COGNITIVE SCIENCE **15,** 293–300 (1991)

COMMENT

# Nonmonotonic Reasoning and Causation: Comment

HERBERT A. SIMON
*Carnegie-Mellon University*

In a recent issue of *Cognitive Science,** Yoav Shoham (1990) proposed a theory of causation based on a nonmonotonic logic of temporal knowledge. It is the purpose of this commentary to show that Shoham's procedure is essentially isomorphic with a theory of causal ordering that requires no special nonmonotonic logic. Although this latter causal ordering theory is compatible with the requirement that causes precede their effects, it can also be used to describe systems in which causes and effects are simultaneous. This discussion will employ a version of the causal ordering theory that uses Boolean truth functions of sentences (Simon, 1952; reprinted in Simon, 1977, chap. 2.2) rather than the more famliar version that uses mathematical functions of real variables (Simon, 1953; reprinted in Simon, 1977, chap. 2.1) as the material of construction. This will make the relation to Shoham's theory easier to see.

The discussion here will be largely informal; the formalities are easily supplied by reference to the original articles by Shoham (1990) and Simon (1952, 1953).

## Causal Ordering

We can illustrate the causal ordering theory with the same example Shoham employed throughout his article. In words:

> Turning the ignition key causes the auto motor to start, provided that the battery is not dead, the spark plugs are not defective, etc. (p. 217)

We can restate this causal theory symbolically, by designating the turning of the key by K, the operability of battery and spark plugs by B and P, respectively, and the starting of the motor by M. The right arrow, $\rightarrow$, is to be read "causes."

---

Correspondence and requests for reprints should be sent to Herbert A. Simon, Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213.

* Shoham, Y. (1990). Nonmonotonic reasoning and causation. *Cognitive Science 14,* 213–252.

$$K \wedge B \wedge P \multimap M$$

In this symbolic translation, we have lost the distinction between "causes" and "provided that" in the original verbal statement. However, although it seems a bit odd to say that "the battery's being charged causes the motor to start," it is quite natural to say that "the battery's being dead causes the motor not to start." In fact, diagnosticians and troubleshooters make statements like this all the time; their major preoccupation being to determine the causes of inoperability. Hence, the causal relations between ignition key and starting on the one hand, and operability of battery or spark plugs and starting on the other, seem to be quite similar to one another, if of opposite "sign." Therefore, I will retain the symmetry.

The next thing to notice about the causal ordering scheme is that the causal arrow, unlike an implication sign, does not reverse direction under negation. That is, from the previous relation, one cannot infer:

$$\neg M \multimap \neg K \vee \neg B \vee \neg P$$

The motor's not starting does not cause the key not to be turned, or the battery to be dead, or the spark plugs to be defective. More generally, the causal relation is not truth functional. What, then, is it, and how is it derived?

We begin with the primitive notion of "mechanism." A mechanism is something that makes the truth value of one sentence depend on the truth values of one or more other sentences, but without the directionality of the dependency being specified. Thus, in the above example, there is postulated a mechanism that makes the truth value of M depend on the truth values of K, B, and P, but without specifying any direction of the causal arrow. The reason for postulating this dependency, of course, is empirical: Experience has shown that such a physical dependency actually exists, and the causal sentence (*sans* its implication of direction) denotes the postulated mechanism.

The reason for introducing mechanisms independently of directionality is that many real mechanisms are bidirectional. For example, a bridge truss "transmits" forces from one node to another in both directions. We can only determine the causal arrow from information about other mechanisms with which these nodes are connected.

Hence, to infer direction, and thereby arrive at the full causal statement, additional information must be provided. This information may (and often does) take the form of specifying that the truth values of certain sentences are given exogenously, independently of the postulated mechanism. In the case of starting the auto, we would normally assume that the turning of the key and the states of operability of the battery and the spark plugs are determined independently of the mechanism that connects them with the motor. Our sum of information, consisting of the mechanism plus the assumptions of exogeneity, can now be depicted in the following matrix:

| | K | B | P | M |
|---|---|---|---|---|
| K | X | | | |
| B | | X | | |
| P | | | X | |
| M | X | X | X | X |

The four columns correspond to the four sentences, "the key has been turned," and so on. The four rows represent the mechanisms that determine the truth values of the sentences. Each row is a truth function of a subset of the sentences, specifically, the subset whose columns contain the symbol X. When only a single *X* appears in a row, this means that the truth value of the corresponding sentence is determined independently of the truth values of other sentences, that is, is exogenously determined.

In the matrix above, the sentences K, B, and P are exogenous. When their truth values have been set, the truth value of M is determined (*caused*) by the fourth mechanism. In such a situation, we say that M is caused by K, B, and P, or that turning the key causes the engine to start, provided that the battery and spark plugs are OK. It is equally true to say that a dead battery ( or defective plugs) causes the engine *not* to start, even if the key is turned.

We can now also construct the truth table for the system, which contains these and other sentences that can be phrased as causal statements (always with the truth value of M as effect), and which sums up everything the diagnostician or troubleshooter needs to know and can know about the system at this level of detail.

| K | P | B | M |
|---|---|---|---|
| T | T | T | T |
| T | T | F | F |
| T | F | T | F |
| T | F | F | F |
| F | T | T | F |
| F | T | F | F |
| F | F | T | F |
| F | F | F | F |

The key to the asymmetry of the causal ordering in this system of mechanisms lies in the fact that the truth values of various subsets of sentences can be determined in a particular order. K, B, and P being endogenous permits us to determine their truth values independently of M and of each other. We will call them sentences of zero order. Having determined their truth values, we can now determine the truth value of M. We will call it a sentence of first order. We will assert a causal relation between a given sentence and any sentences of higher order whose truth values depend on the truth value of the given sentence. In the example, K, B, and P all have a causal relation with

M. (The fact that the truth of M is associated with the truth of K, whereas the falsity of M is associated with the falsity of B or P is irrelevant.)

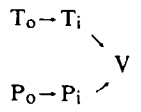
The assumptions we made in order to establish the causal relations in the system were empirical assumptions, presumably to be tested by examining the real world: the existence of specific mechanisms in which particular sentences were implicated, and specifically the exogenous determination of the truth of certain sentences. The formal definition upon which the example is based can be found in Simon (1977, pp. 82–84).

Notice that in this formulation, it is not assumed that the cause precedes the effect in time, although it may very well do so. We may wish to impose at least the restriction that the effect *not* precede the cause, but the formalism does not require this restriction. On the other hand, we may wish to use temporal precedence as one of the cues for determining that a sentence is exogenous.

There are at least two positive reasons for not incorporating temporal precedence directly in the definition of causal ordering. First, we are thereby able to make causal statements about systems in a state of equilibrium, or systems shifting so rapidly from one state of equilibrium to another that we cannot detect the time lag. For example, consider a vertical cylindrical chamber with a piston held in place by a weight resting on it. The chamber is filled with gas, and its walls are highly permeable to heat so that the temperature of the gas within the cylinder cannot deport measurably from the temperature on the outside.

Now the volume of the gas in the cylinder is determined (caused) by the mass of the gas it contains, the temperature of the air outside (and hence inside), and the pressure exerted by the weight. If we change the weight, the pressure will change, consequently the volume also. If we change the outside temperature, the inside temperature will change, and consequently the volume also. $T_o$ and $P_o$, the external temperature and pressure, are exogenous. They cause, in turn, the internal temperature, $T_i$ and pressure, $P_i$, respectively. The latter variables, in turn, cause V.

The reader can easily construct the appropriate matrix of five variables and five mechanisms: two for the exogenous variables, two for transmitting the external influences into the cylinder, and the fifth connecting the internal pressure, temperature, and volume. The first two mechanisms will identify the two variables of zero order, the third and fourth mechanisms, two variables of first order, and the fifth mechanism a variable of second order, yielding the following causal structure:



A second reason for not postulating temporal precedence is that in making causal statements about experimental situations or, more generally, situations in which there is external intervention, the exogenous sentences can be determined directly from knowing which variables can be manipulated directly. Thus, in the piston example, we can change the external temperature and the weight on the piston, but we have no direct manipulation for changing the internal temperature, pressure, or volume. (If we did, then we would be confronted with a different set of mechanisms, and potentially, a different causal ordering.)

On pages 217–220 of his article, Shoham (1990) listed eight "properties associated with the concept of causation." Causal ordering possesses most, but not all, of these properties:

1. *"The causal connective is not a material implication."* Agreed. Causal ordering is both weaker and stronger than a material implication.
2. *"The causal connective is nonmonotonic."* Partial agreement. Causal ordering is "nonmonotonic" in the sense that if $A$ causes $C$, this does not imply that $A$ and $B$ cause $C$. However, causal ordering makes use of the standard sentential calculus and not a special modal logic. It is nonmonotonic because it is not a truth function.
3. *"Causal statements are context sensitive."* Agreed. The context dependency of causal ordering in the example would be achieved by redefining the mechanism when the battery is disconnected. Essentially, this is what Shoham also did: He changed his axiom schema to rerepresent the causal relations.
4. *"Causal ordering is antisymmetric and antireflexive."* Agreed.
5. *"Causes cannot succeed their effects in time."* No agreement. The reasons why causal ordering does not require temporal ordering have been discussed.
6. *"Entities participating in the causal relation have a temporal aspect."* Agreed. The sentences in the causal ordering structure describe events.
7. *"Causal terminology includes other verbs besides 'cause'."* Partial agreement. In causal ordering within the sentential calculus, "enable" and "prevent" are easily distinguished, and can be handled exactly as Shoham suggested. In the case of causal ordering within a calculus of functional relations, this language may not be applicable. For example, in the sentential calculus, we may say: "Rain causes the wheat to grow," or "in the presence of adequate rain, warm weather enables the wheat to grow," or "Absence of warmth will prevent the wheat from growing." If we replace such statements by a function in which the amount of wheat is jointly determined by the amount of rain, the temperature, and so on, then terms like "enabling," "preventing," and so forth, are not readily applicable.

8. *"Causal terminology is amazingly ubiquitous in everyday life."* Agreed. It is ubiquitous in the everyday conversation of physicists, too. The reason Russell thought it superfluous is that he had in mind systems of simultaneous total and partial differential equations, in which, essentially, "everything causes everything else," hence, causal statements do not buy very much. It does not help much to say that the positions of all of the members of the solar system at time $(t+1)$ are caused by their positions at time $t$. Causal ordering becomes interesting (as in trouble-shooting) when only a few things are connected directly (by mechanisms) to each other thing.

**Shoham's Causal Logic**

The causal ordering formalism can now easily be compared with Shoham's formalism by observing how the latter handles the car-starting example. First, Shoham did not deal with the truth of sentences, but with *knowledge* of whether they are true or false. This immediately introduces an asymmetry between K, on the one hand, and B and P, on the other. In his axiomatization, he postulated that it is known that K is operative, but not known that B and P are inoperative.

It is implied, although never clearly stated, that not all the inoperative things need be mentioned explicitly (note the "and so on" toward the bottom of p. 231, and the "..." in the restatement of the axioms on p. 232). Similarly, Old Mother Hubbard declared her cupboard to be bare without enumerating all the groceries not in it. In causal ordering, all variables not included in the model are assumed to be irrelevant to the system's behavior within the limits of interest; events whose failure would be relevant are included explicitly (as in Shoham's axioms).

Second, Shoham required effects to occur at later time intervals than their causes.

Third, Shoham's causal formulas depend on the truth values of the sentences they contain, whereas the formulas in the causal ordering system depend only on the presence or absence of particular sentences in particular mechanisms.

If we look at the details of the formalisms, there are other differences, but these are the most fundamental, and the only ones that seem to have consequences of much substance. It is, in fact, quite easy, in any given illustrative situation, to translate back and forth between the two formalisms.

There are at least three disadvantages in introducing the modal operators "it is known that S" and "it is not known that not S" into the formalism. First, we must then define a modal logic within which to operate, whereas the causal ordering scheme operates within the standard sentential calculus. The (intended) compensation for this inconvenience is that we can assign causal relations without knowing the truth values of sentences. But the

causal ordering scheme also achieves this goal, for it depends only upon which sentences appear in the descriptions of which mechanisms. Hence, this desideratum provides no reason for abandoning the sentential calculus for more esoteric logics.

The second disadvantage in using the modal logic is the asymmetry already noted. There appears to be no good reason why we should define a completely separate causal ordering to define the causal dependence of motor failure on the failure of battery or plugs. There is a single system of mechanisms here, in which the key, the battery, the plugs, and the motor are all components; and it seems desirable to represent it as a single and symmetric system. Either we simply assume that all other components are always operative, in which case we treat the key, an exogenous variable, as the sole cause, and the motor's starting as the effect; or we consider the possibility of elements malfunctioning, in which case we incorporate all these elements in the causal scheme.

A third disadvantage in using modal logic is that the presence of the implication sign suggests that the causal statement could be counterposed, making the nonstarting of the motor the cause of the nonturning of the key. Shoham avoided this by the temporal constraint and the use of the modal operators, but with the causal ordering scheme, the problem never arose because the causal relation is not truth-functional.

**Shoham on Causal Ordering**

Shoham's article contains a brief discussion of the causal ordering scheme, relating it to work in qualitative physics, the context in which he evidently encountered it. Perhaps it is worth mentioning that Simon (1952) was published in the *Journal of Philosophy,* hence, perhaps can be qualified as a "philosophical" theory of causality. As such, it possesses the same advantages over the other philosophical theories Shoham discussed as his own theory does, with qualifications previously noted.

Shoham counted it as a major disadvantage of the causal ordering scheme that it requires an *a priori* specification of the exogenous variables. Indeed it does, as we have seen, but his modal logic formalism does not escape this requirement. It is implicit in the decision as to which variables are to be placed to the left, and which to the right, of the causal arrow. Both formalisms would appear to require the same empirical information about the systems to be formalized. The counterexample Shoham presented on page 246, in which acceleration causes force, is as readily constructible in his system if acceleration, instead of force, is put to the left of the implication sign.

Note also that the $F = MA$ example involves no precedence in time. I think it is commonly said that force causes acceleration (i.e., that the force is exogenous in this equation), but if I read the Newtonian equations correctly, the cause seems to act instantaneously. Of course, if we accept gravity

waves, a time lag can appear, but do we want to rule out instantaneous effects *a priori,* independently of the empirical evidence?

I have already set forth other reasons for regarding the absence of a requirement of temporal precedence as an advantage of causal ordering over formalisms that will only work when the cause is known to precede the effect.

Finally, although it was shown in Simon (1952, 1953) how the causal ordering scheme applies both to the sentential calculus and to systems of functional relations among real and complex variables, it is not equally obvious how the scheme using modal logic would be extended to the latter class of systems. Here, the distinction between ordering systems of variables and ordering *values* of those variables becomes truly critical. On the other hand, I have no proof that a solution to this problem could not be found.

In conclusion, it appears that any causal relation expressible in Shoham's formalism is as readily expressible in terms of causal ordering, and perhaps vice versa. The relative simplicity of the causal ordering formalism, and its dependence only upon the sentential calculus, or (in the functional case) upon standard mathematics, would seem to give it rather decisive advantages over the alternatives.

## REFERENCES

Shoham, Y. (1990). Nonmonotonic reasoning and causality. *Cognitive Science, 14,* 213–252.

Simon, H.A. (1952). On the definition of the causal relation. *The Journal of Philosophy, 49,* 517–528.

Simon, H.A. (1953). Causal ordering and identifiability. In T. Koopmans & W. Hood (Eds.), *Studies in econometric methods.* New York: Wiley & Sons.

Simon, H.A. (1977). *Models of discovery.* Dordrecht, Holland: Reidel.