Bayesianism and Causality, or, Why I Am Only a Half-Bayesian

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

Abstract

The rationale behind the Bayesian paradigm consists on two major assumptions: 1. Background information is valuable, and it is silly not to use what one knows. 2. It is natural and advantageous to express background information in the language of prior probabilities. I will argue that the second assumption is false. Since the bulk of human knowledge consists of causal, not statistical relationships, the grammar of Bayesian priors is too crude for accomplishing the primary goal of the Bayesian program. To bring mathematics closer to where knowledge resides, I propose to enrich the language of probabilities with causal vocabulary, and to admit causal judgement into the Bayesian repretoire.

Keywords: Causal inference, Knowledge representation, Structural equations odels, graphical ethods, counterfactuals.

1 Introduction

I turned Bayesian in 1971, as soon as I began reading Savage's onograph *The Foundations of Statistical Inference* (Savage, 1962). The argu ents were unassailable: (i) It is plain silly to ignore what we know, (ii) It is natural

and useful to cast what we know in the language of probabilities, and (iii) If our subjective probabilities are erroneous, their i pact will get washed out in due ti e, as the nu ber of observations increases.

Thirty years later, I a still a devout Bayesian in the sense of (i), but I now doubt the wisdo of (ii) and I know that, in general, (iii) is false. Like ost Bayesians, I believe that the knowledge we carry in our skulls, be its origin experience, schooling or hearsay, is an invaluable resource in all hu an activity, and that co bining this knowledge with e pirical data is the key to scientific enquiry and intelligent behavior. Thus, in this broad sense, I a a still a Bayesian. However, in order to be co bined with data, our knowledge ust first be cast in so e for al language, and what I have co e to realize in the past ten years is that the language of probability is not suitable for the task; the bulk of hu an knowledge is organized around causal, not probabilistic relationships, and the gra ar of probability calculus is insufficient for capturing those relationships. Specifically, the building blocks of our scientific and everyday knowledge are ele entary facts such as " ud does not cause rain" and "sy pto s do not cause disease" and those facts, strangely enough, cannot be expressed in the vocabulary of probability calculus. It is for this reason that I consider yself only a half-Bayesian.

In the rest of the paper, I plan to review the dichoto y between causal and statistical knowledge, to show the li itation of probability calculus in handling the latter, to explain the i pact that this li itation has had on various scientific disciplines and, finally, I will express y vision for future develop ent in Bayesian philosophy: the enrich ent of personal probabilities with causal vocabulary and causal calculus, so as to bring athe atical analysis closer to where knowledge resides.

2 Statistics and Causality: A Brief Summary

The ai of standard statistical analysis, typified by regression and other esti ation techniques, is to infer para eters of a distribution fro sa ples drawn of that population. With the help of such para eters, one can infer associations a ong variables, esti ate the likelihood of past and future events, as well as update the likelihood of events in light of new evidence or

new easure ents. These tasks are anaged well by statistical analysis so long as experi ental conditions re ain the sa e. Causal analysis goes one step further; its ai is to infer the structure (na ely, the stable building blocks) of the data generation process. With the help of such structure, one can deduce not only the likelihood of events under static conditions, but also the dyna ics of events under *changing conditions*. This capability includes predicting the effect of actions (e.g., treat ents or policy decisions), identifying causes of reported events, and assessing responsibility and attribution (e.g., whether event x was necessary for the occurrence of event y).

Al ost by definition, causal and statistical concepts do not ix. Statistics deals with behavior under uncertain, yet static conditions, while causal analysis deals with changing conditions. There is nothing in the joint distribution of sy pto s and diseases to tell us that curing the for er would not cure the latter. In general, there is nothing in a distribution function that would tell us how that distribution would differ if external conditions were to change—say fro observational to experi ental setup—every conceivable difference in the distribution would be perfectly co patible with the laws of probability theory, no atter how slight the change in conditions. [1]

Drawing analogy to visual perception, the infor ation contained in a probability function is analogous to a precise description of a three-di ensional object; it is sufficient for predicting how that object will be viewed fro any angle outside the object, but it is insufficient for predicting how the object will be viewed if anipulated and squeezed by external forces. The additional infor ation needed for aking such predictions (e.g., the object's resilience or elasticity) is analogous to the infor ation that causal odels provide using the vocabulary of directed graphs and/or structural equations. The role of this infor ation is to identify those aspects of the world that re ain invariant when external conditions change, say due to an action.

These considerations i ply that the slogan "correlation does not i ply

[1] Even the theory of stochastic processes, which provides probabilistic characterization of dynamic phenomena, assumes a fixed density function over time-indexed variables. There is nothing in such a function to tell us how it would be altered if external conditions were to change. If a parametric family of distributions is used, we can represent some changes by selecting a different set of parameters. But we are still unable to represent changes that do not correspond to parameter selection; for example, restricting a variable to a certain value, or forcing one variable to equal another.

causation" can be translated into a useful principle: one cannot substantiate causal clai s fro associations alone, even at the population level—behind every causal conclusion there ust lie so e causal assu ption that is not testable in observational studies. Nancy Cartwright (1989) expressed this principle as "no causes in, no causes out", eaning we cannot convert statistical knowledge into causal knowledge.

The de arcation line between causal and statistical concepts is thus clear and crisp. A statistical concept is any concept that can be defined in ter s of a distribution (be it personal or frequency-based) of observed variables, and a causal concept is any concept about changes in variables that cannot be defined fro the distribution alone. Exa ples of statistical concepts are: correlation, regression, dependence, conditional independence, association, likelihood, collapsibility, risk ratio, odd ratio, and so on.[2] Exa ples of causal concepts are: rando ization, influence, effect, confounding, disturbance, spurious correlation, instru ental variables, intervention, explanation, attribution, and so on. The purpose of this de arcation line is not to exclude causal concepts fro the province of statistical analysis but, rather, to ake it easy for investigators and philosophers to trace the assu ptions that are needed for substantiating causal clai s. Every clai invoking causal concepts ust be traced to so e pre ises that invoke such concepts; it cannot be derived or inferred fro statistical clai s alone.

This principle ay sound obvious, al ost tautological, yet it has so e far reaching consequences. It i plies, for exa ple, that any syste atic approach to causal analysis ust acquire new athe atical notation for expressing causal assu ptions and causal clai s. The vocabulary of probability calculus, with its powerful operators of conditionalization and arginalization, is si ply insufficient for expressing causal infor ation. To illustrate, the syntax of probability calculus does not per it us to express the si ple fact that "sy pto s do not cause diseases", let alone draw athe atical conclusions fro such facts. All we can say is that two events are dependent— eaning that if we find one, we can expect to encounter the other, but we cannot distinguish statistical dependence, quantified by the conditional probability $P(disease|symptom)$ fro causal dependence, for which we have no expres-

[2] "The term 'risk ratio' and 'risk factors' have been used ambivalently in the literature; some authors insist on a risk factor having causal influence on the outcome, and some embrace factors that are merely associated with the outcome."

sion in standard probability calculus. [3] Scientists seeking to express causal relationships ust therefore supple ent the language of probability with a vocabulary for causality, one in which the sy bolic representation for the relation "sy pto s cause disease" is distinct fro the sy bolic representation of "sy pto s are associated with disease." Only after achieving such distinction can we label the for er sentence "false," and the latter "true."

The preceding two require ents: (1) to co ence causal analysis with untested,[4] judg entally based assu ptions, and (2) to extend the syntax of probability calculus, constitute, in y experience, the two ain obstacles to the acceptance of causal analysis a ong statisticians, philosophers and professionals with traditional training in statistics. We shall now explore in ore detail the nature of these two barriers, and why they have been so tough to cross.

2.1 The Barrier of Untested Assumptions

Many statistical studies are based on so e untested assu ptions. For exa - ples, we often assu e that variables are ultivariate nor al, that the density function has certain s oothness properties, or that a certain para eter falls in a given range. The question thus arises why innocent causal assu ptions, say, that sy pto s do not cause disease or that ud does not cause rain, invite istrust and resistance a ong statisticians, especially of the Bayesian school.

There are three funda ental differences between statistical and causal assu ptions. First, statistical assu ptions, even untested, are testable in principle, given sufficiently large sa ple and sufficiently fine easure ents. Causal assu ptions, in contrast, cannot be verified even in principle, unless one resorts to experi ental control. This difference is especially accentuated in Bayesian analysis. Though the priors that Bayesians co only assign to statistical para eters are untested quantities, the sensitivity to these priors tends to di inish with increasing sa ple size. In contrast, sensitivity to priors of causal para eters, say the effect of s oking on lung cancer, re ains

[3] Attempts to define causal dependence by conditioning on the entire past (e.g., Suppes, 1970) violate the statistical requirement of limiting the analysis to "observed variables", and encounter other insurmountable difficulties (see Eells (1991), Pearl (2000), pp. 249-257).

[4] By "untested" I mean untested using frequency data in nonexperimental studies.

finite regardless of (nonexperi ental) sa ple size.

Second, statistical assu ptions can be expressed in the fa iliar language of probability calculus, and thus assu e an aura of scholarship and scientific respectability. Causal assu ptions, as we have seen before, are deprived of that honor, and thus beco e i ediate suspect of infor al, anecdotal or etaphysical thinking. Again, this difference beco es illu inated a ong Bayesians, who are accusto ed to accepting untested, judg ental assu ptions, and should therefore invite causal assu ptions with open ar s—they don't. A Bayesian is prepared to accept experts judg ent, however esoteric and untestable, so long as the judg ent is wrapped in the safety blanket of a probability expression. Bayesians turn extre ely suspicious when that sa e judg ent is cast in plain English, as in " ud does not cause rain." A typical exa ple can be seen in Lindley and Novick's (1981) treat ent of Si pson's paradox.

Lindley and Novick showed that decisions on whether to use conditional or arginal contingency tables should depend on the story behind the tables, that is, on one's assu ption about how the tables were generated. For exa ple, to decide whether a treat ent $X = x$ is beneficial ($Y = y$) in a population, one should co pare $\Sigma_z P(y|x,z)$ to $\Sigma_z P(y|x',z)$ if Z stands for the gender of patients. In contrast, if Z stands for a factor that is affected by the treat ent (say blood pressure), one should co pare the arginal probabilities, $P(y|x)$ vis-à-vis $P(y|x')$, and refrain fro conditioning on Z (see (Pearl, 2000; pp. 174-182) for details). Re arkably, instead of attributing this difference to the causal relationships in the story, Lindley and Novick wrote: "We have not chosen to do this; nor to discuss causation, because the concept, although widely used, does not see to be well-defined" (p. 51). Thus, instead of discussing causation, they attribute the change in strategy to another untestable relationship in the story—exchangeability [DeFinetti, 1974] which is cognitively for idable yet, at least for ally, can be cast in a probability expression. In Section 4.2, we will return to discuss this trend a ong Bayesians of equating "definability" with expressibility in probabilistic language.

The third resistance to causal (vis-à-vis statistical) assu ptions ste s fro their inti idating clarity. Assu ptions about abstract properties of density functions or about conditional independencies a ong variables are, congnitively speaking, rather opaque, hence they tend to be forgiven, rather than debated. In contrast, assu ptions about how variables cause one

another are shockingly transparent, and tend therefore to invite counter-argu ents and counter-hypotheses. A co-reviewer on a paper I have read recently offered the following objection to the causal odel postulated by the author:

> "A thoughtful and knowledgeable epide iologist could write down two or ore equally plausible odels that leads to different conclusions regarding confounding."

Indeed, since the bulk of scientific knowledge is organized in causal sche a, scientists are incredibly creative in constructing co peting alternatives to any causal hypothesis, however plausible. Statistical hypotheses in contrast, having been several levels re oved fro our store of knowledge, are relatively protected fro such challenge.

I conclude this subsection with a suggestion that statisticians' suspicion of causal assu ptions, vis-*à*-vis probabilistic assu ptions is unjustified. Considering the language of scientific knowledge, it akes prefect sense that we per it scientists to articulate what they know in plain causal expressions, and not force the to co pro ise reliability by converting to the "higher level" language of prior probabilities, conditional independence and other cognitively unfriendly ter inology. [5]

2.2 The Barrier of New Notation

If reluctance to aking causal assu ptions has been a hindrance to causal analysis, finding a athe atical way of expressing such assu ptions encountered a for idable ental block. The need to adopt a new notation, foreign to the province of probability theory, has been trau atic to ost persons trained in statistics; partly because the adaptation of a new language is difficult in general, and partly because statisticians have been accusto ed to assu ing that all pheno ena, processes, thoughts, and odes of inference can be captured in the powerful language of probability theory.[6]

[5] Similar observations were expressed by J. Heckman (2001).

[6] Commenting on my $set(x)$ notation [Pearl, 1995a, b], a leading statistician wrote: "Is this a concept in some new theory of probability or expectation? If so, please provide it. Otherwise, 'metaphysics' may remain the leading explanation." Another statistician, commenting on the $do(x)$ notation used in *Causality* [Pearl, 2000a], insisted: "...the calculus of probability is the calculus of causality."

Not surprisingly, in the bulk of the statistical literature, causal clai s never appear in the athe atics. They surface only in the verbal interpretation that investigators occasionally attach to certain associations, and in the verbal description with which investigators justify assu ptions. For exa ple, the assu ption that a covariate is not affected by a treat ent, a necessary assu ption for the control of confounding [Cox, 1958], is expressed in plain English, not in a athe atical equation.

In so e applications (e.g., epide iology), the absence of notational distinction between causal and statistical dependencies see ed unnecessary, because investigators were able to keep such distinctions i plicitly in their heads, and anaged to confine the athe atics to conventional probability expressions. In others, as in econo ics and the social sciences, investigators rebelled against this notational tyrany by leaving ainstrea statistics and constructing their own athe atical achinery (called Structural Equations Models). Unfortunately, this achinery has re ained a ystery to outsiders, and eventually beca e a ystery to insiders as well. [7]

But such tensions could not re ain dor ant forever. "Every science is only so far exact as it knows how to express one thing by one sign," said Augustus de Morgan in 1858 – the harsh consequences of not having the signs for expressing causality surfaced in the 1980-90's. Proble s such as the control of confounding, the esti ation of treat ent effects, the distinction between direct and indirect effects, the esti ation of probability of causation, and the co bination of experi ental and nonexperi ental data beca e a source of endless disputes a ong the users of statistics, and statisticians could not co e to the rescue. (Pearl, 2000) describes several such disputes, and why they could not be resolved by conventional statistical ethodology.

[7] Most econometric texts in the last decade have refrained from defining what an economic model is, and those that attempted a definition, erroneously view structural equations models as compact representations of density functions (see Pearl, 2000, pp. 135-138).

3 Languages for Causal Analysis

3.1 The language of diagrams and structural equations

How can one express athematically the co on understanding that sy p- to s do not cause diseases? The earliest atte pt to for ulate such relationship athematically was ade in the 1920's by the geneticist Sewall Wright (1921). Wright used a co bination of equations and graphs to co unicate causal relationships. For exa ple, if X stands for a disease variable and Y stands for a certain sy pto of the disease, Wright would write a linear equation:

$$y = ax + u \tag{1}$$

supple ented with the diagra $X \longrightarrow Y$, where x stands for the level (or severity) of the disease, y stands for the level (or severity) of the sy pto , and u stands for all factors, other than the disease in question, that could possibly affect Y (U is called "exogeneous", "background", or "disturbance".) The diagra encodes the possible existence of (direct) causal influence of X on Y, and the absence of causal influence of Y on X, while the equation encodes the quantitative relationships a ong the variables involved, to be deter ined fro the data. The para eter a in the equation is called a "path coefficient" and it quantifies the (direct) causal effect of X on Y; given the nu erical value of a, the equation clai s that, ceteris paribus, a unit increase in X would result in an a-unit increase of Y. If correlation between X and U is presu ed possible, it is custo ary to add a double arrow between X and Y.

The asy etry induced by the diagra renders the equality sign in Eq. (1) different fro algebraic equaltity, rese bling instead the assign ent sy - bol (:=) in progra ing languages. Indeed, the distinctive characteristic of structural equations, setting the apart fro algebraic equations, is that they stand for a value-assign ent process – an autono ous echanis by which the value of Y (not X) is deter ined. In this assign ent process, Y is co itted to track changes in X, while X is not subject to such co it ent [8].

[8] Clearly, if we intevene on X, Y would continue to track changes in X. Not so when we intervene on Y, X is not commitment to track changes in Y. Such intervention would alter

Wright's major contribution to causal analysis, aside from introducing the language of path diagrams, has been the development of graphical rules for writing down (by inspection) the covariance of any pair of observed variables in terms of path coefficients and of covariances among disturbances. Under certain causal assumptions, (e.g. if $Cov(U, X) = 0$), the resulting equations may allow one to solve for the path coefficients in terms of observed covariance terms only, and this amounts to inferring the magnitude of (direct) causal effects from observed, nonexperimental associations, assuming of course that one is prepared to defend the causal assumptions encoded in the diagram.

The causal assumptions embodied in the diagram (e.g, the absence of arrow from Y to X, or $Cov(U, X) = 0$) are not generally testable from nonexperimental data. However, the fact that each causal assumption in isolation cannot be tested does not mean that the summation total of all causal assumptions in a model does not have testable implications. The chain model $X \longrightarrow Y \longrightarrow Z$ for example, encodes many causal assumptions, each corresponding to a missing arrow or a missing double-arrow between a pair of variables. None of those assumptions is testable in isolation, yet the totality of all those assumptions implies that Z is unassociated with X, conditioned on Y. Such testable implications can be read off the diagrams (see [Pearl 2000, pp. 16-19]), and these constitute the only opening through which the assumption embodies in structural equation models can be tested in observational studies. Every conceivable statistical test that can be applied to the model is entailed by those implications.

3.2 From path-diagrams to *do*-calculus

Structural equation modeling (SEM) has been the main vehicle for causal analysis in economics, and the behavioral and social sciences [Goldberger 1972; Duncan 1975]. However, the bulk of SEM methodology was developed for linear analysis and, until recently, no comparable methodology has been devised to extend its capabilities to models involving discrete variables, nonlinear dependencies, or situations in which the functional form of the equations is unknown. A central requirement for any such extension is to detach the notion of "effect" from its algebraic representation as a coefficient in an equation, and redefine "effect" as a general capacity to transmit *changes*

the assignment mechanism for Y and, naturally, would cause the equality to be violated.

a ong variables. One such extension, based on si ulating hypothetical interventions in the odel, is presented in Pearl (1995a, 2000)

The central idea is to exploit the invariant characteristics of structural equations without co itting to a specific functional for . For exa ple, the non-para etric interpretation of the chain odel $Z \longrightarrow X \longrightarrow Y$ corresponds to a set of three functions, each corresponding to one of the variables:

$$\begin{aligned} z &= f_Z(w) \\ x &= f_X(z, v) \\ y &= f_Y(x, u) \end{aligned} \tag{2}$$

together with the assu ption that the background variables W,V, U (not shown in the chain) are jointly independent but, otherwise, arbitrarily distributed. Each of these functions represents a causal process (or echanis) that deter ines the value of the left variable (output) fro those on the right variables (input). The absence of a variable fro the right hand side of an equations encodes the assu ption that it has no direct effect on the left variable. For exa ple, the absence of variable Z fro the argu ents of f_Y indicates that variations in Z will leave Y unchanged, as long as variables U and X re ain constant. A syste of such functions are said to be *structural* (or *modular*) if they are assu ed to be autono ous, that is, each function is invariant to possible changes in the for of the other functions [Si on 1953; Koop ans 1953].

This feature of invariance per its us to use structural equation as a basis for odeling actions and counterfactuals. This is done through a athe atical operator called $do(x)$ which si ulates physical interventions by deleting certain functions fro the odel, replacing the by constants, while keeping the rest of the odel unchanged. For exa ple, to represent an intervention that sets the value of X to x_0 the odel for Eq. (2) would beco e

$$\begin{aligned} z &= f_Z(w) \\ x &= x_0 \\ y &= f_Y(x, u) \end{aligned} \tag{3}$$

The distribution of Y and Z calculated fro this odified odel characterizes the effect of the action $do(X = x_0)$ and is denoted as $P(y, z|do(x_0))$. It is not

hard to show that, as expected, the odel of Eq. (2) yields $P(y|do(x_0)) = P(y|x_0)$ and $P(z|do(x_0)) = P(z)$ regardless of the functions f_X, f_Y and f_Z. Additional features of this transfor ation are discussed in the Appendix; see (Pearl, 2000; chapter 7) for full details.

The ain task of causal analysis is to infer causal quantities fro the observed distribution $P(x, y, z)$, or fro sa ples of that distribution. Such analysis requires athe atical eans of transfor ing causal quantities, represented by expressions such as $P(y|do(x))$, into *do*-free expressions derivable fro $P(z, x, y)$, since only *do*-free expressions are esti able fro non-experi ental data. When such a transfor ation is feasible, we say that the causal quantity is *identifiable*. A calculus for perfor ing such transfor ations, called *do*-calculus, was developed in [Pearl, 1995a]. Re arkably, the rules governing this calculus depend erely on the topology of the diagra ; it takes no notice of the functional for of the equations, nor of the distribution of the disturbance ter s. This calculus per its the investigator to inspect the causal diagra and

1. Decide whether the assu ptions e bodied in the odel are sufficient to obtain consistent esti ates of the target quantity;

2. Derive (if the answer to ite 1 is affir ative) a closed-for expression for the target quantity in ter s of distributions of observed quantities; and

3. Suggest (if the answer to ite 1 is negative) a set of observations and experi ents that, if perfor ed, would render a consistent esti ate feasible.

4 On the Definition of Causality

In this section, I return to discuss concerns expressed by so e Bayesians that causality is an undefined concept and that, although the *do*-calculus can be an effective athe atical tool in certain tasks, it does not bring us any closer to the deep and ulti ate understanding of causality, one that is based solely on classical probability theory.

4.1 Is causality reducible to probabilities?

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 ad it extra-probabilistic pri itives (such as "counterfactuals" or "causal relevance") into the analysis of causation (see Eells (1991) and Pearl (2000), Section 7.5). The basic reason was alluded to in Section 2: probability theory deals with beliefs about an uncertain, yet static world, while causality deals with changes that occur in the world itself, or in one's theory of the world. Causality deals with how probability functions change in response to new conditions and interventions that originate fro outside the probability space, while probability theory, even when given a fully specified joint density function on all (te porally-indexed) 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.

Many philosophers have aspired to show that the calculus of probabilities, endowed with a ti e dyna ic, would be sufficient for causation (Suppes, 1970). A well known de onstration of the i possibility of such reduction (following Otte (1981)) goes as follows. Consider a switch X that turns on two lights, Y and Z, and assu e that, due to differences in location, Z turns on a split second before Y. Consider now a variant of this exa ple where the switch X activates Z, and Z, in turns, activates Y. This case is probabilistically identical to the previous one, because both the functional and te poral relationships are identical. Yet few people would perceive the causal relationships to be the sa e in the two situations; the latter represents cascaded process, $X \longrightarrow Z \longrightarrow Y$, while the for er represents a branching process, $Y \longleftarrow X \longrightarrow Z$. The difference shows, of course, when we consider interventions; intervening on Z would affect Y in the cascaded case, but not in the branching case.

The preceding exa ple illustrates the essential role of *mechanisms* in defining causation. In the branching case, although all three variables are sy etrically constrained by the functional relationships: $X = Y$, $X = Y$, $Z = Y$, these relationships in the selves do not reveal the infor ation that the three equalities are sustained by only two echanis s, $Y = X$ and $Z = X$, and that the first equality would still be sustained when the second is violated. A set of echanis s, each represented by an equation, is not

equivalent to the set of algebraic equations that can be asse bled fro those echanis s. Mathe atically, the latter is defined as *one* set of n equations, whereas the for er is defined as n separate sets, each containing one equation. These are two distinct athe atical objects that ad it two distinct types of solution-preserving operations. The calculus of causality deals with the dyna ics of such odular syste s of equations, where the addition and deletion of equations represent interventions (see Appendix).

4.2 Is causality well-defined?

¿Fro a athe atical perspective, it is a istake to say that causality is undefined. The *do* calculus, for exa ple, is based on two well-defined athematical objects: a probability function P and a directed acyclic graph (DAG) D; the first is standard in statistical analysis while the second is a newco er that tells us (in a qualitative, yet for al language) which echanis s would re ain invariant to a given intervention. Given these two athematical objects, the definition of "cause" is clear and crisp; variable X is a *probabilistic-cause* of variable Y if $P(y|do(x)) \neq P(y)$ for so e values x and y. Since each of $P(y|do(x))$ and $P(y)$ is well-defined in ter s of the pair (P, D), the relation "probabilistic cause" is, likewise, well-defined. Si ilar definitions can be constructed for other nuances of causal discourse, for exa ple, "causal effect", "direct cause", "indirect cause" "event-to-event cause", "scenario-specific cause", "necessary cause", "sufficient cause", "likely cause" and "actual cause" (see (Pearl, 2000), pages 222-3, 286-7, 319; so e of these definitions invoke functional odels).

Not all statisticians/philosophers are satisfied with these athe atical definitions. So e suspect definitions that are based on unfa iliar nonalgebraic objects (i.e., the DAG) and so e istrust abstract definitions that are based on unverifiable odels. Indeed, no athe atical achinery can ever verify whether a given DAG really represents the causal echanis s that generate the data – such verification is left either to hu an judg ent or to experi ental studies that invoke interventions. I sub it, however, that neither suspicion nor istrust are justified in the case at hand; DAGs are no less for al than athe atical equations, and questions of odel verification need be kept apart fro those of conceptual definition.

Consider, for exa ple, the concept of a distribution *mean*. Even non-Bayesians perceive this notion to be well-defined, for it can be co puted fro

any given (non-pathological) distribution function, even before ensuring that we can esti ate that distribution fro the data. We would certainly not declare the ean "ill-defined" if, for any reason, we find it hard to esti ate the distribution fro the available data. Quite the contrary; by defining the ean in the abstract, as a functional of any hypothetical distribution, we can often prove that the defining distribution need not be esti ated at all, and that the ean can be esti ated (consistently) directly fro the data. Re arkably, by taking seriously the abstract (and untestable) notion of a distribution, we obtain a lisence to ignore it. 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 fra ework for deciding, given the type of data available, which of the assu ptions e bodied in the DAG are ignorable, and which are absolutely necessary for establishing the target causal quantity fro the data. [9]

The separation between concept definition and odel verification is even ore pronounced in the Bayesian fra ework, where purely judg ental concepts, such as the prior distribution of the ean, are perfectly acceptable, as long as they can be assessed reliably fro one's experience or knowledge. Dennis Lindley has re arked recently (personal co unication) that "causal echanis s ay be easier to co e by than one ight initially think". Indeed, fro a Bayesian perspective, the newco er concept of a DAG is not an alien at all – it is at least as legiti ate as the probability assess ents that a Bayesian decision- aker pronounces in constructing a decision tree. In such construction, the probabilities that are assigned to branches e anating fro a decision variable X correspond to assess ents of $P(y|do(x))$ and those assigned to branches e anating fro a chance variable X correspond to assess ents of $P(y|x)$. If a Bayesian decision- aker is free to assess $P(y|x)$ and $P(y|do(x))$ in any way, as separate evaluations, the Bayesian should also be per itted to express his/her conception of the echanis s that entail those evaluations. It is only by envisioning these echanis s that a decision aker can generate a coherent list of such a vast nu ber of $P(y|do(x))$ type assess ents.[10] The structure of the DAG can certainly be

[9] I have used a similar logic in defense of counterfactuals (Pearl, 2000a), which Dawid (2000) deemed dangerous on account of being untestable. (See, also Dawid (2001), this volume.) Had Bernoulli been constrained by Dawid's precautions, the notion of a "distribution" would have had to wait for another adventurous metaphysician to be created.

[10] Coherence requires, for example, that for any x, y, and z, the inequality

recovered fro judg ents of the for $P(y|do(x))$ and, conversely, the DAG co bined with a probability function P dictates all judg ents of the for $P(y|do(x))$. Accordingly the structure of the DAG can be viewed as a qualitative parsi onious sche e of encoding and aintaining coherence a ong those assess ents. And there is no need to translate the DAG into the language of probabilities to render the analysis legiti ate. Adding probabilistic veneer to the echanis s portrayed in the DAG ay ake the *do* calculus appear ore traditional, but would not change the fact that the objects of assess ent are still causal echanis s, and that these objects have their own special gra ar of generating predictions about the effect of actions. In su ary, recalling the unlti ate Bayesian ission of fusing judg ent, it is not the language in which we cast judg ents that legiti izes the analysis, but whether those judg ents can reliably be assessed fro our store of knowledge and fro the peculiar for in which this knowledge is organized.

If it were not for loss of reliability (of judg ent), one could easily translate the infor ation conveyed in a DAG into purely probabilistic for ulae, using hypothetical variables. (Translation rules are provided in (Pearl, 2000, p. 232) Indeed, this is how the potential-outco e approach of Ney an [Ney an, 1923]and Rubin [Rubin, 1974] has achieved statistical legiti acy: judg ents about causal relationships a ong observables are expressed as state ents about probability functions that involve ixtures of observable and counterfactual variables. The difficulty with this approach, and the ain reason for its slow acceptance in statistics, is that judg ents about counterfactuals are uch harder to assess than judg ents about causal echanis s. For instance, to co unicate the si ple assu ption that sy pto s do not cause diseases, we would have to use a rather roundabout expression and say that the probability of the counterfactual event "disease had sy pto s been absent" is equal to the probability of "disease had sy pto s been present". Judg ents of conditional independencies a ong such counterfactual events are even harder for researchers to co prehend or to evaluate.

$P(y|do(x), do(z)) > P(y, x|do(z))$ be satisfied. This follows from the property of composition (see Appendix, Eq. (6), or (Pearl, 2000; pp. 229)

5 Summary

This paper calls attention to a basic conflict between ission and practice in Bayesian ethodology. The ission is to express prior knowledge ath- e atically and reliably so as to assist the interpretation of data, hence the acquisition of new knowledge. The practice is to express prior knowledge as prior probabilities – too crude a vocabulary, given the grand ission. Considerations of reliability (of judge ent) call for enriching the language of probabilities with causal vocabulary and for ad itting causal judg ents into the Bayesian repertoire. The athe atics for interpreting causal judg ents has atured, and tools for using such judg ents in the acquisition of new knowledge have been developed. The grounds are now ready for ission-oriented Bayesianis .

6 Appendix - Causal Models, Actions and Counterfactuals

This appendix presents a brief su ary of the structural-equation se antics of causation and counterfactuals as defined in Balke and Pearl (1995), Galles and Pearl (1997, 1998), and Halpern (1998). For detailed exposition of the structural account and its applications see [Pearl, 2000a].

Causal odels are generalizations of the structural equations used in engineering, biology, econo ics and social science.[11] World knowledge is represented as a odular collection of stable and autono ous relationships called " echanis s," each represented as a function, and changes due to interventions or un odelled eventualities are treated as local odifications of these functions.

A causal odel is a athe atical object that assigns truth values to sentences involving causal relationships, actions, and counterfactuals. We will first define causal odels, then discuss how causal sentences are evaluated in such odels. We will restrict our discussion to recursive (or feedback-free) odels; extensions to non-recursive odels can be found in Galles and Pearl (1997, 1998) and Halpern (1998).

[11] Similar models, called "neuron diagrams" [Lewis, 1986, p. 200; Hall, 1998] are used informally by philosophers to illustrate chains of causal processes.

Definition 6.1 (*Causal model*)
A causal odel *is a triple*

$$M = \langle U, V, F \rangle$$

where

(i) *U is a set of variables, called* exogenous. *(These variables will represent background conditions, that is, variables whose values are determined outside the model.)*

(ii) *V is an ordered set $\{V_1, V_2, \ldots, V_n\}$ of variables, called* endogenous. *(These represent variables that are determined in the model, namely, by variables in $U \cup V$.)*

(iii) *F is a set of functions $\{f_1, f_2, \ldots, f_n\}$ where each f_i is a mapping from $U \times (V_1 \times \ldots \times V_{i-1})$ to V_i. In other words, each f_i tells us the value of V_i given the values of U and all predecessors of V_i. Symbolically, the set of equations F can be represented by writing* [12]

$$v_i = f_i(pa_i, u_i) \quad i = 1, \ldots, n$$

where pa_i is any realization of the unique minimal set of variables PA_i in V (connoting parents*) sufficient for representing f_i.*[13] *Likewise, $U_i \subseteq U$ stands for the unique minimal set of variables in U that is sufficient for representing f_i.*

Every causal odel M can be associated with a directed graph, $G(M)$, in which each node corresponds to a variable in V and the directed edges point fro e bers of PA_i toward V_i (by convention, the exogenous variables are usually not shown explicitly in the graph). We call such a graph the *causal graph* associated with M. This graph erely identifies the endogenous variables PA_i that have direct influence on each V_i but it does not specify the functional for of f_i.

[12] We use capital letters (e.g., X, Y) as names of variables and sets of variables, and lower-case letters (e.g., x, y) for specific values (called realizations) of the corresponding variables.

[13] A set of variables X is *sufficient* for representing a given function $y = f(x, z)$ if f is trivial in Z—that is, if for every x, z, z' we have $f(x, z) = f(x, z')$.

For any causal odel, we can define an *action* operator, $do(x)$, which, fro a conceptual viewpoint, si ulates the effect of external action that sets the value of X to x and, fro a for al viewpoint, transfor s the odel into a *submodel*, that is, a causal odel containing fewer functions.

Definition 6.2 *(Submodel)*
Let M be a causal model, X be a set of variables in V, and x be a particular assignment of values to the variables in X. A sub odel M_x *of M is the causal model*

$$M_x = \langle U, V, F_x \rangle$$

where

$$F_x = \{f_i : V_i \notin X\} \cup \{X = x\} \tag{4}$$

In words, F_x is for ed by deleting fro F all functions f_i corresponding to e bers of set X and replacing the with the set of constant functions $X = x$.

If we interpret each function f_i in F as an independent physical echanis and define the action $do(X = x)$ as the ini al change in M required to ake $X = x$ hold true under any u, then M_x represents the odel that results fro such a ini al change, since it differs fro M by only those echanis s that directly deter ine the variables in X. The transfor ation fro M to M_x odifies the algebraic content of F, which is the reason for the na e *modifiable structural equations* used in [Galles and Pearl, 1998].[14]

Definition 6.3 (*Effect of action*)
Let M be a causal model, X be a set of variables in V, and x be a particular realization of X. The effect of action $do(X = x)$ *on M is given by the submodel M_x.*

Definition 6.4 (*Potential response*)
Let Y be a variable in V, let X be a subset of V, and let u be a particular

[14]Structural modifications date back to Marschak (1950) and Simon (1953). An explicit translation of interventions into "wiping out" equations from the model was first proposed by Strotz and Wold (1960) and later used in Fisher (1970), Sobel (1990), Spirtes et al. (1993), and Pearl (1995). A similar notion of sub-model is introduced in Fine (1985), though not specifically for representing actions and counterfactuals.

value of U. The potential response *of Y to action $do(X = x)$ in situation u, denoted $Y_x(u)$, is the (unique) solution for Y of the set of equations F_x.*

We will confine our attention to actions in the for of $do(X = x)$. Conditional actions, of the for "$do(X = x)$ if $Z = z$" can be for alized using the replace ent of equations by functions of Z, rather than by constants [Pearl, 1994]. We will not consider disjunctive actions, of the for "$do(X = x$ or $X = x')$", since these co plicate the probabilistic treat ent of counterfactuals.

Definition 6.5 (*Counterfactual*)
Let Y be a variable in V, and let X be a subset of V. The counterfactual expression "The value that Y would have obtained, had X been x" is interpreted as denoting the potential response $Y_x(u)$.

Definition 5 thus interprets the counterfactual phrase "had X been x" in ter s of a hypothetical external action that odifies the actual course of history and i poses the condition "$X = x$" with ini al change of echanis s. This is a crucial step in the se antics of counterfactuals [Balke and Pearl, 1994], as it per its x to differ fro the actual value $X(u)$ of X without creating logical contradiction; it also suppresses abductive inferences (or backtracking) fro the counterfactual antecedent $X = x$.[15]

It can easily be shown [Galles and Pearl, 1997] that the counterfactual relationship just defined, $Y_x(u)$, satisfies the following two properties:
Effectiveness:
For any two disjoint sets of variables, Y and W, we have

$$Y_{yw}(u) = y. \tag{5}$$

In words, setting the variables in W to w has no effect on Y, once we set the value of Y to y.
Composition:
For any two disjoint sets of variables X and W, and any set of variables Y,

$$W_x(u) = w \Longrightarrow Y_{xw}(u) = Y_x(u). \tag{6}$$

[15] Simon and Rescher (1966, p. 339) did not include this step in their account of counterfactuals and noted that backward inferences triggered by the antecedents can lead to ambiguous interpretations.

In words, once we set X to x, setting the variables in W to the sa e values, w, that they would attain (under x) should have no effect on Y. Further- ore, effectiveness and co position are *complete* whenever M is recursive (i.e., $G(M)$ is acyclic) [Galles and Pearl, 1998, Halpern, 1998], that is, every property of counterfactuals that follows fro the structural odel se antics can be derived by repeated application of effectiveness and co position.

A corollary of co position is a property called *consistency* by [Robins, 1987]:

$$(X(u) = x) \Longrightarrow (Y_x(u) = Y(u)) \tag{7}$$

Consistency states that, if in a certain context u we find variable X at value x, and we intervene and set X to that sa e value, x, we should not expect any change in the response variable Y. Co position and consistency are used in several derivations of Section 3.

The structural for ulation generalizes naturally to probabilistic syste s, as is seen below.

Definition 6.6 (*Probabilistic causal model*)
A probabilistic causal odel *is a pair*

$$\langle M, P(u) \rangle$$

where M is a causal model and $P(u)$ is a probability function defined over the domain of U.

$P(u)$, together with the fact that each endogenous variable is a function of U, defines a probability distribution over the endogenous variables. That is, for every set of variables $Y \subseteq V$, we have

$$P(y) \stackrel{\Delta}{=} P(Y = y) = \sum_{\{u \mid Y(u)=y\}} P(u) \tag{8}$$

The probability of counterfactual state ents is defined in the sa e anner, through the function $Y_x(u)$ induced by the sub odel M_x. For exa ple, the *causal effect* of x on y is defined as:

$$P(Y_x = y) = \sum_{\{u \mid Y_x(u)=y\}} P(u) \tag{9}$$

Likewise, a probabilistic causal odel defines a joint distribution on counterfactual state ents, i.e., $P(Y_x = y, Z_w = z)$ is defined for any sets of variables Y, X, Z, W, not necessarily disjoint. In particular, $P(Y_x = y, X = x')$ and $P(Y_x = y, Y_{x'} = y')$ are well defined for $x \neq x'$, and are given by

$$P(Y_x = y, X = x') = \sum_{\{u | Y_x(u)=y \ \& \ X(u)=x'\}} P(u) \tag{10}$$

and

$$P(Y_x = y, Y_{x'} = y') = \sum_{\{u \ | \ Y_x(u)=y \ \& \ Y_{x'}(u)=y'\}} P(u). \tag{11}$$

When x and x' are inco patible, Y_x and $Y_{x'}$ cannot be easured si ultaneously, and it ay see eaningless to attribute probability to the joint state ent "Y would be y if $X = x$ *and* Y would be y' if $X = x'$." Such concerns have been a source of recent objections to treating counterfactuals as jointly distributed rando variables [Dawid, 2000]. The definition of Y_x and $Y_{x'}$ in ter s of two distinct sub odels, driven by a standard probability space over U, de onstrates that joint probabilities of counterfactuals have solid athe atical and conceptual underpinning and, oreover, these probabilities can be encoded rather parsi oniously using $P(u)$ and F.

References

[Balke and Pearl, 1994] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volu e I, pages 230–237. MIT Press, Menlo Park, CA, 1994.

[Balke and Pearl, 1995] A. Balke and J. Pearl. Counterfactuals and policy analysis in structural odels. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 11–18. Morgan Kauf ann, San Francisco, 1995.

[Cartwright, 1989] N. Cartwright. *Nature's Capacities and Their Measurement*. Clarendon Press, Oxford, 1989.

[Cox, 1958] D.R. Cox. *The Planning of Experiments*. John Wiley and Sons, NY, 1958.

[Dawid, 2000] A.P. Dawid. Causal inference without counterfactuals (with co ents and rejoinder). *Journal of the American Statistical Association*, 95(450):407–448, June 2000.

[DeFinetti, 1974] B. DeFinetti. *Theory of Probability: A Critical Introductory Treatment,* 2 volu es (Translated by A. Machi and A. S ith). Wiley, London, 1974.

[Duncan, 1975] O.D. Duncan. *Introduction to Structural Equation Models.* Acade ic Press, New York, 1975.

[Eells, 1991] E. Eells. *Probabilistic Causality.* Ca bridge University Press, Ca bridge, MA, 1991.

[Fine, 1985] K. Fine. *Reasoning with Arbitrary Objects.* B. Blackwell, New York, 1985.

[Fisher, 1970] F.M. Fisher. A correspondence principle for si ultaneous equations odels. *Econometrica*, 38(1):73–92, January 1970.

[Galles and Pearl, 1997] D. Galles and J. Pearl. Axio s of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

[Galles and Pearl, 1998] D. Galles and J. Pearl. An axio atic characterization of causal counterfactuals. *Foundation of Science*, 3(1):151–182, 1998.

[Goldberger, 1972] A.S. Goldberger. Structural equation odels in the social sciences. *Econometrica: Journal of the Econometric Society*, 40:979–1001, 1972.

[Hall, 1998] N. Hall. Two concepts of causation, 1998. In press.

[Halpern, 1998] J.Y. Halpern. Axio atizing causal reasoning. In G.F. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, pages 202–210. Morgan Kauf ann, San Francisco, CA, 1998.

[Heck an, 2001] J.J. Heck an. Econo etrics and e pirical econo ics. *Journal of Econometrics*, 100(1):c–5, 2001.

[Koop ans, 1953] T.C. Koop ans. Identification proble s in econo etric odel construction. In W.C. Hood and T.C. Koop ans, editors, *Studies in Econometric Method*, pages 27–48. Wiley, New York, 1953.

[Lewis, 1986] D. Lewis. *Philosophical Papers*. Oxford University Press, New York, 1986.

[Lindley and Novick, 1981] D.V. Lindley and M.R. Novick. The role of exchangeability in inference. *The Annals of Statistics*, 9(1):45–58, 1981.

[Marschak, 1950] J. Marschak. Statistical inference in econo ics. In T. Koop ans, editor, *Statistical Inference in Dynamic Economic Models*, pages 1–50. Wiley, New York, 1950. Cowles Co ission for Research in Econo ics, Monograph 10.

[Ney an, 1923] J. Ney an. On the application of probability theory to agricultural experi ents. Essay on principles. Section 9. *Statistical Science*, 5(4):465–480, 1923.

[Otte, 1981] R. Otte. A critque of suppes' theory of probabilistic causality. *Synthese*, 48:167–189, 1981.

[Pearl, 1994] J. Pearl. A probabilistic calculus of actions. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence 10*, pages 454–462. Morgan Kauf ann, San Mateo, CA, 1994.

[Pearl, 1995a] J. Pearl. Causal diagra s for e pirical research. *Biometrika*, 82(4):669–710, Dece ber 1995.

[Pearl, 1995b] J. Pearl. Causal inference fro indirect experi ents. *Artificial Intelligence in Medicine*, 7(6):561–582, 1995.

[Pearl, 2000a] J. Pearl. *Causality: Models, Reasoning, and Inference*. Ca - bridge University Press, New York, 2000.

[Pearl, 2000b] J. Pearl. Co ent on A.P. Dawid's, Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):428–431, June 2000.

[Robins, 1986] J.M. Robins. A new approach to causal inference in ortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.

[Robins, 1987] J.M. Robins. A graphical approach to the identification and esti ation of causal para eters in ortality studies with sustained exposure periods. *Journal of Chronic Diseases*, 40(Suppl 2):139S–161S, 1987.

[Rubin, 1974] D.B. Rubin. Esti ating causal effects of treat ents in rando ized and nonrando ized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

[Si on and Rescher, 1966] H.A. Si on and N. Rescher. Cause and counterfactual. *Philosophy and Science*, 33:323–340, 1966.

[Si on, 1953] H.A. Si on. Causal ordering and identifiability. In W . C. Hood and T.C. Koop ans, editors, *Studies in Econometric Method*, pages 49–74. Wiley and Sons, Inc., 1953.

[Sobel, 1990] M.E. Sobel. Effect analysis and causation in linear structural equation odels. *Psychometrika*, 55(3):495–515, 1990.

[Spirtes *et al.*, 1993] P. Spirtes, C. Gly our, and R. Scheines. *Causation, Prediction, and Search.* Springer-Verlag, New York, 1993.

[Strotz and Wold, 1960] R.H. Strotz and H.O.A. Wold. Recursive versus nonrecursive syste s: An atte pt at synthesis. *Econometrica*, 28:417–427, 1960.

[Suppes, 1970] P. Suppes. *A Probabilistic Theory of Causality.* North-Holland Publishing Co., A sterda , 1970.

[Wright, 1921] S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–585, 1921.