Probabilistic Causation Without Probability

Paul W. Holland
Educational Testing Service

Program Statistics Research
Technical Report No. 93-29

Research Report No. 93-19

Educational Testing Service
Princeton, New Jersey 08541

March 1993

Copyright © 1993 by Educational Testing Service. All rights reserved.

Abstract

The failure of Hume's 'constant conjunction' to describe apparently causal relations in science and everyday life has led to various 'probabilistic' theories of causation of which (Suppes 1970) is an important example. A formal model that was developed for the analysis of comparative agricultural experiments in the first quarter of this century can be used to give an alternative account of 'probabilistic causality' that does not take a stand as to the stochastic or deterministic nature of the causal connection. The approach has many applications in social, behavioral and medical science. This paper discusses the formal model in detail, applies it to 'probabilistic causation' and compares the resulting theory to Suppes' theory.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Probabilistic Causation Without Probability

Paul W. Holland
June 25, 1992

> . . . in restricting himself to the concept of constant conjunction, Hume was not fair to the use of causal notions in ordinary language and experience. (Suppes 1970, p. 10)

1. Introduction

As Suppes observes, Hume's condition of constant conjunction may not square well with experience. In practice, the billiard ball analogy is not apt in many cases where causal language seems natural. The use of a medical procedure may often give the intended result, but not always. Cigarette smoking may be associated with an increased risk of lung cancer, but not all smokers get lung cancer and there are victims of lung cancer who never smoked. In many cases in everyday life and in the the less 'exact' sciences, when we use causal language the connection between cause and effect is noisy and uncertain--not at all like the crisp click of one billiard ball striking the next. This state of affairs has led to various theories of probabilistic causation, e.g., (Good 1961, 1962), (Granger 1969), (Suppes 1970), also reviewed in (Skyrms 1988).

In this paper I show how a formal model developed early in this century by statisticians to provide a detailed analysis of the structure of 'comparative experiments' leads very naturally to a 'probability-free' account of probabilistic causation that is quite different from these other theories but which can accomplish many of their goals. My intent is to make this material accessible to philosophers of science. Here, I give an uncluttered description of the formal

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

model of experiments in the simplest case and then use it to develop a theory of probabilistic causation in which 'probability' plays a limited role. I then compare this approach to Suppes' probabilistic theory of causality to illustrate their differences.

I believe that the formal model for experiments has many other potential uses in philosophical analyses of causation and related topics. I concentrate on probabilistic causality here because of Suppes' work on this subject and in the hope that such an introduction might catch the interest of other philosophers of science. Indeed, if the review in (Tiles 1992) is an indication of the role that 'experiments' currently play in philosophical discussions, then the time may be ripe to introduce a useful formal model into the debates. For discussions of the relation of the formal model to 'causal modeling' in the social sciences and to the views of various philosophers see (Sobel 1990, 1992).

After covering some basic terminology, I then describe the formal model in some detail before relating it to probabilistic causation and the theory of Suppes.

The modern statistical theory of comparative experiments was developed during the first quarter of this century in response to the needs of agricultural experimentation, and quickly spread to many other disciplines. Key to the failure of older experimental methods was the attempt to take them out of the lab and apply them to problems where there is a substantial amount of 'uncontrollable variation' in the material being studied (such as the effects of the weather in agricultural field experiments). The new statistical methods deal explicitly with uncontrolled variation, and, thus, breathed new life into experimental research.

Basic terminology: The three key terms used to describe experiments are treatments, units, and outcomes. In a comparative experiment each experimental unit is exposed to one element of a set of treatments and then, some time after this exposure to a treatment, an outcome is observed for each unit. Comparative experiments get their name from the fact that they are used to compare the outcomes of treatments on units.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For example, in a medical experiment (often called a clinical trial) the units are patients of various types and the treatments are new (and old) drug therapies, or various surgical or other medical procedures (including various combinations of medical procedures). In a medical experiment, there are a variety of outcomes of potential interest, e.g., survival status for five years after the therapy, or degree of improvement in some physical activity or in some measure of health. In an educational experiment, the units might be classrooms of elementary school children (instead of individual students), the treatments could be various educational programs for teaching mathematics to whole classrooms of students, and the outcomes might be classroom averages on tests of performance in certain areas of mathematics. (This example illustrates that the units are part of the definition of the experiment and not inherently individual objects or people).

The process of assigning each unit to one of the treatments is called the assignment process of the experiment. Random assignment employs a randomizing mechanism to assign the treatments to the units. Historically, randomization was not the first assignment process used. Systematic assignment processes were used in the earliest agricultural experiments in which, for example, alternating contiguous plots of land (the units) were given different amounts of fertilizer (the treatments). Random assignment is regarded by many practitioners as a great invention in experimental science when there is uncontrolled variation in the material under study. Some reasons for this are discussed in section 2.

R. A. Fisher, who made many contributions to the development of modern statistics and especially to the application of randomized comparative experiments in agriculture and biological science, identified three important principles to be observed in the design of any comparative experiment. These are randomization, replication, and blocking. I have already described randomization. Replication means that more than one unit should be exposed to

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

each treatment. (This can involve many technical issues and further discussion of it is beyond the scope of this paper.) In blocking, we use observed information about the units to group them into homogeneous subsets (the blocks) and then assign the treatments to units within each of these blocks. For example, we might group all the men in one block and all the women in another block. In a complete, randomized block design there are as many units in each block as there are treatments under study and each unit within a block is randomly assigned a treatment that is different from those assigned to any other unit in that block. For example, in comparing the wear on two types of shoe leather, the 'units' might be feet and the 'blocks' people. A person would then have the two types of shoe leather randomly assigned to his or her left and right soles.

The idea behind blocking is adapted from those branches of experimental science in which it is possible to take seriously the idea of distinct-but-identical samples of experimental material for study (the units). For example, in a chemistry laboratory it is often possible to create samples of chemicals that are very nearly identical in purity, volume, mass, etc. When distinct-but-identical units are available for study, the effects of treatments can be studied easily by assigning the different samples to different treatments.

The idea of distinct-but-identical units arises later in my discussion. The point of mentioning it here is to forewarn the reader that although this notion may be a helpful crutch for thinking about experimentation, and it even has a role to play in modern experimental design through blocking, the idea of distinct-but-identical units has no essential role in the theory discussed here, even though it can be expressed within the model as the special case of 'unit homogeneity'.

With this dash through the terminology of comparative experiments I now turn to the formal model that is used to describe them. This model is implicit in the earliest work on comparative experiments, described in (Fisher 1925). It

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was first explicitly developed to analyze the behavior of statistical procedures in Neyman's 1923 doctoral dissertation in Polish, translated and republished posthumously in (Neyman 1990), and was first published in English in an influential paper, (Neyman 1935). It is standard material in those books on experimental design that go into the details of 'permutation' or 'randomization' tests, such as (Kempthorne 1952) or (Cox 1958). However, in all of these early references it is presented purely as a technical tool and applied only to randomized experiments. Its importance for the analysis of non-experimental studies and to causal inference more generally was first pointed out in (Rubin 1974) and was developed extensively in (Rubin 1977, 1978). In (Holland 1986) I call it 'Rubin's model for causal inference' and emphasize its importance as a general tool for studying the relationship between statistics and causal inference. Here I will call it simply 'the model for experiments'.

2. **The Formal Model**

At first glance, the model does not look like the types of things that are usually called 'statistical models'. In particular, at least to begin with, there is no role for probability. This is because this model is intended to be applicable to all types of comparative experiments, even those 'deterministic' cases in which the variation across the units can be well controlled.

We need notation for the three basic elements: units, treatments, and outcomes. The units are indexed by the letter i, i.e., 'unit i', and i ranges over a set (or population) of units, U. The treatments are indicated by labels denoted by the letter t, and t ranges over a (finite) set of possible treatments, T. T must contain at least two treatments or there is nothing to compare (i.e., 'experiments' with only one treatment are not comparative experiments in the sense used here and might better be called 'measurements' or 'controlled observations'). The case of an infinite number of treatments is possible--e.g., doses of a drug--but is not discussed here. Later, I simplify to the case where T

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

consists of only two treatments, t the new treatment, and c the control treatment.

So far so good. Now comes the hard part in the sense that we need to define quantities that do not 'exist' according to some critics of this approach, e.g., (Lund 1991).

In a comparative experiment, each unit could be exposed to any one of the treatments in T and the outcome resulting from that exposure could then be observed. Thus, before setting up notation for the outcome that is actually observed for unit i , I add to U and T notation for all the potential outcomes that could be observed for unit i (also depending on potential exposure to treatment t in T). Thus, for every pair *(i, t),* combining a unit and a treatment, define Y_{it} as

Y_{it} = the outcome that would be observed for unit i if it were exposed to treatment t.

I assume that Y_{it} is a real number for each combination of i and t. This is not essential, but it makes things go smoothly and agrees with standard discussions of experimental design. Thus, Y is a real-valued function defined on the Cartesian product, $U \times T$.

Statisticians are interested in data, but so far there is no data or at least no notation for it. We need to remedy this. It takes two steps. First we need notation for the treatment to which unit i is actually exposed. This is denoted by x_i ,

x_i = the treatment to which unit i is exposed in the experiment.

Formally, x is a mapping from U to T. This gives us part of the data.

Next, we need to define the outcome that is actually observed for unit i in terms of what has been defined so far. We denote the outcome observed for unit

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

i by y_i and define y_i in terms of Y_{it} and x_i as follows,

$$y_i = Y_{i x_i}, \qquad (1)$$

so that y_i is equal to Y_{it} if and only if $x_i = t$. Hence, the data that is observed in this comparative experiment is $\{(y_i, x_i)$, for i in $U\}$. I use the notational device of denoting the observed data, (y_i, x_i), by lower case letters and potential observations, Y_{it}, by upper case letters.

Finally, we need to define the relative effects of the treatments on each unit. These are called the unit-level causal effects, or just the causal effects for short. For unit i, the causal effect on the outcome Y of treatment t_1 relative to treatment t_2 is the difference

$$Y_{it_1} - Y_{it_2}. \qquad (2)$$

In the earlier statistical literature, the causal effects are called 'treatment effects'. The term 'causal effect' was introduced in (Rubin 1974) to emphasize that they are the basic quantities of interest in all problems of causal inference, and I follow his terminology. Thus in terms of this model, causal inference means inferring something about the causal effects.

In summary, the formal model is the quadruple (U, T, Y, x) where U and T are (finite) sets, Y is a real-valued mapping from $U \times T$ and x is a mapping from U to T. The observed response, y_i, is the real-valued mapping from U defined by (1). The causal effects are the quantities defined in (2) and they are the items of ultimate interest in causal inference.

Having set up this formal model and these definitions, I now mention several important points.

First, because it is impossible to apply two different treatments at the same time to a single unit, it is impossible in principle to observe both Y_{it_1} and Y_{it_2}. Thus, the unit-level causal effects cannot be directly observed. I call this the

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fundamental Problem of Causal Inference because if we could observe both Y_{it_1} and Y_{it_2} for all i, t_1, and t_2 then we would know the relative effects of all the treatments on each unit, that is we could infer everything about the causal effects.

Second, the causal parameters, i.e., the causal effects, are defined independently of the way that the data are collected, i.e., the way that x is constructed. This allows a complete separation between the targets of our inferences (the causal effects) and our ability to make sound inferences about them (the experimental design). This separation is very important and not always made.

Third, it is important to separate the model, *(U, T, Y, x),* from its application to a real situation. The model is an ideal case. Applying it to a real situation requires identifying the elements of the model with corresponding elements of the real situation. This correspondence may be imperfect in various ways (see points four and five, below). Deductive analysis takes place within the model. To the extent that the correspondence between the model and the real situation is faithful, these deductions can then carry over to the real situation. I give some examples of these deductions in sections 2, 3 and 4. Criticisms of the model have to do with its logical structure or with the logic of the deductions made from it. Criticisms of an application of the model have to do with the faithfulness of the correspondence between the elements of the model and those of the real situation. It is useful to distinguish between these two types of criticisms.

Fourth, an important assumption, that is implicit in any application of the model to a real situation, goes by a variety of names including 'non-interfering' units. When units interfere with each other, the outcome for one unit can depend not only on the treatment that the unit is exposed to, but also on the treatments received by some or all of the other units. For example, if the units are people in a theatre watching a play, the treatments are to stand up or to sit down, and the

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

outcome is the clarity of one's view of center stage, then the outcome for me depends on what I choose to do (my own self-imposed treatment) but also on what all those in front of me also choose to do (their self-imposed treatments). Interfering economic units are the sine qua non of micro-economic analysis. Interfering units can occur in agricultural experiments when the rain causes runoff of fertilizer from one plot of land to the next. When units interfere with each other, a more complex model is required than the one discussed here, but many of the same ideas are useful in that case, too. This point is discussed more extensively in (Rubin 1980, 1986, and 1990b) under the topic he calls the stable-unit-treatment-value assumption, or SUTVA.

Fifth, there are two assumptions about the set T and the actual 'treatments' in a real experiment that should be mentioned. The first is that the elements of T are sufficient to distinguish between the treatments that are actually applied in the experiment. This can become a serious issue. For example, if T contains only two elements, say t (the new treatment) and c (the control treatment) and in actual fact there are several versions of t and of c used in the experiment, then classifying them all as either t or c may be misleading. In (Neyman 1935), 'errors of experimental technique' are discussed in which treatments labeled the same way by elements of T are not really exactly the 'same' treatment. This leads to a more complex model than the one discussed here. A second assumption about T is that it is actually possible in the real experiment to expose any unit in U to any treatment in T. This assumption often creates apparent conceptual difficulties in applying this model to non-experimental studies where there is no real control over which units are exposed to which treatments.

Sixth, in medical and other types of studies that involve human subjects, ethical questions arise in which it may be ethically inappropriate to expose some subjects to some of the treatments in T. Some patients might be too sick to undergo surgery, but could undergo some form of drug therapy. Or, a physician

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

may believe that a given patient would benefit more from treatment t than from treatment c. When this happens, the mapping x is restricted. This in turn can restrict the types of conclusions that are possible. The question of ethics in medical research can force a trade-off between the treatments a patient can receive and the type of conclusions that can be reached by the study. It is exacerbated when the physician does not really know which treatment the patient ought to receive (because this requires a solution to the fundamental problem of causal inference) but has strong opinions about the matter anyway. (Kadane and Seidenfeld 1990) discuss methods for designing clinical trials that allow some 'ethical' restrictions on x.

Finally, and for my purposes here, most importantly, the model assumes nothing about how t acts on i to result in the value Y_{it}. The only assumption made in using the model is that if i were exposed to t, then some value of the outcome, denoted by Y_{it}, would be observed, i.e., we know exactly how to ascertain the value of Y_{it}. The relationship between i, t and Y is that of a mathematical function (i.e., Y is a function of i and t), but without any specification of the form of this functional relationship. This might appear to be a bit unstatistical in the sense that there is no assumption that there is a stochastic component to the relationship between the pair (i, t) and Y, i.e., at this stage Y_{it} is not represented by a probability distribution and is not a random variable. There are two reasons for this lack of probabilism. The first is that it is unnecessary--Y_{it} simply denotes a value that would be observed under some conditions of the experiment. It does not need to be thought of as having any sort of random or stochastic element in its generation. In fact, we are entirely agnostic about how the value, Y_{it}, comes about. The second is that probability considerations come later in this development for very specific purposes. We will get to probability, but not yet.

Connections to causation: At this point, it may be helpful to connect this formal model to ordinary causal talk. The treatments, indexed by the elements of

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

T , are the causes, and the effects are the relative causal effects defined in (2). The effect of a cause is, therefore, always relative to another cause. Outcomes are not effects, but are intermediate to the definition of effects. Sometimes the distinction between outcomes and effects is blurred when the cause is being compared implicitly rather than explicitly to another treatment (e.g., "my headache went away because I took an aspirin an hour ago"). I think that the confusion of outcomes with effects is a source of many problems in the analysis of causation.

The 'mechanisms' that lead from cause to effect play no role in this discussion and for this reason I argue that this theory is about measuring the effects of causes rather than about identifying the causes of effects or the mechanisms that lead from cause to effect. By taking units, causes and outcomes as the primitive, undefined terms of the model and defining effects and the observed data in terms of them, this approach breaks with the philosophical tradition of taking events as the primitives and attempting to explicate the nature of causation in terms of them.

The notation for the potential observations, Y_{it}, is related to Hume's conditions of 'temporal succession' and of 'spatial and temporal contiguity'. In indicating that Y is a function of both i and t, the unit i becomes the point of spatial (and, to some extent, temporal) contiguity between t and Y--the treatment acts on the unit and the outcome is measured on it as well. Similarly, the notation, Y_{it}, indicates that i and t combine in some way to produce the value of Y thus insuring the temporal succession of the cause, t, and the outcome, Y. The model for experiments is consistent in many ways with Hume's classic analysis of causation, but it does not insist on 'constant conjunction'--see section 3.

Causal models versus experimental design: In terms of the quadruple

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(U, T, Y, x), a causal model is any restriction or assumption that we make about the function Y, whereas, assumptions about U, T and x are parts of the experimental design. A reoccurring theme in the development of statistical methods for experiments is the replacement of causal models (i.e., assumptions) by elements of experimental design (i.e., actions under the control of the experimenter). It should be noted that causal models are assumptions that cannot always be tested directly using the data from the experiment, i.e., $\{(y_i, x_i)$ for i in $U\}$. However, these assumptions allow us to draw conclusions about Y from the data. This gives us information about the unobserved values of Y and this information, in turn, can allow us to describe the causal effects, i.e., allow us to make causal inferences. Two simple causal models arise often enough that I will discuss them briefly, next.

Unit homogeneity: It is sometimes useful to make the assumption that the value of Y_{it} does not depend on the unit i, i.e., the units are homogeneous in their reactions to the various treatments in T. This may be a believable assumption in those branches of science in which it is possible to prepare distinct-but-identical units, i.e., 'identical' samples of material for study in the experiment. There is always a pragmatic element in regarding two distinct things as being 'identical'. They are not identical in all respects, but only in some respects. We might be willing to assume that all electrons are identical, but that is because 'electrons' are one consequence of a long effort to find the fundamental constituents of matter that are, by definition, identical in some ways. It is wise to remember just how tentative such notions of 'identical' are in the history of physical science. Certainly when we turn to the social and biological sciences, the idea of distinct-but-identical units is a much less useful idea. In these branches of science, units differ in many ways only some of which we know about. This is the 'uncontrolled variation' mentioned above.

Regardless of its general tenability, if one can assume that the units are homogeneous then this assumption may be expressed in terms of Y as:

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$Y_{it} = Y_t, \quad \text{for all } t \text{ in } T. \qquad (3)$$

Under the assumption of unit homogeneity in (3), the fundamental problem of causal inference mentioned earlier is easily avoided. We need only expose unit 1 to treatment t_1, unit 2 to treatment t_2, etc., until we have exhausted all the treatments in T and have, therefore, observed all the values that Y can take on for any unit. Once this is done, we can calculate all of the causal effects defined in (2) for any pair of treatments. Unit homogeneity neatly eliminates the fundamental problem of causal inference. It is one reason why the older lab-based experimental methods worked well on material not subject to much uncontrolled variation.

The assumption of unit homogeneity is sometimes disguised under the rubric of ceteris paribus or 'other things being equal'. Ceteris paribus is usually evoked for ideal (and unrealizable) experiments in which we vary only the factor whose effect is of interest and somehow hold all other relevant factors fixed. We get back immediately to the notion of distinct-but-identical units here because these other factors are just ways in which the units can be non-identical. The problem is that we can not ever really know what all the relevant factors are in any real experiment, so the best we can do is hold fixed all the factors that we do know about and which we can fix. I argue that in the experiments in which unit homogeneity might hold, it does not really matter that the units 'appear' to be identical. The crucial assumption is that the outcomes do not depend on which unit is exposed to which treatment. The apparent similarity of units in terms of their 'properties' and 'other factors' is irrelevant except insofar as it makes plausible the assumption of the homogeneity of the outcomes on distinct units, i.e., that equation (3) holds. In practice, the plausibility (and therefore the utility) of the assumption of unit homogeneity is always a matter of degree and it can rarely be expected to hold exactly. Even when it is only approximately true,

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

it is a powerful inferential device.

Fisher's Null Hypothesis: Suppose we are willing to entertain the hypothesis that none of the treatments have effects. For example, in an early article in which R. A. Fisher set down some principles of design for field experiments in agriculture, (Fisher 1926), he makes the following statement about the apparent productive value of treating a given acre of a field with manure:

> What reason is there to think that, even if no manure had been applied, the acre which actually received it would not still have given the higher yield? (p. 504)

This type of statement can be translated into another simple causal model known as Fisher's Null Hypothesis:

$$Y_{it} = Y_i, \text{ for all } i \text{ in } U, \qquad (4)$$

that is, Y_{it} does not depend on the treatment, t, in T that is applied. Note that (4) also avoids the fundamental problem of causal inference in the sense that if (4) were true, then having observed Y_{it_1} for unit i and treatment t_1, we automatically know the value of Y_{it} for all t in T. In this case, all the unit-level causal effects defined in (2) are zero. This is why the hypothesis is called 'null'--the treatments have no effect on any unit. But note that the units may exhibit any amount of 'uncontrolled variation' because the Y_i in (4) may depend, in an arbitrary way, on i.

The purpose of Fisher's null hypothesis is quite different from the assumption of unit homogeneity even though the two are formally quite similar--compare (3) to (4). Unit homogeneity can be used to estimate unknown causal effects using the data, whereas Fisher's null hypothesis is that all of these causal effects are a particular known value--namely zero. If the assignment mechanism is of a particular type--randomization--then Fisher's null hypothesis has consequences

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for the observed data and we may be able to conclude that these consequences are incompatible with the data. This brings us to the well-known, and very large, subject of testing statistical hypotheses and it would take me too far afield to say more about it here, so I won't.

Random assignment: The function x can be determined in a variety of ways. One very important aspect of comparative experiments is that the determination of x is under the control of the experimenter; this is what 'experimental manipulation' means within this model. If we use appropriately a randomizing device such as the toss of a coin, the roll of a die, or a table of random numbers to decide which treatment is applied to each unit, then x is said to be constructed by randomization. I do not want to go into the fine points of the many ways that randomization can be used to construct x. My purpose at this point is merely to indicate how randomization fits into the overall model of experimentation described here.

Probability: It is now time to get serious about probability. I use probability in two ways. First, probability has already entered the discussion through the construction of x via randomization. This is, I believe, a fairly insignificant use of probability in this theory, though it has very significant consequences as we shall see. Randomization is a physical act in which a known chance mechanism is used in particular ways to construct the function x. When U is large, this results in an x with certain known properties that I will discuss in more detail in a moment.

Second, and most importantly, I also use the language of probability to describe the population U in terms of x and Y. In particular, I use the notation of expected values and conditional expected values to denote averages of the values of Y over subsets of the units in U. For example, the notation

$$E(Y_t)$$

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

denotes the average of Y_{it} across all the units in U, and

$$E(\, y \mid x = t \,)$$

denotes the average value of y_i among all those units in U for which $x_i = t$ (the units exposed to treatment t). Note that in this expectation notation, I have suppressed the index i because i is being 'averaged over'. The average value of y_i among all those units exposed to treatment c is denoted by $E(\, y \mid x = c \,)$. (For simplicity of notation I have just slipped into the two treatment case of a treatment t and a control treatment c that I will continue to use throughout the rest of this paper.)

My use of the expectation notation is, in fact, quite standard because expectations are population averages and I average over the population U. While it does not come up until section 3, I also use the standard probability notation to refer to the proportions of units in U that are in subsets of U defined by x and Y in various ways. Thus, the probability that I use here has a simple frequentist interpretation in terms of proportions of units in U. 'Probability' in this model is simply another name for 'unit inhomogeneity'.

ACES and FACES: When unit homogeneity is implausible, the fundamental problem of causal inference must be addressed in other ways and this is the great contribution of the introduction of randomization to the design of comparative experiments.

We begin by changing our target and considering causal parameters other than the basic unit-level causal effects in (2). This leads easily to the idea of an **A**verage **C**ausal **E**ffect or ACE. There are several ways to average the unit-level causal effects to get interesting causal parameters. I describe only one of them. Set

$$\text{ACE} = E(\, Y_t - Y_c \mid x = t \,). \qquad (5)$$

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The ACE in (5) is the average of the causal effects of t relative to c across the units exposed to treatment t. It can be thought of as a measure of the 'effect of the treatment on the treated units'. It is worthwhile to change (5) into a difference of two expectations, i.e.,

$$\text{ACE} = E(Y_t \mid x = t) - E(Y_c \mid x = t). \qquad (6)$$

The first expectation in (6) is easy to interpret. It is the average value of Y_t over all the units exposed to t, i.e.,

$$E(Y_t \mid x = t) = E(y \mid x = t), \qquad (7)$$

and the expectation on the right-hand-side of (7) involves only the data $\{(y_i, x_i)\}$ and can, therefore, be computed from it. But, the second expectation in (6) is counterfactual in the sense that it is the average of the values of Y_c across all the units exposed to t (and for which the value of Y_c is inherently unknown due to the fundamental problem of causal inference). I have heard the counterfactual expectation, $E(Y_c \mid x = t)$, referred to as an 'oxymoron', but I view it as a mathematical formula for the single most important quantity that arises in causal inference. Without introducing it or equivalent notation there is simply no way to describe precisely many important causal parameters that arise in practice.

Before continuing, I will define the FACE or Prima Facie Average Causal Effect that is associated with the experiment:

$$\text{FACE} = E(y \mid x = t) - E(y \mid x = c). \qquad (8)$$

The FACE is the average of the observed outcomes for the units exposed to t minus the average of the observed outcomes for the units exposed to c. It can be put into a form that more closely resembles the ACE by using Y instead of

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

y:

$$\text{FACE} = E(Y_t \mid x = t) - E(Y_c \mid x = c). \qquad (9)$$

Neither of the two expectations in (9) are counterfactual in the sense used above because both of them are based on the data, $\{(y_i, x_i)\}$. The FACE is a measure of the association of y and x across the units in U. It is the 'correlation' in 'correlation is not causation'. The ACE is a causal parameter, because it is determined by the causal effects, and represents the 'causation' in the time-honored, correlation/causation distinction.

This brings us to a basic question of causal inference. When does the FACE (which we can always calculate from the data) equal the ACE (which is a causal parameter and which involves an inherently unobservable counterfactual conditional expectation)? Setting (9) equal to (6) yields the following fundamental condition:

$$E(Y_c \mid x = t) = E(Y_c \mid x = c). \qquad (10)$$

Equation (10) is a statement of the lack of dependence between the variable Y_c and the variable x across the units in U. As such it holds if Y_c and x are statistically independent across the units in U. This brings us back to the process of constructing x using randomization.

A variety of theorems in probability theory have the following form: if U is countably infinite and x is constructed by randomization (including, but not restricted to, simple coin flipping) then x is independent of Y_c across the units in U, for almost all realizations of the randomizing process. This means that when U is a large, but finite, population, randomization insures that (10) is approximately true no matter how Y_c varies across the units, and the larger U is, the more exactly this approximation holds. This fits well with the practitioner's intuition that randomization has the greatest force when the

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

number of units being randomized to treatments is large. The practical matter of how close the approximation holds is solved by the usual statistical tools of standard errors, etc., and is an important, but merely technical, detail from the point of view described here.

The effect of randomization can be described in a slightly different but related way. Table I displays a simple hypothetical example of the model in the case where Y is dichotomous (e.g., 1 = the unit lived, 0 = the unit died) and there are only two treatments, t and c. Table I lists a few of the units of U and their corresponding values for all the quantities that arise in the model. Unit 1 lives ($Y_{1t} = 1$) if it is exposed to t but dies ($Y_{1c} = 0$) if it is exposed instead to c. However, for unit 1 $x_1 = t$ and, therefore, unit 1 is observed to live, i.e., $y_1 = 1$. Similarly for the other units listed in Table I.

(Insert Table I about here)

Randomization can be thought of as producing two simultaneous random samples, one from the Y_t column and one from the Y_c column in Table I. From this perspective, the average y-value for the units exposed to c is an estimate of the column average of the Y_c column in Table I (i.e., $E(Y_c)$). However, randomization also makes the (unobserved) Y_c values of the units exposed to t a random sample from the Y_c column of Table I and hence their average value (i.e., the counterfactual expectation in (6)) also estimates the average of the entire Y_c column of Table I. From this point of view, equation (10) is approximately true under randomization because both sides are estimates of the same quantity, i.e., $E(Y_c)$, and these estimates become more accurate the larger U is.

Thus, randomization is a powerful technique. When the number of units in the study is large, it makes an inherently unobservable causal parameter (the ACE)

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

approximately equal to a quantity that is easy and natural to compute from the observed data (the FACE). Under randomization, equation (10) holds approximately so that 'correlation' and 'causation' essentially become one.

Note also that the use of randomization to estimate the ACE does not depend on assumptions about any causal model for Y (as unit homogeneity does). To reiterate a statement made earlier, one of the contributions of statistical methods for making causal inferences in experiments is to replace untestable causal models (such as unit homogeneity) by features of the experimental design (such as randomization) that are under the control of the experimenter. This is not done without cost, of course. The cost is that we are no longer making inferences about the individual unit-level causal effects in (2). The inferences are now about the ACE's which are averages of the unit-level causal effects.

Averages have both strengths and weaknesses. Their biggest weakness is that an average by itself says little about the individual values that make it up. The patient and his or her physician are interested in the patient's own causal effect, but a research study will often result in information about the ACE of which the patient's causal effect forms only a tiny part. For many problems, however, averages are very useful. In public health, the ACE is often the main causal parameter of interest.

I argue that unit homogeneity (for simple 'deterministic' phenomena) and randomization (for phenomena exhibiting uncontrolled variation) are the twin inferential pillars of experimental science because, in different ways, each can be used to solve the fundamental problem of causal inference. Neither is guaranteed to succeed, of course--unit homogeneity may not be plausible in many situations, randomization is impossible to implement in many studies, and the ACE may not be a relevant parameter in some cases.

Non-experimental studies: 'Observational studies' or 'non-experimental studies' are statistical terms-of-art describing research where the goal is to

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

estimate the causal effects of treatments but for which unit homogeneity is implausible and randomization is infeasible. The active experimenter is replaced by the passive observer who tries to draw causal conclusions without having much control over the data generation process and in the face of significant unit inhomogeneity. For example, a causal finding based exclusively on non-experimental research is the link between cigarette smoking and lung cancer.

Detailed discussion of these important types of studies goes beyond the scope of this paper. I simply make a few points about them here, and refer the interested reader to the references mentioned at the end of section 4 for more details.

First, for reasons that have never been clear to me, many of the examples used in philosophical discussions of causation and related topics are based on observational-study-like examples rather than on comparative experiments. It seems to me that giving an account of causation in the context of an experiment where, in practice, there is confidence in the conclusions is logically prior to attempting to discuss causation in more complex situations where, again in practice, there is always less confidence in the conclusions. That is why I have emphasized experiments in this discussion.

Second, in observational studies, the identification of the units and the outcomes is usually unambiguous. However, it is sometimes difficult to determine what treatment was applied to each unit and whether or not they were applied prior to the determination of the outcomes (in which case Hume's 'temporal succession' becomes a non-trivial issue).

Third, the fundamental problem of causal inference and the consequent a priori potential inequality between the ACE and the FACE poses a major problem for observational studies. In general, in an observational study there is no assurance that equation (10) holds. To proceed with their analyses, one must make assumptions that are closely related to equation (10) but without any real basis for these assumption in the design of the study (such as provided by

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

randomization). An example of such an assumption is 'strong ignorability' discussed in (Rosenbaum 1984a). All is not hopeless, of course, but the measurement of causal effects in non-experimental studies always has an additional level of uncertainty in it beyond that determined by sample size. This additional uncertainty is due to our natural lack of conviction in the validity of untested and often untestable assumptions about the assignment of units to treatments. A key tool in the analysis of observational studies is 'sensitivity analysis' in which we express this additional uncertainty by recomputing estimates of the causal effects under a plausible variety of untestable alternatives to equation (10).

Finally, in an observational study, the question of 'what can be a cause' arises. In an experiment, the causes are the treatments. In observational studies the notion of an experimental manipulation can become confusing. Consequently, estimates of 'causal effects' are sometimes sought in cases where there is ambiguity in what they mean, e.g., the 'causal effect' of gender on test performance, or the effect of 'nature' versus that of 'nurture'.

3. Probabilistic Causation.

The whole purpose of the formal model for comparative experiments is to accommodate 'uncontrollable variation', so it should be no surprise to find that it also gives rise to a type of probabilistic theory of causality that addresses many of the same issues that concern Suppes. I first describe this theory and then discuss its relation to Suppes' work.

Probabilistic causation without probability: To make things simple, let us assume that Y is a 0/1 variable so that we can talk about events occurring or not in terms of the values of Y. For example, suppose the units are patients in a medical study and $Y = 0$ denotes death within some specific future time period while $Y = 1$ denotes survival over this time period. Suppose further that we have two treatments, t and c, with t denoting a new treatment and c a

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

control treatment, either of which can be applied to each patient.

We can conceive of grouping the units in U into four sets depending on the joint values of Y_t and Y_c; we can not actually form these groups in practice, of course, due to the fundamental problem of causal inference, but it is useful to discuss them. The four subsets of U are

$$A_{11} = \{i : Y_{it} = 1 \text{ and } Y_{ic} = 1\},$$

$$A_{10} = \{i : Y_{it} = 1 \text{ and } Y_{ic} = 0\},$$

$$A_{01} = \{i : Y_{it} = 0 \text{ and } Y_{ic} = 1\},$$

and,

$$A_{00} = \{i : Y_{it} = 0 \text{ and } Y_{ic} = 0\}.$$

Then let $P(1, 1)$, $P(1, 0)$, $P(0, 1)$ and $P(0, 0)$ denote the proportions of units that are in these four subsets of U. These four proportions (i.e., probabilities) can be described as follows:

$P(1, 1)$ = the probability that a patient lives throughout the period no matter which of the two treatments he or she receives,

$P(0, 0)$ = the probability that a patient dies within the period no matter which of the two treatments he or she receives,

$P(1, 0)$ = the probability that the new treatment causes a patient to live during the period,

$P(0, 1)$ = the probability that the new treatment causes a patient to die during the period.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The last two 'causal' probabilities address constant conjunction in a sense that is quite compatible with Hume's intent. For example, if

$$P(1, 0) = 1,$$

then the presence of the cause t is always followed by survival, and the absence of the cause t is never followed by survival. Non-constant conjunction or 'probabilistic causality' occurs when either $P(1, 0)$ or $P(0, 1)$ is strictly between 0 and 1. In other words, in terms of this model, probabilistic causality is identical to unit inhomogeneity and the rejection of Fisher's null hypothesis.

Thus, the model for experiments makes it easy to talk about probabilistic causality. The only price we pay for this is that we must refer to sets such as A_{10} that can never be formed in practice. This is closely related to the counterfactual language that has been used to describe the causal relation, for example (Lewis 1973).

This price may be too large for an extreme Positivist, but it is interesting to observe that a combination of randomization and a causal model can allow us, in certain circumstances, to draw plausible inferences about these four sets that we can not actually form in practice. Here is how this analysis goes.

'Observing' the unobservable: First of all, it is useful to express the joint distribution of Y_t and Y_c (over U) as a two-by-two table, as in Table II.

(Insert Table II about here)

The idea is that if U is large, then under randomization we can estimate the values of the margins of this joint distribution, i.e., $Prob\{Y_t = 1\}$ and $Prob\{Y_c = 1\}$, and then we use a causal model to allow us to infer the values of the 'insides' of this two-by-two table (i.e., the $P(i, j)$'s $i, j = 0,1$) from its margins.

The discussion in section 2 shows that we can use the FACE to estimate the ACE defined in (5). However, under randomization x is approximately independent of both Y_t and Y_c across U, so it follows that we can drop the x

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in the conditioning part of the definition of the ACE and hence,

$$\text{ACE} = E(Y_t) - E(Y_c). \qquad (11)$$

However, because Y is a 0/1 variable, we have the following basic fact from Table II:

$$E(Y_t) = Prob\{Y_t = 1\} = P(1, 1) + P(1, 0), \qquad (12)$$

and,

$$E(Y_c) = Prob\{Y_c = 1\} = P(1, 1) + P(0, 1). \qquad (13)$$

Hence, combining (11), (12) and (13) we can express the ACE in terms of the 'causal' probabilities $P(1, 0)$ and $P(0, 1)$, i.e.,

$$\text{ACE} = P(1, 0) - P(0, 1). \qquad (14)$$

Equation (14) is the contribution of randomization to the analysis.

Now suppose that from some a priori consideration it is plausible that the treatment does not cause death in the sense that we regard it as impossible for exposure to t to lead to death but exposure to c not to, i.e., $Y_{it} = 0$ but $Y_{ic} = 1$. This might be a plausible assumption if t were an improved version of c or if c involves a hazard that could lead to death and t does not involve this hazard. In Table I, this assumption would eliminate such units as unit 5. The assumption can be formulated in the model as

$$P(0, 1) = 0. \qquad (15)$$

Equation (15) is the contribution of the causal model to the analysis. From (14) and (15) we have

$$\text{ACE} = P(1, 0). \qquad (16)$$

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Therefore, if we assume that the causal model (15) holds and include random assignment in the experiment's design, then the FACE is an estimate of the ACE and, therefore, of *P(1, 0)*. Thus, while we cannot actually form the set A_{10} we can still learn about its frequency in the population *U*, (a) using a causal model that may be plausible from other considerations, and (b) including randomization in the experimental design. We can then find all four of the *P*'s in Table II.

Finally, I should emphasize that the probability, *P(1, 0)*, that *t* causes survival should not be confused with the probability involved in some stochastic mechanism that might connect *t* to survival. Such a mechanism might or might not be the reason why *P(1, 0)* does not equal unity, i.e., why there is unit inhomogeneity. (Indeed, a stochastic mechanism might be proposed to account for the value of *P(1, 0)* in some population, *U*.) This analysis and the model on which it is based is entirely silent about that possibility. All of the 'probability' in this analysis arises from unit inhomogeneity. The source of unit inhomogeneity may be an underlying stochastic mechanism 'within' each unit (stochastic units), or it may be deterministic differences between the units that are not reflected in the data that has been collected about them prior to their exposure to the treatments. Again, I argue that this agnosticism about stochastic versus deterministic sources of unit inhomogeneity is irrelevant to this theory because it is about measuring the effects of causes and not about identifying causes or the mechanisms that lead from cause to effect.

Suppes' theory of probabilistic causation: Suppes intends his theory to accommodate the possible lack of constant conjunction of cause and effect while retaining temporal succession from Hume's analysis of causation. As far as I can tell, spatial and temporal contiguity play no direct role in Suppes' analysis.

Roughly speaking, Suppes proposes that one event is the cause of another

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that occurs later in time if the occurrence of the first event increases the probability of the occurrence of the second, and there is no third, temporally prior, event that we can use to 'factor out' this 'probabilistic association' between the first and second events.

Suppes' theory is much more general than the one described above. He uses the language of probability theory for events that occur in time (stochastic processes) but is not specific about the probability space or the nature of the events beyond their temporal order of occurrence. The essence of his theory is to identify when one event is a genuine (probabilistic) cause of another. His three basic definitions of prima facie, spurious and genuine causes are given in S_1, S_2. and S_3.

(S_1) If $r < s$ denote two time values, the event C_r is a prima facie cause of the event E_s if

$$\text{Prob}\{ E_s \mid C_r \} > \text{Prob}\{ E_s \}.$$

(S_2) C_r is a spurious cause of E_s if C_r is a prima facie cause of E_s and for some $q < r < s$ there is an event D_q such that

(a) $\text{Prob}\{ E_s \mid C_r , D_q \} = \text{Prob}\{ E_s \mid D_q \}$, and

(b) $\text{Prob}\{ E_s \mid C_r , D_q \} \geq \text{Prob}\{ E_s \mid C_r \}$.

(S_3) C_r is a genuine cause of E_s if C_r is a prima facie cause of E_s, but C_r is not a spurious cause of E_s.

Condition S_1 says that given that the causal event has occurred, the probability that the effect will occur is higher than it would be had the causal event not occurred. A prima facie cause exhibits a positive probabilistic association with the effect.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Condition S_2 is what Suppes means by a temporally prior event 'factoring out' the association between a prima facie cause and the effect. Part (a) of the condition is the conditional independence of the prima facie cause and the effect given the other event. This is also called 'screening off' (D_q screens off C_r from E_s). Part (b) requires the screening-off event to increase the probability of the effect beyond the value attained by conditioning on the occurrence of the original prima facie cause alone, i.e., given the occurrence of the prima facie cause, the screening-off event must act like a prima facie cause itself.

Condition S_3 says that if a prima facie cause is not spurious in the sense of S_2, then it can be elevated to the status of a genuine cause.

The crucial part of these definitions is part (a) of the definition of a spurious cause. It also appears in many probabilistic analyses in the philosophy of science, e.g., Salmon's definition of statistically irrelevant explanations, (Salmon 1989). For other references to similar uses of 'screening off' see (Skyrms 1988). Part (b) is more problematic and may not be important. In (Holland 1986), I give some evidence that it may be inessential to the notion of cause. Suppes' conditions S_1 and S_2(b) both concern positive association between causes and effects. For the most part, positive versus negative association is an inessential artifact of the labeling of outcomes and treatments.

There is a certain unattainable quality to Suppes' definition of a genuine cause. In order to show that a cause is genuine, we need to be able to show that there is no event D_q that satisfies the condition S_2. In practice there could be an infinite number of potential D_q's, and ruling them all out might be impossible. See (Good 1962) for a different criticism of the use of screening-off events in definitions of probabilistic causality. For a criticism of Suppes' theory see (Otte 1981).

4. A Comparison of the Two Theories

I must admit that I stole the 'prima facie' in prima facie average causal effect (FACE) from Suppes' definition of prima facie cause because I believe

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

they represent closely related ideas. Both are surface appearances. The FACE measures the empirical association between treatments and outcomes. If the FACE is positive in the medical example given earlier, then t is a prima facie cause of survival in Suppes' sense.

Except for this, the two theories diverge widely because, I argue, the experimental model focuses on the measurement of effects, while Suppes' theory follows the ancient tradition in philosophy of defining what it means to be a cause of something.

Differences in the permanence of results: The problem of the unattainability of the status of being a genuine cause is a difficulty inherent in the approach of identifying causes rather than of measuring effects. For example, it also appears in a different guise in the econometric notion of Granger causality, (Granger 1969), (Florens and Mouchart 1985), in which conditions are given to define when one variable is the cause of another. Granger causality is an approach similar to that taken by Suppes. Granger and others attempt to get around the problem of unattainability by assuming that the set of all the possible D_q's is limited by available information so that it can be exhaustively searched. The status of being a genuine cause now ceases to be unattainable and becomes merely provisional. A genuine cause can be reduced to a mere spurious cause by new data (i.e., a newly discovered D_q).

In practice the identification of a cause often has a provisional quality about it. 'A is a cause of B' is often simply a theoretical statement used to summarize the current state of our ignorance. On the other hand, 'B is an effect of A' can be an empirical finding that does not change or become false as knowledge of the subject increases. As I am fond of saying,

> old experiments never die, they are just reinterpreted.

I intend this paraphrase to emphasize the permanence that experimental results

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

often have, in comparison to that of proposed causal mechanisms. Of course, the status of old experimental results can change in the light of new experimental findings. For example, the introduction of single and double 'blindness' into experiments on human subjects had a profound effect on the interpretation of previous research on the treatment of pain in medicine, (Beecher, 1966).

In non-experimental research the aim is often to measure causal effects with the recognition that because randomization is infeasible and unit homogeneity is implausible the measurement of an ACE by a FACE is biased. In this type of research, older (more biased) estimates of an ACE are replaced by newer (less biased) ones as the quality of the non-experimental research improves. While it might be argued that this is a lack of permanence in the measurement of the effect of a cause, I argue that it is more appropriate to view it as an improvement in this measurement; an improvement that is necessary only because of the infeasibility of a randomized experiment that would make it unnecessary.

Differences in the use of probability: Another important difference between the theories is the way in which they use the language of probability. This is possibly best seen in the analysis of probabilistic causation in the case of a single unit, say the patient Jones. In this special case, the 'events' in Suppes' theory all describe various aspects of Jones and the things that happen to her. The probability that is invoked in his theory refers to Jones alone and not to some population to which Jones might belong. Jones is viewed as a stochastic being, her survival being only probabilistically related to her exposure to the treatment t.

The situation described by the model for experiments is very different. Suppose that Jones is unit 1, then there are just the values of Y_{1t}, Y_{1c}, and x_1. There is nothing said about Jones being stochastic or not. Under exposure to t the outcome Y_{1t} would be recorded, under exposure to c the outcome Y_{1c}

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

would be recorded, and x_1 is whatever it is. We may introduce probabilities for Jones by considering various subpopulations of U to which Jones belongs. For example, suppose that Jones was exposed to t and survived. Can we say anything about the probability that Jones actually survived because she got treatment t in the experiment? Under some circumstances the answer is yes, as I now demonstrate.

Continuing the medical example from section 3, suppose that from a large randomized experiment comparing t to c we have established that

$$Prob\{Y_t = 1\} = .4 \quad \text{and} \quad Prob\{Y_c = 1\} = .1$$

using the methods mentioned in section 3. The ACE is then equal to .4 - .1 = .3.

Suppose, as before, that we are willing to assume that $P(0, 1) = 0$. We can then interpret the ACE as $P(1, 0)$ and conclude that t causes 30% of the units in U to survive, i.e., that the set A_{10} contains 30% of U, including Jones. We know a bit more about Jones, namely, that she was exposed to t and she survived. That is, we know that she is in the first row of Table II. She could be in either A_{11}, the patients who survive regardless of exposure to t or to c, or in A_{10}, those patients for whom t causes survival. Hence, we might consider answering the question "Did t cause Jones to survive?" by reporting the conditional probability formed by dividing $P(1, 0)$ by the sum of it and $P(1, 1)$, i.e., dividing by $Prob\{Y_t = 1\}$. The answer will then depend not only on the value estimated for $P(1, 0)$, but also on the experimental estimate of the marginal probability, $Prob\{Y_t = 1\}$. In this example,

$$Prob\{Y_t = 1\} = .4.$$

Hence, given that Jones got treatment t and survived (i.e., that she is in the first row of Table II), the probability that Jones survived because she got treatment t (i.e., that she is also in the second column of Table II) is .3 divided by .4 or

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

75%. If another patient, say Smith, received treatment *c* and died, the corresponding probability that Smith died because she received treatment *c* is .3 / .9 = 33%.

This is unit inhomogeneity and probabilistic causality in all their glory, because, while we can say something about individual cases, we typically obtain conditional probability statements whose probabilities are far away from 1 or 0.

The relative risk: It is useful here to point out an interesting connection between the conditional probability just computed and a 'causal' parameter that appears in biometric studies called the relative risk. The relative risk is the ratio of the probability of the outcome under *t* to the probability of the outcome under *c*, i.e.,

$$R = Prob\{Y_t = 1\} / Prob\{Y_c = 1\}.$$

(In biometric applications, the outcome coded '1' is usually *death* rather than *survival* as used here, hence the term 'risk' in relative risk.) The relative risk is the parameter used in statements such as "smokers are 10 times more likely to die of lung cancer than are non-smokers".

If $R > 1$ then, following the calculations made above, it is easy to show that the conditional probability that *t* causes survival among the surviving units exposed to *t* is $(R - 1) / R$. (Note that I am still assuming that $P(1, 0) = 0$.) Hence, for a small relative risk, say $R = 1.01$, this conditional probability is also small, i.e., 1%. When the relative risk is big, say $R = 10$, then this conditional probability is also big, i.e., 90%. In the above example, $R = .4 / .1 = 4$ and $(R - 1) / R = 3 / 4 = 75\%$.

Differences in the role of null effects: A third difference between the model for experiments and Suppes' theory is the use of 'screening off'. For Suppes screening off plays an important role in determining when there is only a spurious statistical association between 'cause' and 'effect'. The model for

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

experiments does not involve the use of probability in this way. Instead, to describe everything causal it uses functional relationships--the function *Y* that associates a value with each pair *(i, t)* and the mapping *x* that describes the assignment of treatments to units. Probabilities are proportions of units in the studied population and nothing more. I formed a conditional probability in discussing the probability that *t* caused Jones' survival, but this is not a fundamental aspect of the model, only a natural consequence of using it. Indeed, I think that the strength of the model for experiments is its simultaneous use of functional relationships and of unambiguous population frequencies as probabilities to express all the items of interest. The language is richer than that of probability alone, and more can be said using it.

The role of 'no relationship' or 'spuriousness', while fundamental to the Suppes' theory, appears in the model of experiments in a fairly minor role, as Fisher's null hypothesis. The null value, zero, is only one of many possibilities, no one of which plays a dominant role in this theory. Again, this difference can be attributed to the differing goals of the two theories. When measuring the effects of causes, null effects are only a special case in the range of all possible effect sizes. On the other hand, identifying a genuine cause means being sure that its apparent association with the effect cannot be explained away by any other temporally prior event.

In summary: The 'probability-free' account of probabilistic causation has many of the same aims as other such theories, and it can accomplish much of what these other theories do as well. It is only one application of the general model for experiments described in section 2 and it illustrates that remaining uncommitted as to the nature of the causal connection does not mean that a detailed analysis of many causal questions is out of reach. It is characterized by a focus on measuring the effects of causes and by formal notation that uses the language of both functional relationships and frequentist probability.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The following references may be of interest to those who want more information about the model for experiments and its application to a variety of problems in causal inference.

General introduction, (Holland 1986); foundations, (Rubin 1974, 1978, 1980, 1986, and 1990a); history, (Neyman 1990), (Rubin 1990b); criticisms, (Glymour 1986), (Granger 1986), (Kadane and Seidenfeld 1990), (Lund 1991); observational studies, (Rubin 1974, 1977), (Holland and Rubin 1983), (Rosenbaum and Rubin 1984), (Rosenbaum 1984a, 1984b, and 1987); retrospective studies, (Holland and Rubin 1988); indirect causation and path analysis, (Holland 1988a, Sobel 1990, 1992); employment discrimination, (Holland 1988b); other related material and extensions, (Pratt and Schlaifer 1984, 1988), (Efron and Feldman 1991), (Robins 1985, 1986, and 1987).

Acknowledgements: This paper was prepared while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences and was supported by the Spencer Foundation and Educational Testing Service. I would like to thank my fellow Fellows at the Center, Geoffrey Garrett, William Lycan, and especially Richard Watson for very helpful comments on earlier drafts of the manuscript. In addition, Vance Berger, Michael Sobel and Howard Wainer made several helpful suggestions. Of course, I am responsible for any foolishness that remains.

I first became aware that the model for experiments described in section 2 gave rise to a probabilistic theory of causation in 1987 at Flinders University, through conversations with Professor John Darroch.

References

Beecher, H. K.: 1966, 'Pain: One Mystery Solved' Science **151**, 840-841.

Cox, D. R.: 1958, The Planning of Experiments, John Wiley, New York.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Efron, B. and Feldman. D.: 1991, 'Compliance as an Explanatory Variable in Clinical Trials', Journal of the American Statistical Association **86**, 9-26.

Fisher, R. A.: 1925, Statistical Methods for Research Workers, 1st ed., Oliver and Boyd, Edinburgh.

Fisher, R. A.: 1926, 'The Arrangement of Field Experiments', Journal of Ministry of Agriculture of Great Britain **33**, 503-513.

Florens, J. P. and Mouchart, M.: 1985, 'A Linear Theory for Noncausality', Econometrica **53**, 157-175.

Glymour, C.: 1986, 'Statistics and Metaphysics', Journal of the American Statistical Association **81**, 964-966.

Good, I. J.: 1961, 'A Causal Calculus', British Journal of Science **11**, 305-318; **12**, 43-51; 1962, **13**, 88.

Granger, C. W. J.: 1969, 'Investigating Causal Relations by Econometric Models and Cross-Spectral Methods', Econometrica **37**, 424-438.

Granger, C. W. J.: 1986, 'Comment', Journal of the American Statistical Association **81**, 967-968.

Holland, P. W.: 1986, 'Statistics and Causal Inference', Journal of the American Statistical Association **81**, 945-960.

Holland, P. W. : 1988a, 'Causal Inference, Path Analysis, and Recursive

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Structural Equations Models', in C. C. Clogg (ed.), Sociological Methodology, 1988, American Sociological Association, Washington, D C, pp. 449-484.

Holland, P. W.: 1988b, 'Causal Mechanism or Causal Effect: Which is Best for Statistical Science?', Statistical Science **3**, 186-188.

Holland, P. W. and Rubin, D. B.: 1983, 'On Lord's Paradox', in H. Wainer and S. Messick (eds.), Principals (sic) of Modern Psychological Measurement, Lawrence Erlbaum, Hillsdale, NJ, pp. 3-25.

Holland, P. W. and Rubin, D. B.: 1988, 'Causal Inference in Retrospective Studies', Evaluation Review **12**, 203-231.

Kadane, J. B. and Seidenfeld, T.: 1990, 'Randomization in a Bayesian Perspective', Journal of Statistical Planning and Inference **25**, 329-345.

Kempthorne, O.: 1952, The Design and Analysis of Experiments, John Wiley, New York.

Lewis, D.: 1973, 'Causation', Journal of Philosophy **70**, 556-567.

Lund, T.: 1991, 'Two Metamodels of Causal Effects', Scandinavian Journal of Psychology **32**, 300-314.

Neyman, J.: 1935, 'Statistical Problems in Agricultural Experimentation', Supplement of the Journal of the Royal Statistical Society **2**, 107-180.

Neyman, J.: 1990, 'On the Application of Probability Theory to Agricultural Experiments. Essay on Principles', Translated and edited by D. M. Dabrowska

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and T. P. Speed, Statistical Science **8**, 465-472.

Otte, R.: 1981, 'A Critique of Suppes' Theory of Probabilistic Causality', Synthese **48**, 167-189.

Pratt, J. W. and Schlaifer, R.: 1984, 'On the Nature and Discovery of Structure', Journal of the American Statistical Association **79**, 9-21.

Pratt, J. W. and Schlaifer, R.: 1988, 'On the Interpretation and Observation of Laws', Journal of Econometrics **39**, 23-52.

Robins, J. M.: 1985, 'A New Theory of Causality in Observational Survival Studies--Application of the Healthy Worker Effect', Biometrics **41**, 311

Robins, J. M.: 1986 'A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period--Application to the Control of the Healthy Worker Survivor Effect', Mathematical Modelling **7**, 1393-1512.

Robins, J. M.: 1987, 'A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods', Journal of Chronic Diseases (Supplement 2) **40**, 139S-161S.

Rosenbaum, P. R.: 1984a, 'From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment', Journal of the American Statistical Association **79**, 41-48.

Rosenbaum, P. R.: 1984b, 'The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment', Journal of the Royal Statistical Society, Series A **147**, 656-666.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Rosenbaum, P. R.: 1987, 'The Role of a Second Control Group in an Observational Study', Statistical Science **2**, 292-316.

Rosenbaum, P. R and Rubin, D. B.: 1984, 'Estimating the effects caused by treatments', Journal of the American Statistical Association **79**, 26-28.

Rubin, D. B.: 1974, 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies', Journal of Educational Psychology **66**, 688-701.

Rubin, D. B.: 1977, 'Assignment of Treatment Group on the Basis of a Covariate', Journal of Educational Statistics **2**, 1-26.

Rubin, D. B.: 1978, 'Bayesian Inference for Causal Effects: The Role of Randomization', Annals of Statistics **6**, 34-58.

Rubin, D. B.: 1980, 'Discussion', Journal of the American Statistical Association **75**, 591-593.

Rubin, D. B.: 1986, 'Which Ifs have Causal Answers?', Journal of the American Statistical Association **81**, 961-962.

Rubin, D. B.: 1990a, 'Formal Modes of Statistical Inference for Causal Effects', Journal of Statistical Planning and Inference **25**, 279-292.

Rubin, D. B.: 1990b, 'Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies', Statistical Science **5**, 465-480.

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Salmon, W. C.: 1989, Four Decades of Scientific Explanation, University of Minnesota Press, Minneapolis, MN

Skyrms, B.: 1988, 'Probability and Causation', Journal of Econometrics **39**, 53-68.

Sobel, M. E.: 1990, 'Effect Analysis and Causation in Linear Structural Equation Models', Psychometrika **55**, 495-515.

Sobel, M. E.: 1992, 'Causal Inference in the Social and Behavioral Sciences', in G. Arminger, C. C. Clogg, and M. E. Sobel (eds.) A Handbook for Statistical Modelling in the Social and Behavioral Sciences, Plenum Press, New York, 1-38.

Suppes, P. C.: 1970, A Probabilistic Theory of Causality, North-Holland, Amsterdam.

Tiles, J. E.: 1992, 'Experimental Evidence vs. Experimental Practice', British Journal of Philosophy **43**, 99-109.

Educational Testing Service
Rosedale Rd., 21-T
Princeton, NJ 08541

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table I

A Hypothetical Example of the Formal Model in the Case of a Dichotomous Outcome

	Unit i	Y_{it}	Y_{ic}	x_i	y_i
	1	1	0	t	1
	2	1	0	c	0
	3	1	1	t	1
	4	0	0	c	0
(?)	5	0	1	t	0
	6	1	1	c	1
	-	-	-	-	-
	-	-	-	-	-
	-	-	-	-	-

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table II

The Joint Distribution of Y_t and Y_c

		Y_c		
		1	0	Total
Y_t	1	$P(1, 1)$	$P(1, 0)$	$Prob\{Y_t = 1\}$
	0	$P(0, 1)$	$P(0, 0)$	$Prob\{Y_t = 0\}$
Total		$Prob\{Y_c = 1\}$	$Prob\{Y_c = 0\}$	1.0

23308516, 1993, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1993.tb01530.x, Wiley Online Library on [20/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License