Comments on Seeing and Doing

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

I am grateful to Professor Lindley for taking the time to study my book *Causality*, and for summarizing its main ideas so crisply and lucidly to the readers of this Journal. I would like to comment on a couple of issues that I believe warrant further emphasis in discussing causality. The first concerns the importance of mathematical notation for distinguishing causal from associational relationships, the second deals with the theoretical foundations of causality.

It is true that many members of Lindley's generation were not tormented by causal questions but, still, one should not take lightly the frustration of those who tried to tackle such questions and who could not find any mathematical machinery for solving, or even posing those questions. Having been part of this frustrated generation, I remember quite clearly how, even ten years ago, we could not express mathematically the simple fact that symptoms do not cause diseases, let alone draw mathematical conclusions from such facts. Asking for the probability that one event "caused" another was considered an ill-posed, metaphysical question that lies outside the province of statistical analysis.

Statisticians who had to interface with researchers in other disciplines have encountered countless barriers of confusion and miscommunication, all rooted in causation. The main users of statistical methods: economists, biologists, and health and social scientists, brought with them a wealth of substantive, $do(x)$-type information (also called "assumptions"), which they were unable to incorporate into statistical methods and techniques. Likewise, these users expected statistical methods to produce meaningful $do(x)$-type conclusions, but all statistics could deliver were study-specific associations of the $see(x)$ variety. In some applications (e.g., epidemiology), the absence of notational distinction between $do(x)$ and $see(x)$-type dependencies seemed unnecessary, because investigators were able to keep such distinctions implicitly in their heads, and managed to confine the mathematics to strictly conventional, $see(x)$-type expressions. In others, as in economics and the social sciences, investigators rebelled against this notational restriction by leaving mainstream statistics and constructing their own mathematical machinery (called Structural Equation Models), This machinery has remained a mystery to outsiders, and eventually became a mystery to insiders as well. "Every science is only so far exact as it knows how to express one thing by one sign," said Augustus de Morgan in 1858, and the results of not having the signs for expressing causality reached a critical point in the 1980-90's. Problems such as the control of confounding, the estimation of treatment effects, the distinction between direct and indirect effects, the estimation of probability of causation, and the combination of experimental and non-experimental data became a source of endless disputes among the users of statistics, and statisticians could not come to the rescue. *Causality* describes several such disputes, and why they could not be resolved by conventional statistical methodology)

One of my main reasons for writing *Causality* was to see such problems handled by mathematical analysis and, indeed, I now find it hard to name even a single problem in causal inference that cannot be expressed and solved by mathematical means. True, the analysis may tell us that the problem has no solution, namely, that additional information (or "assumptions" is needed. Still, the very assurance that more information is needed, coupled with formal facilities to identify the type of information needed is a much healthier state of affairs than the one prevailing in my tormented generation.

My second comments concerns the definition of causality. Some readers have expressed the opinion that causality is still an undefined concept and that, although the *do* calculus can be an effective mathematical tool in certain tasks, it does not bring us any closer to the deep and ultimate understanding of causality, one that is based solely on classical probability theory.

Unfortunately, aspirations for reducing causality to probability are both untenable and unwarranted. Philosophers have given up such aspirations twenty years ago, and were forced to admit extra-probabilistic concepts (such as "counterfactuals" or "causal relevance") into the analysis of causation (see *Causality*, Section 7.5). The reason is quite simple; probability theory deals with beliefs about an uncertain, yet static world, while causality deals with changes that occur in the world itself. Causality deals with how probability functions change in response to new conditions and interventions that originate from outside the probability space, while probability theory, even when given a fully specified joint density function on all variables in the space, cannot tell us how that function would change under external interventions. Thus, "doing" is not reducible to "seeing", and there is no point trying to fuse the two together. Drawing analogy to visual perception, the information contained in a probability function is analogous to a precise description of a three-dimensional object; it is sufficient for predicting how that object will be viewed from any angle outside the object, but it is insufficient for predicting how the object will be viewed if manipulated and squeezed by external forces. The additional information needed for making such predictions (e.g., the object's elasticity) is analogous to the causal information that the *do* calculus extracts from a directed acyclic graph (DAG).

From a mathematical perspective, it is a mistake to say that causality is still undefined. The *do* calculus, for example, is based on two well-defined mathematical objects: a probability function P and a DAG D; the first is standard in statistical analysis while the second is a newcomer that tells us (in a qualitative, yet formal language) which mechanisms would remain invariant to a given intervention. Given these two mathematical objects, the definition of "cause" is clear and crisp; variable X is a $probabilisticcause$ of variable Y if $P(y|do(x)) \neq P(y)$ for some values x and y. Since each of $P(y|do(x))$ and $P(y)$ is well-defined in terms of the pair (P, D), the relation "probabilistic cause" is, likewise, well-defined. Similar definitions can be constructed for other nuances of causal discourse, for example, "causal effect", "direct cause", "indirect cause" "event-to-event cause", "necessary cause", "sufficient cause", "likely cause" and "actual cause" (see $Causality$, pages 222-3, 286-7, 319; some of these definitions invoke functional models).

Not all statisticians are satisfied with these mathematical definitions. Some suspect definitions that are based on unfamiliar non-algebraic objects (i.e., the DAG) and some mistrust abstract definitions that are based on unverifiable models. Indeed, no mathematical machinery can ever verify whether a given DAG really represents the causal mechanisms that generate the data – such verification is left either to human judgment or to experimental studies that invoke interventions. I submit, however, that neither suspicion nor mistrust are justified in the case at hand; DAGs are no less formal than mathematical equations, and questions of model verification need be kept apart from those of conceptual definition. Consider, for example, the concept of a distribution *mean*. We certainly perceive this notion to be well-defined, for it can be computed from any given (non-pathological) distribution function, even before ensuring that we can estimate that distribution from the data. We would certainly not declare the mean "ill-defined" if, for any reason, we find it hard to estimate the distribution from the available data. Quite the contrary; by defining the mean in the abstract, as a functional of any hypothetical distribution, we can often prove that the defining distribution need not be estimated at all, and that the mean can be estimated (consistently) directly from the data. An analogous logic applies to causation. Causal quantities are first defined in the abstract, using the pair (P, D), and the abstract definition then provides a theoretical framework for deciding, given the type of data avaialable, whether the assumptions embodied in the DAG are sufficient (or necessary) for establishing the desired causal quantity from the data.

The separation between concept definition and model verification is even more pronounced in the Bayesian framework, where purely judgmental concepts, such as the prior distribution of the mean, are perfectly acceptable, as long as they can be assessed reliably from one's experience or knowledge. Professor Lindley's observation that "causal mechanisms may be easier to come by than one might initially think" further implies that, from a Bayesian perspective, the newcomer concept of a DAG is not an alien after all. If a Bayesian decision-maker is free to assess $p(y|see(x))$ and $p(y|do(x))$ in any way, as separate evaluations, the Bayesian should also be permitted to assess and assert his/her beliefs in the validity of the mechanisms portrayed in the DAG. Alternatively, the DAG can be viewed merely as a parsimonious scheme of encoding and maintaining coherence among those beliefs. And there is no need to cast these beliefs in the language of probabilities to render the analysis legitimate. Adding probabilistic veneer to these beliefs may make the *do* calculus appear more traditional, but would not change the fact that the objects of assessment are still causal mechanisms, and that these objects have their own special grammar of generating predictions about the effect of actions. Professor Lindley's observation reminds us that it is not the language in which we cast judgments that legitimizes the analysis, but whether those judgments can reliably be assessed from our store of knowledge and from the peculiar form in which this knowledge is organized.

If it were not for loss of reliability (of judgment), one could easily translate the information conveyed in a DAG into purely probabilistic formulae, using hypothetical variables. (Translation rules are provided in Section 7.3 of *Causality*, p. 232) Indeed, this is how the potential-outcome approach of Neyman and Rubin has achieved statistical legitimacy: judgments about causal relationships among observables are expressed as statements about probability functions that involve mixtures of observable and counterfactual variables. The difficulty with this approach, and the main reason for its slow acceptance in statistics, is that judgments about counterfactuals are much harder to assess than judgments about causal mechanisms. For instance, to communicate the simple assumption that symptoms do not cause diseases, we would have to use a rather unnatural expression and say that the probability of the counterfactual event "disease had symptoms been absent" is equal to the probability of "disease had symptoms been present" Judgments of conditional independencies among such counterfactual events are even harder for researchers to comprehend or to evaluate.

In summary, I suggest that it is through friendly conceptual semantics and powerful mathematical machinery that causal analysis will regain its proper place in statistics. I also submit that the theoretical foundations of causality are sharper and stronger when viewed as supplement to, not as part of, probability theory.