Contents

1. Introduction

In a famous essay on causality, Bertrand Russell (1913) asserts the following.

"All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word 'cause' never occurs . . . . The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm . . . . The principle 'same cause, same effect,' which philosophers imagine to be vital to science, is therefore utterly otiose. As soon as the antecedents have been given sufficiently fully to enable the consequent to be calculated with some exactitude, the antecedents have become so complicated that it is very unlikely they will ever recur. Hence, if this were the principle involved, science would remain utterly sterile . . . . No doubt the reason why the old 'law of causality' has so long continued to pervade the books of philosophers is simply that the idea of a function is unfamiliar to most of them, and therefore they seek an unduly simplified statement. There is no question of repetitions of the 'same' cause producing the 'same' effect; it is not in any sameness of causes and effects that the constancy of scientific law consists, but in sameness of relations. And even 'sameness of relations' is too simple a phrase; 'sameness of differential equations' is the only correct phrase."

Perhaps the most amusing thing about this passage from Russell is that its claim about the use of the word 'cause' in physics no longer holds. Contrary to the days when Russell wrote this essay, the words 'causality' and 'cause' are commonly and widely used by physicists in their most advanced work. There is scarcely an issue of *Physical Review* that does not contain at least one article using either 'cause' or 'causality' in its title. A typical sort of title is that of a recent volume edited by the distinguished physicist, E. P. Wigner, "Dispersion relations and their connection with causality" (1964). Another

good example is the recent article by E. C. Zeeman (1964), "Causality implies the Lorentz group".[1] The first point I want to establish, then, is that discussions of causality are now very much a part of contemporary physics. The reasons for this are, I think, very close to the reasons why notions of causality continue to be an important ingredient of ordinary talk, and undoubtedly will remain so.

At the end of the passage quoted from Russell, there is an emphasis on replacing talk about causes by talk about functional relationships, or more exactly, by talk about appropriate differential equations. This remark is very much in the spirit of classical physics, when the physical phenomena in question were felt to be much better understood at a fundamental level than they are today. One has the feeling that in contemporary physics the situation is very similar to that of ordinary experience, namely, it is not possible to apply simple fundamental laws to derive exact relationships such as those expressed in differential equations. What we are able to get a grip on is a variety of heterogeneous, partial relationships. In the rough and ready sense of ordinary experience, these partial relationships often express causal relations, and it is only natural to talk about causes in very much the way that we do in everyday conversation.

From the standpoint of the philosophical analysis of causality probably the most confusing episode in the history of thought was the reign of Newtonian mechanics, from the beginning of the eighteenth century until the end of the nineteenth. The apparent universality and certainty of this mechanics led Kant and other philosophers into a mistaken notion of causality. The overwhelming empirical success of Newtonian mechanics, particularly in accounting for the motions of the solar system, inevitably yoked the notions of causality and determinism. In the heyday of classical mechanics in the nineteenth century, it was impossible to talk about causes without thinking of them as deterministic in character. Perhaps the clearest expression of this view is to be found in Laplace's treatise on probability.

[1] As additional evidence, the following five titles are cited: B. Ferretti, "On the possibility of a macroscopically causal quantum-relativistic theory" (1963); M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, "Use of causality conditions in quantum theory" (1954); W. Schützer and J. Tiomno, "On the connection of the scattering and derivative matrices with causality" (1951); Yu. M. Shirokov, "Microcovariance and microcausality in quantum theory" (1963); and N. G. van Kampen, "S-matrix and causality condition" (1953).

The famous passage on the deterministic character of the universe is the classical statement of the nineteenth-century position, but this classical statement is very much at variance with the notions of causality used in ordinary talk, and even with those developed in practice by Laplace in applying probability theory.

It is an important point to establish for the subsequent discussion that the everyday concept of causality is not sharply deterministic in character. All of us have said things like the following on one occasion or another: "His reckless driving is bound to lead to an accident". What we mean in asserting such a statement is that the probability of the person's having an accident is high, and his own manner of driving will be at least a partial cause of the accident. The phrase 'is bound to' means that the probability is high of having an accident. The phrase 'lead to' conveys the causal relation between the reckless driving and the predicted accident. If the person in question does not have an accident over an extended period of time, his reporting the fact in conversation would naturally lead to head-shaking and the remark, "It is hard to believe. His driving is a sight to behold. I can't understand why he has not yet had a serious accident".

There are a large number of phrases in English that convey causal notions in the same sort of way as the phrase 'is bound to' lead to'. A common example is the phrase 'due to'. A teacher finds that she must say to a student, "Due to your own laziness you will in all likelihood fail this course". There is, of course, the natural and widely used 'because'. A mother says to a child, "Because it is getting colder today, we probably won't be able to go to the circus tomorrow". Or, a mother says, "The child is frightened because of the thunder", or at another time, "The child is afraid of thunder". She does not mean that on each and every occasion that the child hears thunder, a state of fright ensues, but rather that there is a fairly high probability of its happening; and when it does, of course, the cause of the fright is the thunder. If a tired saleswoman says, "I've been standing all day and my feet are killing me", she does not mean to imply that absolutely every time she is standing at work for a long time, her feet will bother her at the end of the day. It is rather that this will happen with a high probability, and "the" cause of the pain is the prior event of standing a good part of the day. In this last example we have a familiar case in which the ubiquitous connective "and" is used to connect two statements, the first of which expresses the cause of the event described by the second. When the causal anal-

ysis is partial it is perhaps particularly natural to use 'and' rather than 'because'.

It is easy to manufacture a large number of additional examples of ordinary causal language, which express causal relationships that are evidently probabilistic in character. One of the main reasons for this probabilistic character is the open-textured nature of analysis of events as expressed in ordinary language. The completeness and closure conditions so naturally a part of classical physics are not at all a part of ordinary talk. Thus in describing a causal relation in ordinary circumstances, we do not explicitly state the boundary conditions or the limitations on the interaction between the events in question and other events that are not mentioned. In contrast, in classical physics it is standard to state the boundary conditions of the system and to enumerate explicitly all the forces operating on the system at a given time. It is this very lack of knowledge of causes that leads naturally to the introduction of probability concepts in the expression of causal relationships. As Laplace remarked for different reasons, probability theory is designed to discover and to analyze partial causes in complex situations for which a complete causal analysis is not feasible. Laplace had in mind applications of systematic physical theory to such complex situations, but it is for precisely the same reason that probability concepts are so natural in ordinary talk about causes. A complete causal analysis is far too complex and subtle, and not to the point for which ordinary talk is designed.

It is interesting to note in the present context that the analysis of causes in legal contexts given in the recent book by Hart and Honoré (1959) seems closer to classical physics than to ordinary talk about causes, with respect to the role of probability concepts. The reason seems clear. It is characteristic of legal analysis, as well as of classical physics, not to be satisfied with open-ended, probabilistic results. A jury is not permitted to render the verdict that the accused probably committed the crime. In a civil suit concerning breach of contract the judge cannot say, "It seems likely that a breach of contract did occur, but it is not certain, and therefore, we shall apply the following expected utility rule in the award of damages". In other words, the law uses in practice a fiction very similar to that enshrined in classical physics. It is significant that in the rather long and detailed discussion of causation in the book by Hart and Honoré there is scant mention of probability questions.[1]

[1] In discussion of this passage at two seminars on causality I gave at the London School of Economics (Spring, 1966), it was pointed out that in civil damage cases, as for example, those concerned with automobile accidents, there is sometimes a judgment allocating blame between the two parties. However, as far as I know, the allocation is meant to reflect a judgment on the proportional share of the blame and is never formulated in explicit probabilistic terms. Above all, the concept of expected blame, corresponding to the concept of expected utility, is not used.

The omission of probability considerations is perhaps the single greatest weakness in Hume's famous analysis of causality. As is well known, Hume said that the relation between cause and effect has three essential characteristics, namely, contiguity, succession in time, and constant conjunction. In other words, causes and their effects are contiguous in space and time, a cause precedes its effect in time, and causes are followed by their effects in a constant fashion. The important passage on constant conjunction in the *Treatise on Human Nature* (Selby-Bigge edition, 1888, pp. 86–87) is the following.

" 'Tis easy to observe, that in tracing this relation, the inference we draw from cause to effect, is not deriv'd merely from a survey of these particular objects, and from such a penetration into their essences as may discover the dependence of the one upon the other. There is no object, which implies the existence of any other if we consider these objects in themselves, and never look beyond the ideas which we form of them. Such an inference wou'd amount to knowledge, and wou'd imply the absolute contradiction and impossibility of conceiving any thing different. But as all distinct ideas are separable, 'tis evident there can be no impossibility of that kind. When we pass from a present impression to the idea of any object, we might possibly have separated the idea from the impression, and have substituted any other idea in its room.

" 'Tis therefore by EXPERIENCE only, that we can infer the existence of one object from that of another. The nature of experience is this. We remember to have had frequent instances of the existence of one species of objects; and also remember, that the individuals of another species of objects have always attended them, and have existed in a regular order of contiguity and succession with regard to them. Thus we remember to have seen that species of object we call *flame*, and to have felt that species of sensation we call *heat*. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one *cause* and the other *effect*, and infer the existence of the one from that of the other. In all those instances, from which we learn the conjunction of particular causes and effects, both the causes and effects have been perceiv'd

by the senses, and are remember'd: But in all cases, wherein we reason concerning them, there is only one perceiv'd or remember'd, and the other is supply'd in conformity to our past experience.

"Thus in advancing we have insensibly discover'd a new relation betwixt cause and effect, when we least expected it, and were entirely employ'd upon another subject. This relation is their CONSTANT CONJUNCTION. Contiguity and succession are not sufficient to make us pronounce any two objects to be cause and effect, unless we perceive, that these two relations are preserv'd in several instances."

Hume follows this passage with an analysis of why the concept of constant conjunction is the appropriate one to replace the fallacious idea of a necessary connection existing between a cause and its effect. The claim being made here is that in restricting himself to the concept of constant conjunction, Hume was not fair to the use of causal notions in ordinary language and experience. Roughly speaking, the modification of Hume's analysis I propose is to say that one event is the cause of another if the appearance of the first event is followed with a high probability by the appearance of the second, and there is no third event that we can use to factor out the probability relationship between the first and second events.

It is the objective of this monograph to work out the technical details of this fundamental idea and to apply the results to some of the typical philosophical problems that arise in discussions of causality. Section 2 develops an analysis of causal relations among events within a standard probabilistic framework. Section 3 examines how much of this analysis can be retained when only qualitative probability relations are used, and Section 4 develops a qualitative causal algebra. Section 5 analyzes causal relations among quantitative variables or properties, and as might be expected, uses as a central concept the probabilistic concept of random variable. In Section 6, the final section, a large number of problems about causality are discussed, some more extensively than others, but they range from problems about the direction of time to those about the freedom of the will.

2. Causal relations among events

Reasons for defining causality in terms of probability have already been given. The deepest and in many ways the most substantial reason lies in the wide use of probabilistic causal concepts in ordinary talk; but after a formal theory of causality has been defined in terms of probability notions, I shall examine some systematic applications of the theory to various branches of science. The point will be to see in what way causal notions of a probabilistic sort occupy an intuitive and natural place in scientific work.

The formal theory itself is a complex and subtle matter. We must consider a number of examples and counterexamples in order to test the intuitive correctness of the definitions. It should be emphasized that the deterministic concept of cause prominent in classical physics simply occupies the place of a special case in the theory to be outlined here. Roughly speaking, we obtain the deterministic theory by letting all the probabilities in question be either 1 or 0. After working out most of the details of the definitions given here in lectures at Stanford, I discovered that a closely related analysis of causality had been given in an interesting series of articles by I. J. Good (1961, 1962), and the reader is urged to look at Good's articles for a development similar to the one given here, although worked out in rather different fashion formally and from a different viewpoint.

We shall consider later various philosophical caveats that can be made against the underlying formal machinery taken for granted here. For example, to what extent is it necessary to assume that a cause precedes its effect in time? Is it possible to define the direction of time in terms of causality? In the present discussion the direction of time is assumed prior to the definition of any causal concepts. Moreover, David Armstrong has indicated in several conversations on these matters that he feels that the notion of event already requires some causal concepts; yet it will soon be clear that the notion of event, as part of standard probability theory, is essential to the definition of causal concepts given here. Without attempting to settle any of these

matters definitively at this point, I think we can certainly agree that the use of causal concepts in the definition of events can be separated from ordinary or scientific talk about the causes of the events themselves. I shall be in a position to discuss these rather delicate epistemological issues with greater precision after the formal developments have been completed.

For the definitions and theorems of this section we assume the events referred to are all subsets of a fixed probability space, that the events are instantaneous, and that their times of occurrence are included in the formal characterization of the probability space. (The philosophically unsatisfactory aspect of treating all events as instantaneous rather than as "chunks" of time will be discussed later.) We write '$P(A_t)$' for the probability of event A occurring at time t, '$P(A_t|B_{t'})$' for the probability of A occurring at time t given that event B occurred at time t', and so forth, in the standard notation of probability theory. A review of probability concepts is given in the Appendix.

Prima facie causes.

The first definition characterizes prima facie causes.[1] A number of remarks about this definition follow some illustrative examples of its application.

Definition 1. *The event $B_{t'}$ is a prima facie cause of the event A_t if and only if*

(i) $t' < t$,
(ii) $P(B_{t'}) > 0$,
(iii) $P(A_t|B_{t'}) > P(A_t)$.

A familiar application of Definition 1 gives rise to 2 × 2-contingency tables. A classical example is the following study of the efficacy of inoculation against cholera (Greenwood and Yule, 1915, cited in Kendall and Stuart, 1961). The data from the 818 cases studied are as follows.

	Not attacked	Attacked	Totals
Inoculated	276	3	279
Not-inoculated	473	66	539
Totals	749	69	818

[1] The term *prima facie* was suggested by Jaakko Hintikka and is certainly more appropriate than my own use of *naive* in the lectures given at Vaasa, Finland in the summer of 1966.

These data clearly show the efficacy of inoculation, for the mean probability of not being attacked is 749/818 = 0.912, whereas the conditional probability of not being attacked, given that an individual was inoculated, is 276/279 = 0.989. Here A_t is the event of not being attacked by cholera and $B_{t'}$ the event of being inoculated. As statisticians would put it, these data show that inoculation is positively associated with exemption from attack; 'measures of association' is the term commonly used in the statistical literature for measures of the causal relationship exhibited here and required by Definition 1. In discussing measures of association, it is important to emphasize that from a causal standpoint the temporal order of the events is assumed. For example, even though measures of association are commonly defined in such a way that the relation could be symmetric, no one proposes that interpretation of the measure be that attacks of cholera cause earlier inoculations. The ordinary intuitive simple ordering of causal events is assumed and used in any interpretation of such data.

It should be clear that within many conceptual frameworks no positive results about causality can be inferred. A familiar example is provided by the standard coin-tossing experiment in which past outcomes have no effect on future ones. Let h_n be the event of a head on trial n, t_m the event of a tail on trial m, and so forth. Then for a fair coin used in a properly conducted experiment, for $m < n$

$$\text{(1)} \qquad \begin{aligned} P(h_n) &= P(t_n) = P(h_n|h_m) = P(h_n|t_m) \\ &= P(t_n|h_m) = P(t_n|t_m) = 1/2. \end{aligned}$$

The only past events are h_m and t_m, so that for this conceptual framework there are no prima facie causes, as we may infer directly from the equations (1). It is important to emphasize that the determination of a causal relationship between events or kinds of events is always relative to some conceptual framework.

There are at least three different kinds of conceptual frameworks within which it seems appropriate to make causal claims. Within each of these, a rather different basic probability measure will be used in any application of Definition 1. One conceptual framework is that provided by a particular scientific theory; the second is of the sort that arises in connection with a particular experiment or class of experiments; and the third is the most general framework expressing our beliefs with respect to all information available to us.

Learning theory example. To illustrate a theoretical conceptual framework, we may take as an example of suitable complexity the

theory of linear learning models set forth in Estes and Suppes (1959a). For simplicity, let us assume that on every trial the organism can make exactly one of two responses, A_1 or A_2, and after each response it receives a reinforcement, E_1 or E_2, of one of the two possible responses. A learning parameter θ, which is a real number such that $0 < \theta \leq 1$, describes the rate of learning in a manner to be made definite in a moment. A possible realization of the theory is an ordered triple $\chi = < X, P, \theta >$ of the following sort. X is the set of all sequences or ordered pairs such that the first member of each pair is an element of some set A and the second member an element of some set B, where A and B each have two elements. Intuitively, the set A represents the two possible responses and the set B the two possible reinforcements. P is a probability measure on the σ-algebra of cylinder sets of X, and θ is a real number as already described. (Actually there is a certain arbitrariness in the characterization of possible realizations of theories whose models have a rather complicated set-theoretical structure, but this is a technical matter into which we shall not enter here.) To define the models of the theory, we need a certain amount of notation. Let $A_{i,n}$ be the event of response A_i on trial n; $E_{j,n}$ the event of reinforcement E_j on trial n, where $i, j = 1, 2$; and for x in X let x_n be the equivalence class of all sequences in X which are identical with x through trial n. We may then characterize the theory by the following set-theoretical definition.

A triple $\chi = < X, P, \theta >$ is a linear learning model if and only if the following two axioms are satisfied:

A1. *If* $P(E_{i,n}A_{i',n}x_{n-1}) > 0$ *then*
$P(A_{i,n+1}|E_{i,n}A_{i',n}x_{n-1}) = (1 - \theta)P(A_{i,n}|x_{n-1}) + \theta.$

A2. *If* $P(E_{j,n}A_{i',n}x_{n-1}) > 0$ *and* $i \neq j$ *then*
$P(A_{i,n+1}|E_{j,n}A_{i',n}x_{n-1}) = (1 - \theta)P(A_{i,n}|x_{n-1}).$

As is clear from the two axioms, this linear response theory is intuitively very simple. The first axiom just says that when a response is reinforced, the probability of making that response on the next trial is increased by a simple linear transformation. The second axiom says that if some other response is reinforced, the probability of making the response is decreased by a second linear transformation. In spite of the simplicity of this theory, it gives a reasonably good account of a number of experiments, and from a mathematical standpoint it is by no means trivial to characterize asymptotic properties of its models.

The theoretical models of the theory of linear learning are determined by three types of parameters. First, a numerical value for the learning parameter θ must be selected; second, the initial probability of an A_1 response must be selected, that is, the probability $P(A_{1,1})$; and third, a reinforcement schedule must be chosen. One of the simplest reinforcement schedules that has been much studied experimentally is the case of simple noncontingent reinforcement. On every trial the probability of an E_1 reinforcement, independent of any preceding events, is π. Once π is selected, then the probabilistic character of the model is determined uniquely. If, for example, we set $\Theta = P(A_{1,1}) = \pi = 1/2$, then a unique model of the theory is determined, and all probability questions that can be asked meaningfully have a unique numerical answer.

Within this framework we may then ask causal questions in the sense of Definition 1. For example, is an earlier response a prima facie cause of a later response of the same kind; is an earlier reinforcement of a response a prima facie cause of that response occurring on a subsequent trial? We shall consider the analysis of causal relations only at asymptote; that is, when n approaches ∞. This simplification is just a convenient simplification in the mathematics, and is not critical to the conceptual distinctions being made. Nine typical asymptotic predictions in terms of π and θ are the following:

$$
\begin{cases}
P(A_{1,n+1}) = \pi \\
P(A_{1,n+1}|A_{1,n}) = \dfrac{1-\theta}{2-\theta}[2\pi(1-\theta)+\theta] + \theta\pi \\
P(A_{1,n+1}|A_{2,n}) = 1 - \dfrac{1-\theta}{2-\theta}[2(1-\pi)(1-\theta)+\theta] - \theta(1-\pi) \\
P(A_{1,n+1}|E_{1,n}) = (1-\theta)\pi + \theta \\
P(A_{1,n+1}|E_{2,n}) = (1-\theta)\pi
\end{cases}
\tag{2}
$$

$$
\begin{cases}
P(A_{1,n+1}|E_{1,n}A_{1,n}) = \dfrac{1-\theta}{2-\theta}[2\pi(1-\theta)+\theta] + \theta \\
P(A_{1,n+1}|E_{1,n}A_{2,n}) = 1 - \dfrac{1-\theta}{2-\theta}[2(1-\pi)(1-\theta)+\theta] \\
P(A_{1,n+1}|E_{2,n}A_{1,n}) = \dfrac{1-\theta}{2-\theta}[2\pi(1-\theta)+\theta] \\
P(A_{1,n+1}|E_{2,n}A_{2,n}) = 1 - \dfrac{1-\theta}{2-\theta}[2(1-\pi)(1-\theta)+\theta] - \theta.
\end{cases}
$$

The derivations of these nine expressions are to be found in Estes and

Suppes (1959a) and Suppes and Atkinson (1960). As an example of how the derivations proceed from the axioms, we may take as typical the derivation of $P(A_{1,n+1}|E_{1,n}A_{1,n})$ when $n \to \infty$.

Even this simple example is not exactly obvious. To begin with, we need an asymptotic expression for a certain second moment, defined as follows:

$$V_{2,n} = \sum_{x_{n-1}} P(A_{1,n}|x_{n-1})^2 P(x_{n-1}).$$

We proceed recursively, applying the axioms to appropriate cases:

$$V_{2,n+1} = \sum_{i}\sum_{j}\sum_{x_{n-1}} P(A_{1,n+1}|E_{i,n}A_{j,n}x_{n-1})^2 P(E_{i,n}|A_{j,n}x_{n-1}) \cdot P(A_{j,n}|x_{n-1})P(x_{n-1})$$

$$= \sum_{x_{n-1}} \Big\{[(1-\theta)P(A_{1,n}|x_{n-1}) + \theta]^2 \pi + (1-\theta)P(A_{1,n}|x_{n-1})^2 (1-\pi)\Big\} \cdot P(x_{n-1})$$

$$= (1-\theta)^2 V_{2,n} + 2\pi\theta(1-\theta)P(A_{1,n}) + \pi\theta^2 .$$

Now it may be shown that $\lim_{n\to\infty} V_{2,n} = V_2$ exists, whence at asymptote, i.e., as $n \to \infty$

$$V_{2,n+1} = V_{2,n},$$

and therefore from the above recursive expression,

$$V_2 = \frac{\pi[2\pi(1-\theta) + \theta]}{2-\theta}, \tag{3}$$

using the fact that

$$\lim_{n\to\infty} P(A_{1,n}) = \pi . \tag{4}$$

We are now ready to find the asymptotic probability $P(A_{1,n+1}|E_{1,n}A_{1,n})$ as a function of θ and π. We begin by applying elementary probability theory, and by using (4) as well as the fact that $P(E_{1,n}|W_{n-1}) = \pi$ for any past event W_{n-1} with $P(W_{n-1}) > 0$.

$$P(A_{1,n+1}|E_{1,n}A_{1,n}) = \frac{1}{\pi}\sum_{x_{n+1}} P(A_{1,n+1}|E_{1,n}A_{1,n}x_{n-1})P(A_{1,n}|x_{n-1}) \cdot P(x_{n-1})$$

$$= \frac{1}{\pi}\sum_{x_{n-1}} [(1-\theta)P(A_{1,n}|x_{n-1}) + \theta] \cdot P(A_{1,n}|x_{n-1})P(x_{n-1})$$

$$= \frac{(1-\theta)}{\pi} V_{2,n} + \theta ,$$

and so, using (3), as $n \to \infty$

$$P(A_{1,n+1}|E_{1,n}A_{1,n}) = \frac{1-\theta}{2-\theta}[2\pi(1-\theta) + \theta] + \theta ,$$

as desired. The approach used here applies with minor modification to the derivation of the other conditional probabilities in (2).

If we take both θ and π to be numbers strictly between 0 and 1, then on the basis of these nine expressions we may establish inequalities justifying the following statements in terms of Definition 1. (It is to be remembered that the expressions and following statements hold without restriction only at asymptote.)

1. $A_{1,n}$ is a prima facie cause of $A_{1,n+1}$.
2. $A_{2,n}$ is not a prima facie cause of $A_{1,n+1}$.
3. $E_{1,n}$ is a prima facie cause of $A_{1,n+1}$.
4. $E_{2,n}$ is not a prima facie cause of $A_{1,n+1}$.
5. The event $E_{1,n}A_{1,n}$ is a prima facie cause of response $A_{1,n+1}$.
6. The event $E_{2,n}A_{2,n}$ is not a prima facie cause of $A_{1,n+1}$.

The inequalities that will establish the statements just made are straightforward. The two cases that are somewhat more sophisticated are the cases in which a like response but unlike reinforcement is given on the preceding trial or an unlike reinforcement but like response is given. The relevant theoretical equations are the seventh and eighth equations of (2). In the first of these two cases an inference about prima facie causation depends upon the relative value of θ and π, as the following argument shows.

From the seventh theoretical equation of (2), it is clear that in order to assert that the joint event $E_{1,n}A_{2,n}$ is a prima facie cause of $A_{1,n+1}$, we need to establish that the following inequality holds:

$$1 - \frac{1-\theta}{2-\theta}[2(1-\pi)(1-\theta)] > \pi ,$$

which is equivalent to:

$$(2-\theta)(1-\pi) > 2(1-\pi)(1-\theta)^2 + \theta - \theta^2 ,$$

which in turn, after some simplification, is equivalent to

$$(2\pi - 1)\theta > 3\pi - 2,$$

so that in order for the causation to hold theoretically, we must have

$$\theta > \frac{3\pi - 2}{2\pi - 1},$$

provided $\pi \neq \frac{1}{2}$.

This inequality is satisfied, for example, by $\pi = .9$ and $\theta = .9$, but not by $\pi = .7$ and $\theta = .1$. Our general causal assertion in this case is thus:

7. The joint event $E_{1,n}A_{2,n}$ is a prima facie cause of $A_{1,n+1}$ if and only if $\theta > \frac{3\pi - 2}{2\pi - 1}$, provided $\pi \neq \frac{1}{2}$, and if $\pi = \frac{1}{2}$ the joint event is a prima facie cause for all permissible values of θ.

By a similar approach we may establish from the eighth equation this theoretical causal statement:

8. The joint event $E_{2,n}A_{1,n}$ is not a prima facie cause of $A_{1,n+1}$ for any permissible values of θ and π, i.e., any values strictly between 0 and 1.

The comparison of these last two causal statements is instructive in that we immediately see that a like preceding reinforcement is causally more important than a like preceding response, as intuitively we would expect.

The analysis of the kind of causal statements that can be made within this particular theoretical framework will be continued later, but at this point I want to move on to the consideration of the closely related causal statements that can be made in connection with a particular experiment or class of experiments designed to test the theory. In the present case, I shall draw upon data reported in Suppes and Atkinson (1960, Chapter 10). Data are reported there for all of the nine theoretical quantities defined by (2).

In the experiment to be reported here, the subjects were thirty undergraduates from introductory courses at Stanford University. The experimental apparatus may be described as follows (Suppes and Atkinson, 1960, p. 81).

> "The subjects, run in pairs, sat at opposite ends of an 8 × 3-foot table. Mounted vertically on the table top facing each subject was a 50-inch-wide by 30-inch-high black panel placed 22 inches from the end of the table. The experimenter sat between the two panels and was not visible to either subject. The apparatus, as viewed from the subject's side, consisted of two silent operating keys mounted 8 inches apart on the table top and 12 inches from the end of the table; upon the panel, three milk-glass panel lights were mounted. One of these lights, which served as the signal for the subject to respond, was centered between the keys at a height of 17 inches from the table top. Each of the two remaining lights, the reinforcing signals, was at a height of 11 inches directly above one of the keys. The presentation and duration of the lights were automatically controlled. The subjects were not visible to one another and could not see one another's responses or panel lights."

Each pair of subjects was read instructions, of which the following is an excerpt (pp. 196–197):

> "This experiment is a game in which you will be playing against each other. The game is similar to a real-life situation in that what you gain or lose depends not only on what you do, but also on what someone else — your opponent — does . . . ."
>
> " 'To keep things straight I will call you "Player A" and you "Player B". The experiment for each of you consists of a series of trials. The top center light on each of your panels will go on for two seconds to indicate the start of each trial. When this light goes on, you will each press one or the other of the two keys in front of you. That is, Player A will press either his A_1 or his A_2 key; Player B will press his B_1 or B_2 key. Then wait until one of the lower lights goes on. If the light above the key you pressed goes on, your prediction was correct; if the light above the key opposite the one you pressed goes on, you were incorrect . . . .
>
> " 'Being correct or incorrect on a given trial depends on the key you press and also on the key the other player presses. With some combinations of your key choice with the other player's key choice, you may both be correct; with other combinations, you may both be incorrect . . . .
>
> " 'Your job, then is to obtain as many correct responses as you possibly can . . . . The trials move along rapidly, and you must make your key choice as soon as the signal light goes on. That is, when the signal light goes on, press one or the other key and release it before the signal light goes off. Then wait until one of the lower white lights goes on. If the light above the key you pressed goes on, you are correct; and if the light above the key opposite from the one you pressed goes on, you are incorrect'."

The experiment was then carried out as follows (p. 197):

> "After the instructions were read, 240 trials were run in continuous sequence. Fifteen reinforcement schedules were used. Each was constructed randomly, with the restriction of exactly 144 E_1's and 96 E_2's. Both members of a subject-pair received the same schedule of E_1's and E_2's. Further, once a given schedule had been used for a subject-pair, it was not used again in that particular experimental group. Thus, within each group two subjects received

identical reinforcement schedules; and across groups each schedule was represented exactly once".

It is evident from the statement about the number of E_1 reinforcements that the value of π in the experiment was set at .6. Using a pseudomaximum-likelihood estimate of θ; which will not be described here but is discussed in detail in Suppes and Atkinson (1960), the estimated value of θ for the data was determined as .19. Using these values of π and θ, the theoretically predicted and observed relative frequencies corresponding to the nine conditional probabilities given in (2) are shown in Table 1.

Because the observed data and the predicted values of the asymptotic quantities are very close indeed, all of the qualitative statements about prima facie causes derived from the theory also apply to the experimental data themselves. In the case of $P(A_1|E_1A_2)$ the inequality derived earlier is supported by the data, but not as strongly as theoretically predicted.

The experimental probability measure used in order to decide within the second framework of analysis what sort of causality statements are appropriate is not the only probability measure that could be derived from the experiment, but it is certainly the most common one. It is the measure that is derived simply from the relative frequency data themselves. Such a measure can be justified as being the maximum-likelihood estimate of the true experimental probabilities. This is a technical point that need not be expanded upon here. Because of the large number of experimental observations, essentially the same probabilities shown as "observed" in Table 1 would be obtained by any other standard method of estimation. In the present instance the experimental probabilities justify precisely

Table 1. Comparison of observations and predictions of the linear model over the last 100 trials of the experiment

Asymptotic	Predicted	Observed
$P(A_1)$	.600	.596
$P(A_1\|A_1)$	.634	.641
$P(A_1\|A_2)$	.549	.532
$P(A_1\|E_1)$	.676	.667
$P(A_1\|E_2)$	.486	.489
$P(A_1\|E_1A_1)$	.710	.715
$P(A_1\|E_1A_2)$	.625	.602
$P(A_1\|E_2A_1)$	.520	.535
$P(A_1\|E_2A_2)$	.435	.413

the same qualitative causal statements as the theoretical predictions shown in Table 1, but in general this would not be the case. It is easy enough to produce data which lead to a clash between the causal statements derived from the theory and those observed in a given experiment.

Let us now turn to the third and most general framework within which we might apply Definition 1. This is the framework expressing our beliefs with respect to all information available to us. In the case of the particular experiment we have been discussing, it is, I think, hard to derive from our general beliefs any quantitative probabilistic predictions that would lead to an additional set of causal statements or causal predictions as an application of Definition 1. On the other hand, it is easy enough on the basis of general beliefs to have certain hypotheses about the behavior of subjects, and not to be content with the evidence offered against these hypotheses, either by the experiment itself or the supporting theoretical analysis. In the present case, for example, we might hold with many cognitive psychologists that the kind of reinforcement theory expressed in the linear model is simply not appropriate for the analysis of human behavior; and therefore, even though the predicted and observed values shown in Table 1 are extremely close, a deeper and more adequate general theory would assume that subjects are using hypotheses that are being tested and rejected over the course of the experiment. Such a theory, we might argue, will provide ultimately a more adequate analysis of this kind of experiment. In particular, we might hold that we certainly are not prepared to accept the subject's own preceding responses as causes of a response. In this case, we might say that even though the conditional probabilities indicate the appropriate relations for Definition 1, a better explanation is to be found. That search for a better explanation takes us to Definition 2, which provides a criterion for judging prima facie causes as spurious.

Spurious causes.

The intuitive idea of a spurious cause is that an earlier event may be found which accounts for the conditional probability of the effect just as well. Formally we have the following preliminary definition.

Let $B_{t'}$ be a prima facie cause of A_t. Then $B_{t'}$ is a spurious cause of A_t if and only if there is a $t'' < t'$ and an event $C_{t''}$ such that $P(B_{t'}C_{t''}) > 0$ and

$$P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''}) . \tag{5}$$

It is to be admitted at once that (5) defines spurious causes in too simple a manner. Subsequently we shall want to elaborate on this definition. Questions can be raised even about the simplest examples of the definition. Some discussions argue for an equation in which $\bar{B}_{t'}$, the complement of $B_{t'}$, is included in the defining equation, so that (5) is replaced by

$$P(A_t|B_{t'}C_{t'}) = P(A_t|\bar{B}_{t'}C_{t'}), \tag{6}$$

provided, of course, that $P(\bar{B}_{t'}C_{t'}) > 0$. However, under the provision just made, we may show that (5) and (6) are equivalent, simply as a consequence of elementary probability theory, and thus we do not need to bring in the complement of $B_{t'}$ in defining the spuriousness of $B_{t'}$. To show the equivalence, first let us suppose that (5) holds. Then from obvious relations for conditional probabilities, we have:

$$\begin{aligned} P(A_t|C_{t'}) = {} & P(A_t|B_{t'}C_{t'})\,P(B_{t'}|C_{t'}) \\ & + P(A_t|\bar{B}_{t'}C_{t'})\,P(\bar{B}_{t'}|C_{t'}) \end{aligned} \tag{7}$$

whence, using (5), we have from (7)

$$P(A_t|B_{t'}C_{t'})\,(1 - P(B_{t'}|C_{t'})) = P(A_t|\bar{B}_{t'}C_{t'})\;P(\bar{B}_{t'}|C_{t'}), \tag{8}$$

but of course,

$$1 - P(B_{t'}|C_{t'}) = P(\bar{B}_{t'}|C_{t'}),$$

and so the desired result follows from (8).

Now assume that (6) holds, then from (6) and (7) we have

$$\begin{aligned} P(A_t|C_{t'}) &= P(A_t|B_{t'}C_{t'})\,[P(B_{t'}|C_{t'}) + P(\bar{B}_{t'}|C_{t'})] \\ &= P(A_t|B_{t'}C_{t'}), \end{aligned}$$

as claimed.

It is a temptation to make the temporal inequality in the preliminary definition weak ($t'' \leq t'$) rather than strict ($t'' < t'$) and thereby permit $B_{t'}$ and $C_{t'}$ to occur simultaneously. However, it is easy to show that without the introduction of some other restriction, this weakening will not do, because it would permit us to show that every cause is spurious. Let $t'' = t'$ and let $C_{t'} = C_{t'} = B_{t'}$. Then for any $B_{t'}$ with $P(B_{t'}) > 0$

$$P(A_t|B_{t'}C_{t'}) = P(A_t|C_{t'}).$$

Now if we impose the natural requirement that $B_{t'} \neq C_{t'}$, we can still show any event is spurious if we permit the weak temporal inequality and if we can find any other event $D_{t'}$ distinct from $B_{t'}$ such that $P(B_{t'}D_{t'}) > 0$; for we may then take

$$C_{t'} = B_{t'} \cap D_{t'},$$

and obviously (5) holds for this definition of $C_{t'}$, because

$$C_{t'} = B_{t'} \cap C_{t'}.$$

However, this last example suggests a real defect of the preliminary definition. Suppose that

$$P(A_t|B_{t'}) > P(A_t|B_{t'}C_{t'}).$$

It hardly seems reasonable to call $B_{t'}$ spurious when it alone predicts the occurrence of A_t with higher probability than does the joint event $B_{t'} \cap C_{t'}$. It is apparent that this inequality is consistent with the preliminary definition. In the revised definition it seems intuitively sound to require that its negation hold, i.e., to impose the weak inequality

$$P(A_t|B_{t'}C_{t'}) \geq P(A_t|B_{t'}).$$

Even with the inequality imposed it still seems desirable to impose the strict inequality $t'' < t'$. The reason for this is that we would make spurious any causes that are not maximal if we permitted $t'' = t'$. Thus, for example, if $B_{t'}$ were a prima facie cause that was not spurious in the sense being defined, but if there were an event $C_{t'}$ such that $B_{t'} \neq C_{t'}$, $P(B_{t'}C_{t'}) > 0$ and

$$P(A_t|B_tC_{t'}) > P(A_t|B_{t'}),$$

then $B_{t'}$ would not be the maximal cause of A_t at time t'. It seems clearly desirable, however, to distinguish spurious from nonmaximal causes. These various remarks about spurious causes are brought together in the following definition.

Definition 2. *An event $B_{t'}$ is a spurious cause in sense one of A_t if and only if $B_{t'}$ is a prima facie cause of A_t and there is a $t'' < t'$ and an event $C_{t''}$ such that*

(i) $P(B_{t'}C_{t''}) > 0$,
(ii) $P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''})$,
(iii) $P(A_t|B_tC_{t''}) \geq P(A_t|B_{t'})$.

I am still not certain that the three conditions of Definition 2 are precisely the right ones. I am particularly uneasy about (ii), because there seem to be some arguments in favor of replacing it by the inequality

$$P(A_t|B_{t'}C_{t''}) \leq P(A_t|C_{t''}),$$

so that $B_{t'}$ is spurious if the other conditions are satisfied and the occurrence of $B_{t'}$ after $C_{t''}$ actually lowers the probability of A_t. When the strict inequality holds, one is inclined to call $B_{t'}$ a negative cause of A_t after the earlier occurrence of $C_{t''}$, rather than a spurious cause. My intuition is that *spurious* should mean no real influence at all, either positive or negative, and therefore I shall stand by condition (ii) for the present, but with no strong feeling of correctness about the decision.

We may define a prima facie cause that is not spurious as *genuine*. Familiar examples of both spurious and genuine causes are easily produced. Some simple artificial examples may be clarifying. Consider first the three-state Markov chain whose transition matrix is

n \ $n+1$	0	1	2
0	0	1/2	1/2
1	1/3	0	2/3
2	1/4	3/4	0 .

Here every prima facie cause is genuine. Now consider the transition-matrix

	0	1	2
0	0	1/2	1/2
1	1/4	3/4	0
2	1/4	0	3/4 .

The process has the Markov property; but according to Definition 2, and also, I believe, according to intuition, the event of being in either state 1 or state 2 is a spurious cause of being in state 0 on the next trial, because $P(0_n) = .2$ as $n \to \infty$ and

$$P(0_n|1_{n-1}0_{n-2}) = P(0_n|2_{n-1}0_{n-2}) = P(0_n|0_{n-2}) = 1/4.$$

I realize that the application of causal terminology to these simple Markov examples makes some people uneasy, and I shall want to explore this problem in some detail later. For the moment I remark only that this uneasiness probably comes from the very strongly felt need to identify what seem to be ultimate causes and not to use causal terminology at all in dealing with processes that intentionally catch only a partial aspect of the real world.

A deeper problem may be raised about Definition 2 and its intuitive adequacy for characterizing spurious causes. Definition 2 makes a prima facie cause spurious if there *exists* an earlier event that eliminates the effectiveness of the cause when that event occurs. It is true that condition (iii) imposes a rather strong constraint on this earlier event. But an intuitively appealing alternative approach is to drop (iii) and demand a partition of the past before the spurious cause such that for every element in the partition, conditions (i) and (ii) hold. Intuitively this amounts to requiring that if we can observe a certain kind of earlier event, then knowledge of the spurious cause is predictively uninformative. The existential requirement is now moved from a demand for an event to a demand for a kind of event or property. For formal purposes we note that a partition of the sample space or universe is a collection of pairwise disjoint, nonempty sets whose union is the whole space. For our purposes we shall also require that a partition π_t consists of events that can be defined by references to times no later than t.

Definition 3. *An event $B_{t'}$ is a spurious cause of A_t in sense two if and only if B_t is a prima facie cause of A_t and there is a $t'' < t'$ and a partition $\pi_{t''}$ such that for all elements $C_{t''}$ of $\pi_{t''}$*

$$\text{(i)}\quad P(B_{t'}C_{t''}) > 0\,,$$
$$\text{(ii)}\quad P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''}).$$

It is probably intuitively evident that spuriousness in sense two implies spuriousness in sense one, but explicit statement and proof of the theorem are perhaps desirable.

Theorem 1. *If event $B_{t'}$ is a spurious cause in sense two of A_t, then $B_{t'}$ is a spurious cause in sense one of A_t.*

Proof: What we need to prove is that the two conditions of Definition 3, holding as they do for all events $C_{t''}$ that are elements of the partition $\pi_{t''}$, imply that for some $C_{t''}$ in $\pi_{t''}$ condition (iii) of Definition 2 is satisfied. The simplest proof seems to be a reductio ad absurdum. Suppose that for every $C_{t''}$ of $\pi_{t''}$

$$P(A_t|B_{t'}C_{t''}) < P(A_t|B_{t'})\,. \tag{9}$$

Now by elementary probability theory

$$P(A_t|B_{t'}) = \frac{1}{P(B_{t'})}\sum_{\pi t''} P(A_t|B_{t'}C_{t''})P(B_{t'}|C_{t''})P(C_{t''})\,,$$

whence using (9)

$$P(A_t|B_{t'}) < \frac{P(A_t|B_{t'})}{P(B_{t'})} \cdot \sum P(B_{t'}|C_{t''})P(C_{t''})$$

$$< \frac{P(A_t|B_{t'})}{P(B_{t'})} \cdot P(B_{t'})$$

$$< P(A_t|B_{t'}),$$

which is a contradiction. Q.E.D.

It is easy to construct examples to show that the converse of the theorem does not hold.

It may be worthwhile to examine the application of Definition 3 to the linear learning theory considered earlier. Perhaps the most interesting conceptual point is whether earlier responses become spurious causes of later responses. We know from our earlier analysis that $A_{t,n}$ is a prima facie cause of $A_{t,n+1}$, and this result can be easily generalized to show that $A_{t,m}$ is a prima facie cause of $A_{t,n}$ for $m < n$. As has already been indicated, the intent of the linear learning theory with the single parameter θ is to make only reinforcements causally relevant. A direct application of Definition 3 would suggest that we look at the partition made up of the two events $E_{1,n-1}$ and $E_{2,n-1}$. Condition (iv) is readily satisfied in the context carried over from our earlier discussion (noncontingent reinforcement with $\pi = .6$ and θ estimated as .19). In particular,

$$P(A_{t,n}E_{j,n-1}) > 0$$

for $i, j = 1, 2$. But the equality expressed in condition (v) is not satisfied in either of the two possible cases. For $i = 1, 2$

$$P(A_{t,n+1}|A_{t,n}E_{t,n-1}) > P(A_{t,n+1}|E_{t,n-1}). \tag{10}$$

The verification of (10) involves tedious arguments in terms of inequalities and will therefore be omitted here.[1] Thus, contrary to our initial hope, $A_{t,n}$ has not been shown to be spurious.

The general assumption about events thus far has been that they are instantaneous. However, any serious probabilistic analysis must

[1] For example, at asymptote,

$$P(A_{1,n+1}|A_{1,n}E_{1,n-1}) = \frac{1}{(1-\theta)\pi+\theta}\left\{\frac{(1-\theta)^2\pi\,[2\pi(1-\theta)+\theta]}{2-\theta} + [2\theta(1-\theta)^2+\theta\pi(1-\theta)]\pi+\theta^2(1-\theta)+\theta^2\pi\right\}$$

and

$$P(A_{1,n+1}|E_{1,n-1}) = (1-\theta)^2\pi+\theta(1-\theta)+\theta\pi.$$

deal with joint events like $A_{1,n}E_{1,n-1}$ which occur not at an instant, but only if the appropriate atomic events occur at n and n-1. In other words, we assume that the class of instantaneous events is closed under intersection and complementation of events, which themselves may occur at different times. Keeping these remarks in mind, we may search for much more elaborate partitions of the past in order to render $A_{t,n}$ spurious. The natural partition to consider is the finest one possible that may be defined solely in terms of the reinforcements preceding $A_{t,n}$. In this case, we shall not initially analyze the situation at asymptote, for as $n \to \infty$, $P(x_n) = 0$ for every x_n, and therefore it is not possible to satisfy condition (i) of Definition 3. For the noncontingent reinforcement schedule we are considering, as long as $0 < \pi, \theta, P(A_{1,1}) < 1$, it is straightforward to show that for every $x_{E,n-1}$ and for $i = 1, 2$

$$P(A_{t,n}x_{E,n-1}) > 0 \tag{11}$$

and

$$P(A_{t,n+1}|A_{t,n}x_{E,n-1}) = P(A_{t,n+1}|x_{E,n-1}), \tag{12}$$

where $x_{E,n-1}$ is the equivalence class of past histories x'_{n-1} all of which have exactly the same reinforcement sequence as x_{n-1}. In other words $x_{E,n-1}$ is a unique sequence of reinforcements from trial 1 to trial n-1. On the basis of (11) and (12) we may then assert that $A_{t,n}$ is a spurious cause in the second sense of $A_{t,n+1}$.

What about the asymptotic case? The intuitive situation seems clear. As we consider longer and longer finite strings of preceding reinforcements, the effects of the immediately preceding response become smaller and smaller. This intuition may be shown to be correct, and it suggests a useful generalization of Definition 3. We may define an event as an ϵ-spurious cause of A_t when its effect on the probability of occurrence of the event A_t is less than ϵ. This definition permits us to rule out as significant genuine causes events that play only a minor role.

Definition 4. An event $B_{t'}$ is an ϵ-spurious cause of A_t if and only if there is a $t'' < t'$ and a partition $\pi_{t''}$ such that for all elements $C_{t''}$ of $\pi_{t''}$

(i) $t' < t$,
(ii) $P(B_{t'}) > 0$,
(iii) $P(A_t|B_{t'}) > P(A_t)$,
(iv) $P(B_{t'}C_{t''}) > 0$,
(v) $|P(A_t|B_{t'}C_{t''}) - P(A_t|C_{t''})| < \epsilon$.

In terms of this definition we may show in the asymptotic learning case that for any $\epsilon > 0$ there exists a partition of strings of immediately preceding reinforcements of finite length which show that $A_{t,n}$ is an ϵ-spurious cause of $A_{t,n+1}$.

It is also easy to think of many physical examples of causes that are ϵ-spurious for very small epsilon. Although the explicit consideration of quantitative variables is outside the framework of this section, classical examples of ϵ-spurious causes are to be found in the theory of motion of the planets. The moons of Jupiter, for example, are ϵ-spurious causes, for extremely small ϵ, of the motion of Earth, or any of the planets except Jupiter. They are even ϵ-spurious causes of the motion of Jupiter for fairly small ϵ.

Tacit use of the concept of ϵ-spurious causes is essential to many branches of science in designing and appraising the outcome of experiments aimed at testing theories. It is important to distinguish ϵ-spurious causes from random errors of measurement and sampling. Random errors of measurement and sampling do not lead to rejection of a theory but provide a framework for showing how discrepancies between observed and predicted results may be explained. Causes that are ϵ-spurious, on the other hand, do lead to rejection of the theory or hypothesis in question under any strict interpretation of the statistical analysis of the data; but when ϵ is small, we know that the discrepancies unexplained by the theory are of relatively small order. In many practical cases, information about the magnitude of ϵ is much more important than the knowledge that a theory is significantly deviant from the facts. (For an empirical application of these ideas see Suppes and Rouanet (1964), and for a working out of the statistical theory, Kraemer (1965).)

Direct causes.

The concept of an indirect cause is of less importance but similar in structure to the concept of a spurious cause. In this case, however, it is somewhat more natural to define direct rather than indirect causes. Of course, a prima facie cause is indirect if it is not direct.

Definition 5. *An event $B_{t'}$ is a direct cause of A_t if and only if $B_{t'}$ is a prima facie cause of A_t and there is no t'' and no partition $\pi_{t''}$ such that for every $C_{t''}$ in $\pi_{t''}$*

(i) $t' < t'' < t$,
(ii) $P(B_{t'}C_{t''}) > 0$,
(iii) $P(A_t|C_{t''}B_{t'}) = P(A_t|C_{t''})$.

The conditions of Definition 5 are almost precisely those of Definition 3 except that now time t'' comes between t and t' rather than before t'.

Indeed, the symmetry of the conditions in Definitions 3 and 5 suggest that causes $B_{t'}$ and $C_{t''}$ of A_t, with $t'' < t'$, could have the following relation. Event $B_{t'}$ is a spurious cause of A_t because of the prior partition $\{C_{t''}, \overline{C_{t''}}\}$, and concurrently $C_{t''}$ is an indirect cause of A_t because of the later partition $\{B_{t'}, \overline{B_{t'}}\}$. If this line of reasoning worked, one has the feeling we might be able to show that in many circumstances there is a natural linking of spurious and indirect causes. The following theorem shows, however, that this cannot happen.

Theorem 2. *Let $B_{t'}$ and $C_{t''}$ be prima facie causes of A_t, with $t'' < t'$. Then it cannot be the case that jointly $C_{t''}$ is an indirect cause of $A_{t'}$ because of the partition $\{B_{t'}, \overline{B_{t'}}\}$, and $B_{t'}$ is a spurious cause (in sense two) of A_t because of the partition $\{C_{t''}, \overline{C_{t''}}\}$.*

Proof: To simplify notation I drop the temporal subscripts in the proof. I derive a contradiction from the hypothesis that C is an indirect cause of A because of $\{B, \bar{B}\}$ and that B is a spurious cause of A because of $\{C, \bar{C}\}$. From the hypothesis about C we have

$$P(A|BC) = P(A|B), \tag{1}$$
$$P(A|\bar{B}C) = P(A|\bar{B}); \tag{2}$$

and from the hypothesis about B we have

$$P(A|BC) = P(A|C), \tag{3}$$
$$P(A|B\bar{C}) = P(A|\bar{C}). \tag{4}$$

From (1) and (3), and the hypothesis that B and C are prima facie causes of A, we have

$$P(A|B) = P(A|C) > P(A). \tag{5}$$

From elementary probability theory

$$P(A|C) = P(A|BC)P(B|C) + P(A|\bar{B}C)P(\bar{B}|C), \tag{6}$$

which, using (3), we may rewrite

$$P(A|C)\,(1 - P(B|C)) = P(A|\bar{B}C)\,(1 - P(B|C)). \tag{7}$$

Now from the joint hypothesis about B and C, and the definitions of spurious causes and direct causes, we have

$$P(BC),\ P(B\bar{C}),\ P(\bar{B}C) > 0, \tag{8}$$

whence from (7) and (8) we infer

$$P(A|\bar{B}C) = P(A|C), \tag{9}$$

but then from (2) and (9)

$$P(A|C) = P(A|\bar{B}), \tag{10}$$

but since

$$P(A|B) > P(A),$$

we must have, in view of (8),

$$P(A|\bar{B}) < P(A), \tag{11}$$

and so from (10) and (11)

$$P(A|C) < P(A),$$

which contradicts (5). Q.E.D.

Direct causes that have any degree of remoteness in time violate Hume's criterion of contiguity. The existence of such causes has been a subject of debate, dogma, theory, and experimentation in almost every branch of human thought. Within physical theories the idea of direct causes remote in time has not been as prominent as analysis of the problem of action at a distance — with the action being promulgated instantaneously. However, the theory of relativity has made any sharp separation of spatial remoteness from temporal remoteness impossible, because the concept of instantaneous action at a distance is not relativistically meaningful. It should be emphasized, however, that the concept of direct remote causes is quite consistent with relativity; and the definitions given can be modified to become relativistically invariant, although I shall not pursue this technical point here. But as Maxwell pointed out many years ago, most of us are particularly bothered by remote direct causes, once remoteness in both space and time is required by finite speeds of propagation of energy. This is what he said in the final paragraph of his *Treatise on Electricity and Magnetism* (3rd edition, 1892).

> "But in all of these theories the question naturally occurs: — If something is transmitted from one particle to another at a distance, what is its condition after it has left the one particle and before it has reached the other? If this something is the potential energy of the two particles, as in Neumann's theory, how are we to conceive this energy as existing in a point of space, coinciding neither with the one particle nor with the other? In fact, whenever energy is transmitted from one body to another in time, there must be a medium or substance in which the energy exists after it leaves one body and before it reaches the other, for energy, as Torricelli remarked, 'is a quintessence of so subtile a nature that it cannot be contained in any vessel except the inmost substance of material things.' Hence all these theories lead to the conception of a medium in which the propagation takes place, and if we admit this medium as an hypothesis, I think it ought to occupy a prominent place in our investigations, and that we ought to endeavour to construct a mental representation of all the details of its action, and this has been my constant aim in this treatise."

The conceptual viewpoint expressed by Maxwell is so persuasive that most physical theories were field theories rather than action-at-a-distance theories, once electromagnetic concepts came to the fore. Although this is true in physics, it is profoundly not true in many other parts of science, especially any part that emphasizes historical knowledge of the phenomena studied.

It is a widespread scientific dogma that all aspects of historical knowledge can be replaced ultimately and in principle by a sufficiently deep structural knowledge of the current state of the phenomena in question. This is the dogma so well expressed in the famous quotation of Laplace mentioned earlier. The depth of general conviction that this dogma asserts a fundamental truth about the character of the real world is difficult to overestimate.

In terms of distinctions already drawn the matter can be put this way. Within many theoretical or strictly experimental frameworks the existence of direct remote causes will be affirmed with great certainty and assurance; but within the framework of fundamental beliefs about the general character of the universe their existence will be strongly denied.

An example whose general nature is familiar to everyone may be drawn from psychiatry. Almost all psychiatrists, whether or not they are followers of Freud, will assert that early childhood experiences causally determine major aspects of the character and nature of every adult human being. For practical purposes various kinds of knowledge about the past childhood of an adult are regarded as useful in understanding his current behavior and in making qualitative predictions about his future behavior. However, it is unlikely that any psychiatrist believes that a direct unmediated link with the past of

several years earlier is the causal mechanism. Rather, almost certainly everyone believes in at least a vague way that if the current states of the memory, of other parts of the nervous system, and of other organs of the body were known and understood in a theoretical way, then knowledge of the past by other means of observation could be dispensed with. There rightly is considerable skepticism of carrying out, in anything like the framework of contemporary science, the fundamental investigations required for this structural knowledge. But just as rejection of action at a distance was a fundamental tenet of Cartesian physics in the seventeenth century, embraced even by Newton, so is the rejection of remote temporal action a fundamental tenet of almost all contemporary views of the universe.

If the assessment of the extent to which remote temporal action is rejected is approximately correct, it is appropriate to ask why a concept of remote direct causation is needed. The best answer perhaps can be given by looking at the comparable status of the concept of probability in the nineteenth century. Laplace's famous statement on the deterministic character of the universe was made in his treatise on probability, not, as one might expect, in his treatise on celestial mechanics. For Laplace, probability was a tool of analysis and prediction which enabled scientists systematically to take account of their ignorance of complex causes. The concept of remote direct causation is a tool of a similar kind and is essential for practical and scientific analysis of many sorts. Its usefulness will not disappear in the foreseeable future in disciplines ranging from political history to meteorology.

A simple theoretical example of remote direct causation may be found in linear learning theory. For the noncontingent reinforcement schedule with $0 < \theta,\ \pi,\ P(A_{1,1}) < 1$, we may show that for every $m < n$ and for $i = 1, 2$ $E_{i,m}$ is a direct cause of $A_{i,n}$ in the sense of Definition 5. Qualitatively a remote reinforcement has a direct causal effect, but quantitatively an important aspect of the remoteness is that the effect of the remote cause decreases essentially geometrically, where the parameter of the geometric distribution is $1 - \theta$. As an example that is algebraically simple, if $k > h$ then

$$P(A_{1,n}|E_{1,n-1}E_{1,n-2}\ldots E_{1,n-k}) - P(A_{1,n}|E_{1,n-1}E_{1,n-2}\ldots E_{1,n-h}) = (1-\pi)\,[(1-\theta)^h - (1-\theta)^k].$$

In the case of linear learning theory the simple concepts of response and reinforcement used are not structurally rich enough to permit us

to define a memory mechanism in which the current residual effect of past reinforcements might be stored. Looked at another way, the theory is too simple to enable us to define any methods of memory reduction of the past. The unique past $x_{E,n}$ of reinforcements is needed for the most precise predictions of response behavior.

On the other hand, the geometrical fading away of the past suggests an ϵ-type definition of direct cause. A cause is ϵ-direct when its irreducible effect is greater than or equal to ϵ with respect to any later partition. The formal definition requires a change only in the final condition of Definition 5.

Definition 6. *An event $B_{t'}$ is an ϵ-direct cause of A_t if and only if $B_{t'}$ is a prima facie cause of A_t and there is no t'' and no partition $\pi_{t''}$ such that for every $C_{t''}$ in $\pi_{t''}$*

(i) $t' < t'' < t$,
(ii) $P(B_{t'}C_{t''}) > 0$,
(iii) $|\,P(A_t|C_{t''}B_{t'}) - P(A_t|C_{t''})| < \epsilon$.

The relation between Definitions 5 and 6 may be expressed in the following theorem whose proof is obvious.

Theorem 3. *If for some $\epsilon > 0$ an event $B_{t'}$ is an ϵ-direct cause of A_t, then $B_{t'}$ is a direct cause of A_t, but the converse does not hold in general.*

Supplementary causes.

Related to the notion of a direct cause is the concept of two prima facie causes supplementing each other in producing a given effect.

Definition 7. *Events $B_{t'}$ and $C_{t''}$ are supplementary causes of A_t if and only if*

(i) *$B_{t'}$ is a prima facie cause of A_t,*
(ii) *$C_{t''}$ is a prima facie cause of A_t,*
(iii) $P(B_{t'}C_{t''}) > 0$,
(iv) $P(A_t|B_{t'}C_{t''}) > \max\,(P(A_t|B_{t'}), P(A_t|C_{t''}))$.

Note that the definition does not require that times t' and t'' be distinct, although in many cases they will be. A simple example of supplementary causation is provided by linear learning theory. Like reinforcements work together as supplementary causes. Thus, for every trial n, reinforcements $E_{i,n-1}$ and $E_{i,n-2}$ are supplementary causes of $A_{i,n}$.

In analogy with Definition 6, we can also define ϵ-supplementary causes.

Definition 8. *Events $B_{t'}$ and $C_{t''}$ are ϵ-supplementary causes of A_t if and only if*

(i) *$B_{t'}$ is a prima facie cause of A_t,*
(ii) *$C_{t''}$ is a prima facie cause of A_t,*
(iii) $P(B_{t'}C_{t''}) > 0$,
(iv) $P(A_t|B_{t'}C_{t''}) - \max(P(A_t|B_{t'}), P(A_t|C_{t''})) \geq \epsilon$.

In many practical contexts we are interested only in ϵ-supplementary causes, i.e., causes which combine to yield a significantly better quantitative prediction of the occurrence of a given event. Medicine provides many such examples. For instance, in the case of many diseases the quantitative study of the body temperature of patients would no doubt yield statistically significant but not ϵ-significant predictions about the course of the disease.

Sufficient causes.

As a limiting case of the probabilistic analysis given here, we may define sufficient or determinate causes as causes that produce their effects with probability one.

Definition 9. *An event $B_{t'}$ is a sufficient (or determinate) cause of A_t if and only if $B_{t'}$ is a prima facie cause of A_t and*

$$P(A_t|B_{t'}) = 1.$$

Casual inspection of the definition of spurious causes suggests the speculation that in a chain of sufficient or determinate causes only the first member of the chain is a genuine cause, if, indeed, there is a first member. If this speculation were correct, if the universe were without beginning and Laplacean in character, we would be faced with the conclusion that there are no genuine causes. It is not to the point in the present context to deny this paradoxical conclusion by denying that the universe is Laplacean. As I emphasize throughout this monograph, the theory of causality advanced here is not meant to be tailored to the latest physics. It is designed to provide a framework for the analysis of causality in a wide variety of theories and, hopefully, in a way that will usually fit the intuitions about causality that go with a given theory.

The spectre to be laid to rest here is that because

$$P(A_t|B_{t'}) = 1$$

and also

$$P(A_t|C_{t''}) = 1,$$

with $t'' < t'$, we must have a determinate cause like $B_{t'}$ be spurious. The following theorem shows that this can never happen.

Theorem 4. *No sufficient (or determinate) cause can be spurious (in sense two).*

Proof: To make the proof complete, the first thing to prove is a relatively familiar fact about conditional probability.

(1) If $P(A|B) = 1$ and $P(BC) > 0$ then $P(A|BC) = 1$.

Suppose, by way of contradiction, that

(2) $$P(A|BC) < 1.$$

Now from (2) and the definition of conditional probability, we have at once

(3) $$P(ABC) < P(BC).$$

Adding $P(AB\bar{C})$ to both sides of (3) and simplifying we have

(4) $$P(AB) < P(BC) + P(AB\bar{C}).$$

We now take conditional probabilities with respect to B, and divide both sides of (4) by $P(B)$, for by the hypothesis of (1), $P(B) > 0$, and thus we obtain

$$P(A|B) < P(C|B) + P(A\bar{C}|B),$$

but

$$P(C|B) + P(A\bar{C}|B) \leq 1$$

and by hypothesis of (1)

$$P(A|B) = 1,$$

whence we have derived the absurdity that $1 < 1$. Thus (1) is established.

Now assume that $B_{t'}$ is a sufficient cause of A_t, and suppose, contrary to the theorem, that there is a partition $\pi_{t''}$ that renders $B_{t'}$ spurious according to Definition 3. Then by virtue of (1), for every event $C_{t''}$ in $\pi_{t''}$,

(5) $$P(A_t|B_{t'}C_{t''}) = P(A_t|C_{t''}) = 1,$$

but then since

$$P(A) = \sum_{\pi_{t''}} P(A_t|C_{t''})\, P(C_{t''}),$$

we have that

$$P(A) = 1,$$

which contradicts the hypothesis that $B_{t'}$ is a sufficient cause of A_t, and thus that $B_{t'}$ is a prima facie cause of A_t, for the definition of prima facie cause requires that

$$P(A_t|B_{t'}) > P(A_t)\,. \qquad Q.E.D.$$

The most crucial assumption in the proof is the requirement of Definition 3 that for every event $C_{t''}$ in $\pi_{t''}$

$$P(B_{t'}C_{t''}) > 0\,.$$

An omniscient God might object to this aspect of the definition of spurious, but for limited human knowers it seems wholly defensible. What must be recognized, of course, is that in dealing with deterministic classical physics a probability measure must be imposed from outside that framework itself even to apply the definitions. For the classically minded it is easy enough to provide that interpretation in terms of the Laplacean idea that probability is the expression of ignorance. Only a God who knows everything would have a distribution that assigns only probability one or zero to any event, and only such a distribution could never satisfy the conditions of Definition 3.

The causal concepts introduced thus far by no means exhaust the list of essential distinctions, nor have the relations holding between the concepts been pursued as thoroughly as possible. If the direction of analysis that has been started is correct, then what is needed is a work of larger scope replete with scientific examples drawn from many domains.

There is also need for a more systematic general theory of causality than I have been able to set forth here. It is clear that the definitions introduced thus far and related definitions not stated here may be used to classify stochastic processes in terms of their causal properties. For example, any process that is a chain of infinite order will contain direct causes of unbounded temporal remoteness. In any Markov process that is not of zero order, some events must have genuine prima facie causes; but in many continuous-time Markov processes there are no direct causes at all. In other words, from the standpoint developed here the general theory of causality, in its mathematical aspects at least, is a theory about the causal classification and characteristics of stochastic processes.

The definitions of prima facie causes and the like given above correspond closely to what is either explicit or implicit in the investigation of causal relations in the more empirical branches of science. From a philosophical standpoint many different sorts of objections may be raised about the adequacy of these definitions. The more general issues are discussed later. Two objections, however, of a more narrow scope and greater definiteness have been brought to my attention by my students, Edward Bolton and Deborah Rosen.

Concept of occurrence. The first objection is to having as causes or effects events that did not occur. In ordinary contexts, when we say that B is a cause of A, we take this assertion to imply that at the very least B and A actually occurred. In the standard set-theoretical versions of probability theory used as a conceptual framework in the discussion thus far, there is no way to indicate the actual occurrence of an event. Miss Rosen gave the following example to show the reason for insisting on some notion of occurrence. Accepting the current data about the relationship between smoking and lung cancer, we might very well be led to say that John's smoking three packs a day of unfiltered cigarettes in his teen years is a prima facie cause of his getting cancer at the age of 60, but then John may die prematurely in his twenties as the result of an automobile accident. Our causal statement then seems peculiar in terms of ordinary distinctions, although if we replace 'is' by the subjunctive 'may be', and thus say that smoking may be a cause of his getting cancer, the demand for the actual occurrence of either cause or effect is much weakened. I expand on this matter later.

I have remarked elsewhere (Suppes 1966) that this inability to express what actually occurs is a difficulty for Bayesian theories of rational beliefs and rational change of belief. The typical Bayesian attitude is to say that new information is absorbed by conditionalization on the event that occurred. In the article just referred to I try to give some reasons why conditionalization is not a sufficient device. At a logical level there is a different point to be made. It is that the mere ability to pass from a probability measure P to a conditional measure P_A does not provide a method of indicating that the event A rather than, for example, event B occurred. An additional formal apparatus must be added to the standard axioms of probability to express systematically the idea of an event's actually occurring.

Some axioms of occurrence have been given by Domotor (1969) for another purpose, and it is easy to derive his axioms from those given below, but not conversely. It is obvious that the concept of the occurrence of an event is formally similar to the concept of a proposi-

tion's being true. The four axioms given below assume the algebra of events as given in the usual set-theoretical fashion. The new additional concept of occurrence is expressed by a one-place predicate Θ. From the four axioms we can derive Huntington's five axioms (1934) for formalizing the "informal" part of Whitehead and Russell's *Principia Mathematica* (1925). The predicate 'Θ' corresponds to his predicate 'C', where $C(x)$ is interpreted to mean that proposition x is true.

Axioms of Occurrence

Axiom 1. *If* ΘA *then* $\Theta(A \cup B)$.
Axiom 2. *If* ΘA *and* ΘB *then* $\Theta(A \cap B)$.
Axiom 3. ΘA *or* $\Theta\bar{A}$.
Axiom 4. *If* ΘA *then it is not the case* $\Theta\bar{A}$.

I shall only prove two theorems about these axioms.

Theorem 5. *The Axioms of Occurrence imply Huntington's eight axioms.*

Proof: His first three axioms are closure axioms that are immediately satisfied. Since $A \cup B = B \cup A$, we infer at once that if $\Theta(A \cup B)$ then $\Theta(B \cup A)$, which is Huntington's fourth axiom. His Axiom 5 is just Axiom 1 here. His Axiom 6 asserts that if $\Theta\bar{A}$ then it is not the case ΘA, but this is just the contrapositive of Axiom 4 above. His Axiom 7 asserts that if it is not the case $\Theta\bar{A}$ then ΘA, and this is equivalent to Axiom 3 above.

His Axiom 8 may be expressed in the notation used here as follows.

(1) If $\Theta(A \cup B)$ and $\Theta(\bar{A})$ then ΘB.

To derive (1) from the axioms given here it is convenient first to prove:

(2) If $\Theta(A \cup B)$ then ΘA or ΘB.

Suppose the consequent is false. We then have

not ΘA and not ΘB,

whence by Axiom 3

$\Theta\bar{A}$ and $\Theta\bar{B}$,

and thus by Axiom 2

$\Theta(\bar{A} \cap \bar{B})$.

Since $\bar{A} \cap \bar{B} = \overline{A \cup B}$, we can then infer

$\Theta(\overline{A \cup B})$,

which together with Axiom 4 and the hypothesis of (2) yields a contradiction. Thus (2) is established. We may now immediately derive (1). From the hypothesis that $\Theta(A \cup B)$ and (2) we infer

ΘA or ΘB,

and from the hypothesis $\Theta\bar{A}$ by Axiom 3, not ΘA, whence by *tollendo ponens*, ΘB, as desired. Q.E.D.

Huntington proves a large number of theorems from his axioms that will not be discussed here. It is worth remarking, however, that the following four theorems correspond to four of the five axioms for the propositional calculus given by Whitehead and Russell. (Their fifth axiom was derived from these four by Bernays.)

$\Theta(\overline{(A \cup \bar{A})} \cup A)$
$\Theta(\bar{A} \cup (A \cup B))$
$\Theta(\overline{(A \cup \bar{B})} \cup (B \cup A))$
$\Theta[\overline{(\bar{A} \cup B)} \cup [\overline{(C \cup A)} \cup (C \cup B)]]$.

The proofs of these four assertions in the present context are quite simple. They just depend on noting that ΘX follows at once from Axiom 1 and that each of the events, such as $\overline{A \cup \bar{A}} \cup A$, is identical to X.

Because the explicit need for a concept of occurrence is not usually admitted in discussions of the foundations of probability, and because some philosophers might want to try to define the concept of occurrence in terms of causal concepts (a similar sort of thing has been tried — without formal success — for the direction of time), an explicit impossibility proof seems desirable.

Theorem 6. *The concept of occurrence satisfying Axioms* 1—4 *above is not definable in terms of standard probability concepts. Moreover, it is not definable in terms of the causal concepts set forth in Definitions* 1—8.

Proof: The standard formal theory of probability is stated explicitly in the Appendix in Definitions $A1-A17$, but no formal details are needed for the proof. We use Padoa's method for proving the independence of concepts (for an exposition, see Suppes (1957, Chapter 8)). Take any probability space $\mathfrak{X} = (X, \mathfrak{F}, P)$ in which X has more than one element, say a and b. In both models keep X, $\mathfrak{F}$, and P fixed. In one model, for any event A in $\mathfrak{F}$, ΘA if and only if $a \in A$. In the second model for any event A in $\mathfrak{F}$, $\Theta' A$ if and only if

$b \in A$. Obviously both Θ and Θ' will satisfy the Axioms of Occurrence, but $\Theta \neq \Theta'$, and thus Θ is not definable in terms of X, $\mathfrak{F}$ and P.

Because all the causal concepts of Definitions 1—8 are definable in terms of a fixed probability space, by the argument just given these causal concepts also cannot be used to define Θ. Q.E.D.

As part of the definiens we may add to the causal definitions previously given the clause that the two events in question, namely, the cause and the effect, both occurred, although reasons for not doing this are given below.

There is a rather interesting question about the theoretical use of the notion of occurrence. In the kind of theoretical application considered earlier in which the theory is applied to a linear learning model, it might be objected that in such theoretical investigations no notion of occurrence is required. It seems to me that the point is moot, and in any case, no harm is done by introducing into the formal definition the requirement that the cause and effect both occur. In making causal statements in a theoretical framework we may always make them explicitly conditional: if the events in question occur then so and so will be the case. For example, instead of simply saying that $A_{1,n}$ is a prima facie cause of $A_{1,n+1}$, we may wish to make the more restrictive statement: If $A_{1,n}$ and $A_{1,n+1}$ occur, then $A_{1,n}$ is a prima facie cause of $A_{1,n+1}$. It is also convenient to omit the explicit occurrence requirements, even as an hypothesis, and simply say that $B_{t'}$ is a *potential* prima facie cause of A_t. When both events occur the potential becomes *actual.*

I have already remarked on the possible use of the subjunctive rather than the indicative to avoid problems of occurrence. There is much in ordinary usage to sanction this change. Many readers probably would be more comfortable if in the initial definition of prima facie causes (Definition 1), 'is' were replaced by 'may be', or, as another solution, 'cause' were replaced by 'causal tendency'. Still another alternative is to replace 'prima facie cause' by 'potential prima facie cause'. However, the search for an exact match to ordinary usage can be difficult to terminate. For example, in much ordinary usage existence of the events, or at least of the potential cause, seems implied even when the subjunctive is used. A wife says to her husband, "Eating too many clams may be the cause of your stomachache." We immediately infer that indeed the event of eating clams has occurred, and the *may be* refers to uncertainty, or at least politeness, in making a causal claim.

The language of *potential causes* may be the most reasonable formulation, especially if it is usually coupled with interpretation of the algebra of events as kinds of events rather than as particular happenings. The extensive use of kinds or classes of events rather than individual events in many scientific investigations of causal relations is one of the strongest arguments against introducing the concept of occurrence explicitly in the formal definitions. More is said on this point in Section 6.

It is beyond the scope of my endeavor here to push these considerations further, but they certainly do warrant a much more careful analysis than I have given them. Having discussed the logic of occurrence this extensively, I shall subsequently ignore it except for an occasional aside.

Improbable consequences. The second objection raised by Mr. Bolton and Miss Rosen is more fundamental and more important to deal with directly in the standard probability framework. The objection is of the following sort. A course of events may begin, in the middle of which a curious event may occur with improbable consequences. From an intuitive standpoint the curious event that occurred in the middle of the course of events is actually the cause of a subsequent event but not one that we would have admitted as being a cause if we had evaluated the probability of occurrence at the beginning. By the "beginning" I mean at least a time no later than the actual occurrence of the event to which causal powers are are being attributed. There is a tangle of problems here and it will be important to be explicit in the analysis.

To modify slightly Miss Rosen's example, suppose a golfer makes a shot that hits a limb of a tree close to the green and is thereby deflected directly into the hole, for a spectacular birdie. Let the event to be explained, A_t, be the event of making a birdie, and let $B_{t'}$ be the event of hitting the limb earlier. If we know something about Mr. Jones' golf we can estimate the probability of his making a birdie on this particular hole. The probability will be low, but the seemingly disturbing thing is that if we estimate the conditional probability of his making a birdie, given that the ball hit the branch, that is, given that event $B_{t'}$ occurred, we would ordinarily estimate the probability as being still lower. Yet when we see the event happen, we recognize immediately that hitting the branch in exactly the way it did was essential to the ball's going into the cup.

Two different observations may be made about this kind of

example. The first is that $B_{t'}$ can be a prima facie cause of A_t, i.e., $t' < t$ and

$$P(A_t|B_{t'}) > P(A_t),$$

and yet an event $C_{t''}$, with $t' < t'' < t$, may occur such that

$$P(A_t|C_{t''}\,B_{t'}) < P(A_t).$$

A simple numerical example of this phenomenon is the following. Let the conditional probability of A or $\bar{A}$ be given by the matrix

	A	$\bar{A}$
BC	.4	.6
$B\bar{C}$	.9	.1
$\bar{B}C$	.5	.5
$\bar{B}\bar{C}$	.6	.4

and let the conditional probability of C or $\bar{C}$ be:

	C	$\bar{C}$
B	.5	.5
$\bar{B}$	.5	.5

and let $P(B) = .5$. Then it is easy to check that $P(A) = .6$, $P(A|B) = .65$ and $P(A|BC) = .4$. It is important that the occurrence of C and the reduced probability of A occurring does not render B a spurious cause. The difficulties of accurate economic, political or social predictions from an analysis of potential causes are to a considerable extent due to the continued intrusion of unanticipated events. This is just another way of saying that our theoretical analysis of causal structures in these domains is as yet rather superficial.

The second observation is this. The definitions of causal concepts given can easily be relativized to conditional probabilities expressing some background information. Thus Definition 1 could be changed to read: $B_{t'}$ is a prima facie cause of A_t with respect to the information $C_{t'}$ if and only if

(i) $t' < t$,
(ii) $P(B_{t'}C_{t'}) > 0$,
(iii) $P(A_t|B_{t'}C_{t'}) > P(A_t|C_{t'})$.

Such relativization can be useful, especially in theoretical contexts. In practical applications if $C_{t'}$ has occurred, then we may insist that

(1) $$P(A_t) = P(A_t|C_{t'}),$$

from the requirements of coherence on the probability measure P that takes account of this occurrence. This measure P must satisfy the following additional axiom of occurrence:

If θA then $P(A) = 1$.

From this axiom, the theorem on total probability and the assertion that C occurred, (1) follows immediately[1].

This axiom was not stated as part of the general axioms of occurrence, because in theoretical contexts especially the measure P remains the same regardless of what events actually occur. Indeed, if we simultaneously adopted this axiom and the requirement that both A_t and $B_{t'}$ must occur in order to assert that $B_{t'}$ is a prima facie cause of A_t, the probabilistic viewpoint of this monograph would not be applicable, because we would then have $P(A_t) = P(B_{t'}) = 1$. Of course, if A_t and $B_{t'}$ are interpreted as kinds of events, even after the fact the probabilities are not one, but are estimated relative frequencies or posteriori Bayesian estimates, and direct problems of occurrence do not arise.

Negative causes.

In the literature of causality there has been a fair amount of discussion of negative causes. It should be clear how prima facie and genuine negative causes can be defined in purely probabilistic terms. The intuitive idea of a negative cause is that it tends to prevent an event from happening, and this concept can be expressed by little more than the reversal of inequalities in the earlier definitions. Formally, I consider only prima facie negative causes.

Definition 10. *The event $B_{t'}$ is a prima facie negative cause of A_t if and only if*

(i) $t' < t$,
(ii) $P(B_{t'}) > 0$,
(iii) $P(A_t|B_{t'}) < P(A_t)$.

To us an example discussed earlier, inoculation is a negative cause of cholera, or, to put the matter in explicit event-language, the event of being inoculated for cholera is a negative cause of the event of getting

[1] This discussion is related to the problem of total evidence in inductive inference. For a more detailed statement see Suppes (1966).

cholera. To generalize on this example, the theory and practice of preventive medicine concentrates on certain types of negative causation.

Because events form a Boolean algebra, it is easy to give a necessary and sufficient condition on events' being negative causes in terms of their complements' being causes.

Theorem 7. $B_{t'}$ is a prima facie negative cause of A_t if and only if $\bar{B}_{t'}$ is a prima facie cause of A_t.

Proof: Assume first that $B_{t'}$ is a prima facie negative cause of A_t. We note first that (omitting subscripts in the proof)

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}), \tag{1}$$

whence on the assumption that

$$P(A) > P(A|B), \tag{2}$$

we must have

$$P(\bar{B}) > 0, \tag{3}$$

for otherwise we would have $P(B) = 1$, and therefore $P(A) = P(A|B)$. It then follows at once from (1), (2) and (3) that

$$P(A) < P(A|\bar{B}),$$

whence $\bar{B}$ is a prima facie cause of A.

The argument in the other direction is analogous and will be omitted. Q.E.D.

It is perhaps worth noting that if A occurs and if $P(A_t|B_{t'}) > P(A_t)$ with $t' < t$, then either a prima facie cause or a prima facie negative cause of A must occur. Indeed, the occurrence of one of the two does not depend on the occurrence of A_t.

Alternative approaches. This is a good point at which to review some alternative approaches to causality. Donald Davidson has put to me the good question of how to explain that the relation of causality, as it involves probability, is different from the geometrical relation of one object being on top of another, for example. According to this line of argument, the concept of causality in no way depends for its definition or axiomatic characterization on any probabilistic concepts. Probability enters only in inferring from evidence the presence of a causal relation. The answer to this view is already given by Hume in his insistence on constant conjunction of the cause and effect. What has been done here is to generalize this concept of

constant conjunction to a probabilistic relationship. What is important to emphasize is exactly what Hume emphasized. The notion of frequent co-occurrence is at the very heart of the idea of causality, as it is not in the case of geometric concepts. It is not a matter of presenting evidence for causality by offering probabilistic considerations but it is part of the concept itself to claim relative frequency of co-occurrence of cause and effect. When we put the matter this way in terms of relative frequency, we are of course speaking in terms of kinds of events. We can by the usual process reduce this discussion of kinds of events to an analysis of particular events.

It is my own assessment that the relationship of frequent co-occurrence is inextricably part of the notion of cause. There are at least two other viewpoints that have had prominence in the philosophical literature and that need to be considered as alternatives. One is the notion of causal relations being necessary relations, which Hume criticized in such a devastating fashion. In spite of the historical importance of this concept, going back at least to Aristotle, and its reinstitution in various forms by many philosophers, I shall not try to enlarge upon Hume's negative critique.

A more modern and modest approach, generated undoubtedly from the necessary notion of causality but conceptually now distinct from it, is the definition of causes in the context of theory; on this view, the laws of the theory determine the analysis of causes. For example, causal notions in classical physics are determined by the fundamental physical laws of phenomena whether they be mechanical, optical, or electromagnetic in nature.

I do think a case can be made for this lawlike concept of cause in systematic scientific contexts, but I am skeptical that it will work for the less theoretical and more empirically oriented sciences. Perhaps more importantly, as far as I can see there is no reason that the "theory-laden" concept of cause cannot be considered a special case of the probabilistic concept made central in the present analysis. The special position of classical deterministic physics has already been mentioned. The causal analysis of classical physics can be regarded in quite a direct way as a limiting case of probabilistic analysis, one in which all the probabilities are either zero or one. And, as I have already observed, once we turn to the consideration of data in support of these deterministic theories, probabilistic considerations enter at once through errors of observations, and the propagation of these errors in future predictions.

One of the more detailed and sophisticated discussions of the lawlike theory of causality is to be found in papers by Coleman (1966) and Malinvaud (1966) in a volume edited by Herman Wold (1966a). Wold's own contribution to this volume on causality provides a good general discussion of the place of causal notions in current social science, especially economics (Wold, 1966b).

Malinvaud develops an axiomatic theory of causality based on the theory of graphs. He assumes that the causal relation is transitive and antisymmetric; some of his examples do not require that the cause precede the effect in time. In a deterministic setting, the transitivity and antisymmetry are easy to accept, but it seems to me that his explicit assumptions are too weak to serve as a genuine characterization of causality. In spite of the examples that can be given to show the use of causal notions in nontemporal settings, for example, the classical economic laws of supply and demand, it still seems to me that the fundamental concepts of causality imply temporal order, and that Malinvaud has not imposed sufficient structure to yield a satisfactory characterization of causality. Malinvaud does use the general notion of a function to replace the unsatisfactory syntactic notion of a law and he imposes a realistic coherence condition on functional relationships in order to have a causal structure.

Coleman's article presents an excellent general discussion of the mathematical representation of causal relations among continuous variables in continuous time. Let $x_1, x_2, \ldots, x_n$ be n functions of time and let the relationships between these functions be expressed as a system of equations of the form

$$\frac{dx_1}{dt} = f_1(x_1, x_2, \ldots, x_n),$$

$$\frac{dx_2}{dt} = f_2(x_1, x_2, \ldots, x_n),$$

$$\ldots$$

$$\frac{dx_n}{dt} = f_n(x_1, x_2, \ldots, x_n).$$

The analysis of causality that accompanies these equations does not satisfy our restriction on causes preceding effects in time, and there are many classical deterministic analyses of a similar sort. One way of looking at the matter within the framework of this monograph is to regard the instantaneous temporal relation between cause and

effect as a limiting case introduced for purposes of mathematical simplicity. I shall not pursue these matters further here, but the reader is urged to have a detailed look at the three articles just mentioned if he wishes to obtain a more detailed and systematic overview of the lawlike or functional approach to causality. Probably the issue of most importance that is least satisfactorily dealt with in the discussion in this monograph is the view stemming from the functional approach that two events or properties can mutually cause each other. In classical physics and elsewhere it is quite common to see the viewpoint expressed that one variable x affects y and also y in turn affects x without any temporal sequence implied. I also would concede that the issues here are deeper than one simply of an assumption introduced for mathematical simplicity. I do think the insistence on the temporal order of cause preceding effect can be held to, but a detailed analysis of the issues would take us too far afield. It might be mentioned that much of the controversy in recent years in econometrics about the relative virtues of recursive models and interdependent systems, as they are called, centers around this issue. The volume edited by Wold provides an introduction to the controversy and some references into the econometric literature. A related analysis directed toward political science, but fully cognizant of a much broader literature on causality, is the article by Alker (1966), and a useful introductive reference in sociology is the book by Blalock (1961).

One approach that fits in naturally with the differential-equation characterization is to adopt the viewpoint of non-standard analysis (Robinson, 1966) and require that a cause precede its effect at least infinitesimally in time. Conceptually it appears that a natural extension of the definitions given here could be worked out, but I shall not pursue the technical details.

Some further remarks about the functional or lawlike approach to causality in a probabilistic setting are to be found at the end of Section 5.

x \ y	1	2	3
1	p	0	0
2	0	q	0
3	s	0	r ,

with, of course, $p + q + r + s = 1$. Thus, clause (ii) of Definition 15 is satisfied by this distribution if $r \geq (q + r)(s + r)$. On the other hand, regress causality is satisfied only if $s = 0$, for

$$P(X \geq 3 | Y = 1) = \frac{s}{p + s},$$

but

$$P(X \geq 3 | Y = 2) = 0. \qquad \text{Q.E.D.}$$

References

Alker, Jr., H. R. Causal inference and political analysis. In J. L. Bernd (Ed.), *Conference on mathematical applications in political science.* Vol. 2. Dallas, Texas: Arnold Foundation, Southern Methodist University, 1966. Pp. 7—43.

Anderson, J. The problem of causality. *Australasian Journal of Philosophy*, 1938, **2**, 127—142.

Berofsky, B. (Ed.) *Free will and determinism.* New York: Harper & Row, 1966.

Blalock, Jr., H. M. *Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press, 1961.

Boscovich, R. J. *Theoria philosophiae naturalis.* Latin-English edition translated by J. M. Childs, Chicago, 1922.

Burks, A. W. The logic of causal propositions. *Mind*, 1951, **60**, 363—382.

Cohen, M. R., & Nagel, E. *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.

Coleman, J. S. Causal models for qualitative attributes. In H. O. A. Wold (Scientific Organizer), *Model building in the human sciences.* Entretiens de Monaco: Union Europienne D'Editions, 1966. Pp. 241—263.

Davidson, D. Actions, reasons, and causes. *The Journal of Philosophy*, 1963, **60**, 685—700.

de Finetti, B. La prevision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincare*, 1937, **7**, 1—68. English translation in H. E. Kyburg, Jr., & H. E. Smokler (Eds.), *Studies in subjective probability.* New York: Wiley, 1964. Pp. 93—158.

Domotor, Z. Probabilistic relational structures and their applications. Technical Report No. 144, May 14, 1969, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Eddington, A. S. *Fundamental theory.* Cambridge: Cambridge University Press, 1949.

Estes, W. K., & Suppes, P. Foundations of linear models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univ. Press, 1959. Pp. 137—179. (a)

Estes, W. K., & Suppes, P. Foundations of statistical learning theory, II. The stimulus sampling model. Technical Report No. 26, October 22, 1959, Stanford University, Institute for Mathematical Studies in the Social Sciences. (b)

Feller, W. *Probability theory and its applications.* New York: Wiley, 1950.

Ferretti, B. On the possibility of a macroscopically causal quantum-relativistic theory. *Nuovo Cimento*, 1963, **27**, 1503—1505.

Føllesdal, D. Quantification into causal context. In R. S. Cohen and M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science*. Vol. 2. New York: Humanities Press, 1965. Pp. 263—274.

Foot, P. Free will as involving determinism. *The Philosophical Review*, 1957, **66**, 439—450.

Gell-Mann, M., Goldberger, M. L., & Thirring, W. E. Use of causality conditions in quantum theory. *Physical Review*, 1954, **95**, 1612—1627.

Good, I. J. A causal calculus (I). *The British Journal for the Philosophy of Science*, 1961, **11**, 305—318.

Good, I. J. A causal calculus (II). *The British Journal for the Philosophy of Science*, 1962, **12**, 43—51.

Greenwood, M., & Yule, G. U. The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine*, 1915, **8**, 113—190.

Grünbaum, A. *Philosophical problems of space and time*. New York: Knopf, 1963.

Hart, H. L. A., & Honoré, A. M. *Causation and the law*. Oxford University Press, 1959.

Höffding, W. Maszstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 1940, **5**, 181—233.

Hume, David. *A treatise on human nature*. Edited by L. A. Selby-Bigge. Oxford: Clarendon Press, 1888.

Huntington, E. V. Independent postulates for the "informal" part of *Principia Mathematica*. *Bulletin of the American Mathematical Society*, 1934, **40**, 127—136.

Kant, I. *Prolegomena and metaphysical foundations of natural science*. London: Bell, 1883. English translation by Belford Bax.

Kant, I. *Critique of pure reason*. Translated by M. Muller. (2nd ed., revised) New York: Macmillan, 1949.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics*. Vol. 2. *Inference and relationship*. London: Griffin, 1961.

Keynes, J. M. *A treatise on probability*. London: Macmillan, 1921.

Kingman, J. F. C. Ergodic properties of continuous-time Markov processes and their discrete skeletons. *Proceedings of the London Mathematical Society*, 1963, **13**, 593—604.

Kolmogorov, A. *Grundbegriffe der wahrscheinlichkeitsrechnung*. Berlin, 1933. English edition, New York: Chelsea, 1950.

Koopman, B. O. The axioms and algebra of intuitive probability. *Annals of Mathematics*, 1940, **41**, 269—292. (a)

Koopman, B. O. The bases of probability. *Bulletin of the American Mathematical Society*, 1940, **46**, 762—774. (b)

Kraemer, H. C. The average error of a learning model, estimation and use in testing the fit of models. *Psychometrica*, 1965, **30**, 343—352.

Laplace, P. S. *Theorie analytique des probabilités*. Troisième édition, revue et augmentée par l'auteur. Paris: Courcier, 1820.

Lehmann, E. L. Some concepts of dependence. *The Annals of Mathematical Statistics*, 1966, **37**, 1137—1153.

Luce, R. D. On the numerical representation of qualitative conditional probability. *The Annals of Mathematical Statistics*, 1968, **39**, 481—491.

Mackie, J. L. I. Causes and conditions. *American Philosophical Quarterly*, 1965, **2**, 245—264.

Malinvaud, E. Pour une axiomatique de la causalite. In H. O. A. Wold (Scientific Organizer), *Model building in the human sciences*. Entretiens de Monaco: Union Europienne D'Editions, 1966. Pp. 297—302.

Maxwell, J. C. *A treatise on electricity and magnetism*. London: Oxford University Press, 1892. (3rd ed.)

Maxwell, J. C. *Matter and motion*. London: Macmillan, 1920.

Melden, A. I. *Free action*. London: Routledge and Kegan Paul, 1961.

Newton, I. *Principia*. Berkeley, California: University of California Press, 1946. Cajori translation.

Ogburn, W. F. Factors in the variation of crime among cities. *Journal of American Statistical Association*, 1935, **30**, 12—34.

Parzen, E. *Modern probability theory and its application*. New York: Wiley, 1960.

Ptolemy, C. *Tetrabiblos*. English translation by J. M. Ashmand. London: Davis & Dickson, 1822.

Reichenbach, H. *Philosophie der Raum-Zeit-Lehre*. Berlin: Walter de Gruiter, 1928.

Robb, A. A. *Geometry of space and time*. Cambridge: Cambridge University Press, 1936.

Robinson, A. *Non-standard analysis*. Amsterdam: North-Holland, 1966.

Russell, B. On the notion of cause. *Proceedings of the Aristotelian Society*, 1913, **13**, 1—26.

Ryle G., *Concept of mind*. London: Hutchinson University Library, 1949. Pp. 86—87.

Savage, L. J. *The foundations of statistics*. New York: Wiley, 1954.

Schützer, W., & Tiomno, J. On the connection of the scattering and derivative matrices with causality. *Physical Review*, 1951, **83**, 249.

Scott, D., & Suppes, P. Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, 1958, **28**, 113—128.

Shirokov, Yu M. Microcovariance and microcausality in quantum theory. *Žurnal Èksperimentalnoĭ i Teoretičeskoĭ Fiziki*, 1963, **44**, 203—224. (Russian, English summary) trans. as. *Soviet Physics JETP*, 1963, **17**, 140—153.

Simon, H. A. Causal ordering and identifiability. In W. C. Hood and T. C. Koopmans (Eds.), *Studies in econometric method*. New York: Wiley, 1953. Pp. 49—74.

Simon, H. A. Spurious correlation: A causal interpretation. *Journal of American Statistical Association*, 1954, **49**, 467—492.

Simon, H. A., & Rescher, N. Cause and counterfactual. *Philosophy of Science*, 1966, **33**, 323—340.

Suppes, P. Descartes and the problem of action at a distance. *Journal of the History of Ideas*, 1954, **15**, 146—152.

Suppes, P. *Introduction to logic*. Princeton: Van Nostrand, 1957.

Suppes, P. Probabilistic inference and the concept of total evidence. In J. Hintikka and P. Suppes (Eds.), *Aspects of inductive logic*. Amsterdam: North-Holland, 1966. Pp. 198—207.

Suppes, P., & Atkinson, R. C. *Markov learning models for multiperson interactions*. Stanford: Stanford University Press, 1960.

Suppes, P., & Rouanet, H. A simple discrimination experiment with a continuum of responses. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, California: Stanford University Press, 1964. Pp. 317—357.

Suppes, P., & Zinnes, J. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. H. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. I. New York: Wiley, 1963. Pp. 3—76.

van Kampen, N. G. S-matrix and causality condition. *Physical Review*, 1953, **89**, 1072.

Whitehead, A. N., & Russell, B. *Principia Mathematica*. Vol. I (2nd ed.) Cambridge: Cambridge University Press, 1925.

Wigner, E. P. (Ed.) *Dispersion relations and their connection with causality*. New York: Academic Press, 1964.

Wilks, S. *Mathematical statistics*. New York: Wiley, 1962.

Wold, H. O. A. (Scientific Organizer) *Model building in the human sciences*. Entretiens de Monaco: Union Europienne D'Editions, 1966. (a)

Wold, H. O. A. On the definition and meaning of causal concepts. In H. O. A. Wold (Scientific Organizer), *Model building in the human sciences*. Entretiens de Monaco: Union Europienne D'Editions, 1966. Pp. 265—295. (b)

Wold, H. O. Forecasting and scientific method. *Proceedings of an international summer institute, Curia, Portugal, September*, 1966. Lisbon: Centro de Economia e Fiancas, 1967. Pp. 1—65.

Zeeman, E. C. Causality implies the Lorentz group. *Journal of Mathematical Physics*, 1964, **4**, 490—493.

Index of Names