using the intervention calculus described in *Causality* (ch. 3), using $G(I)$ as guide. Thus, Woodward is right in noting that most practical manipulations are non-atomic, in that they carry side effects and may modify several mechanisms at once. However, this fact does not restrict the usefulness of the $do(x)$ formalism – the formalism permits us to characterize those compound manipulations in terms of their atomic components and submit them to formal analysis using the $do(x)$ calculus. This decomposition can be likened to the prevailing practice in chemistry, where chemical compounds are characterized in terms of their constituent chemical elements, though pure chemical elements are rarely found in nature.

In *Causality*, I demonstrate the ease with which this approach can be executed. For example, section 3.4.4 (p. 88) gives sufficient 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 **EI** as a very special case. The randomized variable Z plays of course the role of the variable I in the augmented graph $G(I)$, and may modify several mechanisms. Thus, Woodward's problem of distinguishing acceptable from defective manipulations reduces to a mathematical exercise in the $do(x)$ calculus. If our aim is to assess quantitatively the causal effect of X on Y, we ask whether the target quantity $P(y \mid do(x))$ can be derived from $P(y, x, z, w, \ldots \mid do(I))$ – the joint distribution obtained under manipulations of I.[3] More modestly, if our

[2] This formulation applies as well to manipulations that merely change the nature of the mechanisms involved, without dictating in advance the values of the corresponding variables.

[3] Formal machinery for deriving such quantities is provided in ch. 3 of *Causality*.

aim is merely to verify qualitatively whether X has causal influence on Y, an aim that seems to be at the center of Woodward's concerns, we ask whether the truth value of the proposition '$P(y \mid do(x)) = P(y)$' can be deduced from

$$P(y, x, z, w, \ldots \mid do(I)).$$

Applying this analysis to Woodward's example *EX.* 1, we first note that, contrary to the conclusion of **EI**, manipulation I is not defective at all; the model permits us to evaluate the causal effect of D on R despite the side effect represented by the arrow from A to R. This becomes clear by applying the back-door criterion (*Causality*, p. 79) which legitimizes the standard adjustment for confounder A, and gives $P(R \mid do(D)) = \sum_A P(R \mid A, D)P(A)$. In fact, we can derive the causal effect of D on R without resorting to any experiment; it is derivable from passive observations alone.[4]

The distinction between *EX.* 1 and *EX.* 2 is not that the former is 'defective from the point of view of assessing the causal relationship between D and R', but rather, that the latter permits the assessment of this relationship through a simpler formula, $P(R \mid do(D)) = (R \mid D, do(I))$, which requires no measurement of A (see *Causality*, p. 88).

In summary, we see that the notion of atomic intervention, $P(y|do(x))$, provides the formal basis for defining what it means to 'assess the causal relationships between X and Y' and for explicating in what sense Woodward's criterion, **EI**, is in fact correct.[5]

Woodward asserts that, in the **PI** formalism, in order to know whether a manipulation I qualifies as atomic intervention on X we must ensure that I does not modify the causal relationship between X and Y and, hence, we must know already whether there is a causal relationship between X and Y, i.e., whether there is a directed path from X to Y. This assertion is inaccurate. Ensuring that I does not modify the causal relationship between X and Y does not require any prior information about that relationship. It requires merely the assumption that I does not change Y if we hold all other variables fixed (i.e., that no arrow should be drawn between I and Y), an assumption invoked in **EI** as well. This leads to a simple and general criterion of what it means to draw an arrow from one variable to another in the causal graph: such an arrow is drawn when we find a manipulation I that is not linked directly to Y and is capable of producing changes in Y when we hold fixed all parents of Y except X.

To conclude, I have found that, invariably, questions about interventions and experimentation, ideal as well as practical, interpretive

[4] Technically, randomizing I may increase estimation power, but is not needed for consistency.

[5] Note that Woodward posits the condition of **EI** without proof; how do we know that conditions (*i*)–(*iv*) ensure that I is not defective in the sense intended by Woodward?

as well as epistemological, can be formulated precisely and managed systematically using the atomic intervention as a primitive notion. I will thus end this commentary with a conjecture (or a challenge) that any intervention-related problem that one can articulate unambiguously can be expressed formally in the language of atomic interventions and reduced to a mathematical exercise in the calculus of $P(y \mid do(x))$.

REFERENCE

Pearl, J. 2000. *Causality: Models, reasoning, and inference.* Cambridge University Press