11

Causality in the Social Sciences

Margaret Mooney Marini and Burton Singer†*

1. INTRODUCTION

The words *cause* and *causal* are often used by social scientists. In using these words, most social scientists seek to distinguish causation from association, recognizing that causes are responsible for producing effects, whereas noncausal associations are not. Although causal terminology has been imprecise and has waxed and waned in popularity (Bunge 1979; Bernert 1983), the ideas of agency and productivity which it conveys have continued to be viewed as distinctive and important in social science. Thus, when the word *cause* fell into disfavor in the early part of the twentieth century, sociologists used such synonyms as *forces*, *controls*, and *energies* to capture the meaning of what had been referred to formerly as causes (Bernert 1983).

In all branches of social science, the identification of genuine causes is accorded a high priority because it is viewed as the basis for understanding social phenomena and building an explanatory science. Causal judgments are made to *explain* the occurrence of events, to understand *why* particular events occur. With causal knowledge it is

This paper was written while M. M. Marini was supported by National Institute on Aging grants K04-AG00296 and R01-AG05715 and while B. Singer was supported by NICHD grant R01-HD19226. We are indebted to Richard A. Berk, James J. Heckman, Paul W. Holland, Clark Glymour, and an anonymous reviewer for helpful comments on an earlier draft.

* University of Minnesota.
† Yale University.

often possible to predict events in the future or new observations and to exercise some measure of control over events. It is knowledge of causes that makes intervention for the production of desired effects possible.

Despite the recognized importance of identifying causes, relatively little attention has been devoted by social scientists to considering what causality actually means and how knowledge of causes is acquired. Thinking is usually guided by a natural, or intuitive, idea of causality, leading to attempts to consider directional relationships that are not spuriously determined. However, the "causal" effects estimated in the social sciences often do not provide much causal understanding, not only because the methodology is faulty but also because some of the effects hypothesized fit our natural view of causality better than others.

The purpose of this paper is to consider both ontological and epistemological aspects of the problem of causality in the social sciences. We begin by examining philosophical thinking on the ontological problem, which focuses on the question, What *is* causality: What are causal relationships; are such relationships real; what is the nature of causal laws; what can be a cause? We examine ideas on the meaning of causality to arrive at a more precise understanding that can inform attempts at operationalization. Recently, several statisticians (see Holland [1986] and Holland's paper in this volume for discussion and references) have suggested that methodological issues surrounding the measurement of causal effects should dictate what we think of as a cause. This position, which establishes precise guidelines for valid causal inference in narrowly circumscribed studies and argues that talk of causality outside that framework is misguided, is antithetical to our own. In our view, an understanding of the meaning of causality gives rise to a diversified and flexible research approach, in which subject-matter considerations dictate the kind of evidence that should be sought to establish a basis for causal inference. Often the evidence is observational rather than experimental and is accumulated across multiple studies in multiple settings. Regardless of the research approach taken, the degree of belief in a causal hypothesis depends on the strength of the evidence available to support it.

Consideration of the meaning of causality draws attention to several specific aspects of the concept that have largely unrecognized implications for research practice. First, causal relationships are always identified against the background of some causal field, and specification of the field is critical to interpretation of an observed relationship.

Second, causes are often disjunctions of conjunctions, and failure to consider the conjunctive properties of relationships may lead to failure to detect a causal relationship. Third, causes are of different types, involving, for example, extrinsic determination, intrinsic determination, self-determination, and teleological determination, and different types of causes require different approaches to empirical analysis. Fourth, because much human behavior is purposive, the temporal ordering of behavior or even of behavioral intentions may not be a valid indication of causal direction.

After examining the meaning of causality, we turn to the epistemological, or methodological, problem of causality by considering how we acquire causal knowledge: How do we learn about causal relationships; how do we test causal claims and hypotheses? This problem has been a major concern in all branches of social science, although, as indicated by Cook and Campbell (1979, p. 10), "the epistemology of causation, and of the scientific method more generally, is at present in a productive state of near chaos." We consider this problem by examining the process of causal inference in detail. We discuss the criteria upon which causal inferences are based and operational strategies for building a body of evidence to support such inferences.

We argue that initial ideas of causal relationships are usually triggered by empirical cues and inductive reasoning. We discuss the detection and interpretation of empirical cues, including covariation of various types, temporal plausibility, and contiguity. Because our discussion of the meaning of causality suggests that causes are often conjunctions, after general consideration of covariation we describe a couple of new approaches to detecting conjunctions. The role played by the existing body of relevant knowledge in forming causal hypotheses, including knowledge of the world gained through previous experience with similar empirical relations, is also discussed.

After examining the use of empirical cues and inductive reasoning in the formation of causal hypotheses, we consider the process by which cumulation of a body of evidence leads to the confirmation of causal hypotheses. This process involves demonstrating that an association is consistently observed, that the association cannot be attributed to an alternative explanation, and that there is an identifiable mechanism by which the cause produces the effect. We discuss the general process of induction by which evidence leads to acceptance of a causal

claim and the methodological issues that arise in gathering evidence to support an inductive inference. An important implication of our analysis is that subject-matter considerations play a critical role in identifying the evidence needed to support a causal inference and, therefore, must play a critical role in designing research to obtain that evidence. Although statistical tools also play a critical role in the gathering of evidence, there is no context-free statistical method or set of methods that defines causality. The process of causal inference usually involves multiple studies, which successively increase the degree of belief attached to a causal hypothesis.

To illustrate key aspects of the process of causal inference, we present an example at the end of the paper which focuses on attempts to identify an effective treatment for the rehabilitation of heroin addicts. It illustrates the importance of specifying the causal field and considering the possibility that a cause may be a disjunction of conjunctions. It also illustrates the way in which evidence from multiple studies, most of which are observational, can be combined to provide a basis for choosing between competing theories.

2. WHAT IS CAUSALITY?

2.1. *Causal Criteria*

Because philosophers have debated the meaning of causality for centuries and continue to do so, there is no universally accepted definition of causality. At one time, attempts to define causality focused on the idea of logical necessity, and causes were seen as logically entailing their effects (Ducasse 1966). An effect was considered to be "necessitated" by a preceding cause in accordance with the laws of logic and mathematics that establish relations among ideas rather than empirical entities. By the logical relation of implication, the truth of a conclusion is necessitated by the truth of given premises, and there is a contradiction if the premises are true and the conclusion is false.

In one of the most important contributions to our understanding of causality, Hume ([1739] 1896, [1740] 1938, [1748] 1900) set out to show that causes do not logically entail their effects and that the "idea" of causation arises from the *empirical* relations of contiguity, temporal succession, and constant conjunction—i.e., when empirical objects are contiguous in time and space, when one follows the other, and when

they always appear together. Influenced by a general empiricist thesis, Hume argued that a relation of necessity is not something that is perceived empirically but that exists only in the mind. The constant conjunction of empirical entities establishes a mental association between them, leading to a feeling of inevitability when one moves in thought from a cause to its effect. It is this subjective feeling projected upon the empirical world that constitutes the "necessary connection" that Hume considered an essential criterion of causation. Although few later philosophers would defend Hume's argument exactly as he proposed it, the idea that causation involves regularity in relations between empirical entities is widely accepted. Those who accept this criterion as a *sine qua non* are often referred to as "regularity" theorists (Beauchamp 1974).

One problem with Hume's argument is that a causal relation cannot be recognized without a backlog of relevant experience. Since *individual* causal relatedness cannot be perceived, a causal relation cannot be detected in a single case.[1] This problem was addressed by later regularity theorists who claimed that singular causal statements were derivable from causal regularities, or laws, of which they were instances (Beauchamp 1974). Thus, even in the analysis of historical events, which in their full particulars are unique, or unrepeatable, it was argued that a degree of causal understanding was accomplished by reference to general laws (Hempel 1942). An individual sequence was regarded as causal if it could be considered an instance of a general law, which might be known or unknown. As Popper ([1959] 1972, app. X) speculated, "One might suppose that it is this logically necessary dependence upon true statements of higher universality, conjectured to exist, which suggested in the first instance the idea of 'necessary connection' between cause and effect." Because of the logical connection required between individual and general causal statements, regularity theorists emphasized the proper analysis of causal laws. Laws were seen as true, contingent, and universal generalizations. Talk of "necessity" arose because natural laws were used as premises

[1] Surprisingly, Hume ([1739] 1896, p. 104) admits that "not only in philosophy, but even in common life, we may attain the knowledge of a particular cause merely by one experiment, provided it be made with judgement, and after a careful removal of all foreign and superfluous circumstances." The relationship of this statement to his argument that constant conjunction is an essential criterion for causation is unclear.

for inference. Although these laws implied necessary connections between antecedent and consequent events, such connections were seen as gratuitous (Beauchamp 1974). Other theorists have rejected the view that individual causal statements are derivable from causal laws. Mackie (1974), for example, argues that singular causal statements are *prior to* general ones and are supported by the assumption that a singular causal sequence is an instance of some perhaps as yet unknown or unsuspected regularity.

Another criticism of regularity analyses of causation is that not all regularities of sequence express causal relations. The following of night by day and of day by night, the regular motions of the planets, the occurrence of hair growth on babies before the growth of teeth are examples of noncausal regularities. To distinguish these regularities from those expressing causal relations, arguments have been advanced to permit a distinction between factual statements expressing constant conjunctions, or "accidental universals," and nomological generalizations, or universal laws, which express empirically necessary connections (Kneale 1950, 1961; Popper [1959] 1972; Beauchamp 1974). The principle approach to capturing this idea of "natural necessity" has been the requirement that causal laws sustain counterfactual conditional statements (Kneale 1950, 1961). Because this requirement "connects the notion of natural law with that of the validity of states of affairs other than the actual" (Kneale 1961, p. 99), it is seen as distinguishing causal laws from accidental generalizations. However, reduction of the distinction to purely syntactical considerations relating to the form of lawlike statements has been widely criticized (Ayer 1956; Nagel 1961; Mackie 1966, 1974). Mackie (1966, 1974) further explicated the meaning and use of counterfactuals by arguing that counterfactuals must not be construed as statements with truth values or as statements that follow logically from other statements. Rather, they reflect imagined situations and have the form of condensed arguments that are only entertained and not argued. One is justified in advancing them only if the beliefs that support them are justified—i.e., if there is relevant *empirical evidence*. In the case of law-governed counterfactuals, the problem then becomes that of deciding when there is evidence to support an inductive generalization—the general problem of induction.

When causal relations are characterized as counterfactual relations, "X caused Y" means that "X occurred and Y occurred and in

the circumstances Y would not have occurred if X had not" (Lewis 1973; Mackie 1974). This concept of causation has been prominent in the singular judgements of causation required in criminal law. For example, in the case of a crime defined in terms of harmful results, the prosecution must show that the defendant's act was the "cause in fact" of the harm, and for an act to be a "cause in fact" of the result, it must be the "but for" antecedent of the result. This means that "if the result *would have happened anyway*, even had the act not occurred, the act is *not a cause in fact* of that result" (Emanuel 1979, p. 41). Consider the following example:

> D shoots at V, but only grazes him, leaving V with a slightly bleeding flesh wound. X then comes along and shoots V through the heart, killing him instantly. D's act is clearly not a "cause in fact" of V's death, since V would have died, and in just the manner he did, even if D had not shot him. (Emanuel 1979, p. 41)[2]

When causal relations are characterized as counterfactual relations, a cause is something that is both necessary and sufficient in the circumstances for the production of its effect (Bunge 1979; Beauchamp 1974; Mackie 1974). A cause is sufficient for the production of its effect when "sufficient in the circumstances" is taken to mean "given the circumstances, if X occurs then Y will." However, this sufficiency criterion is met by any sequence in which X and Y actually occurred and therefore does not distinguish causal sequences. Mackie (1974) has argued that it is only if "sufficient in the circumstances" is taken in the strong counterfactual sense to mean that "if Y had not been going to occur, X would not have occurred" that causal sequences can be distinguished from noncausal sequences. A cause is usually, although not always, sufficient for its effect in this strong counterfactual sense, whereas this relation does not hold for noncausal sequences.

In defining causal relations, the phrase "in the circumstances" is used to indicate that causal statements are usually made in some

[2] Attention has also been given to considering the assignment of responsibility in more complex cases involving alternative over-determination, or "fail-safe" causes (see, e.g., Hart and Honore 1984; Mackie 1974, p. 44–46).

context. As Mackie (1974, p. 34) states, causal statements are made

> against a background of some *causal field*.... Both cause and effect are seen as differences within a field; anything that is part of the assumed (but commonly unstated) description of the field itself will, then, be automatically ruled out as a candidate for the role of cause.

Thus, although there may be a set of factors that are jointly sufficient and severally necessary to produce a result, we are more willing to say that an event, particularly one that is seen as intrusive, caused the effect than that either a standing condition or an event that occurs within some going concern did. For example, we are more likely to say that a spark rather than the presence of flammable material caused a fire, that the severing of an artery rather than the pumping of the heart caused a loss of blood, or that divorce rather than the division of labor within marriage caused a woman to be poor. These preferences are sometimes related to conceptions of what is normal, right, and proper, since what is viewed as abnormal or wrong is more likely to arouse causal interest. Such preferences do not reflect the meaning of causal statements but the uses to which causal statements are put. Because judgments of causal relevance reflect the degree to which a variable is a "difference in a background" (Einhorn and Hogarth 1986), the determination of causal relevance depends critically on context. For example, if we are told that a watch face has been hit by a hammer and the glass breaks, we tend to assume that the hammer caused the glass to break. However, if we are told that the same event occurred during a testing procedure in a watch factory, we tend to view a defect in the glass as the cause of the breakage. The breadth of the causal field determines whether specific alternatives are ruled in or out and thereby affects the number and salience of alternative explanations. In a recent article on statistics and causal inference, Holland (1986) shows awareness of the important role of a causal field or background when he argues that the effect of a cause, X, is always measured relative to other causes, including not X, or $\overline{X}$.

A single cause rarely, if ever, produces an effect. Usually a plurality of causes is involved in two ways. As described by Bunge (1979), *conjunctive plurality of causes* occurs when various factors, sym-

bolized by A, B, C, etc., must be there jointly to produce an effect, Y. Thus, in field Z, Y occurs whenever some conjunction of A and B and C, symbolized as ABC, occurs, but Y does not occur when only the conjunction of A and B, symbolized as AB, occurs. *Disjunctive plurality of causes*, which is often identified as genuine "multiple causation," occurs when the effect is produced by each of several factors alone, and the joint occurrence of two or more factors does not alter the effect (Bunge 1979; Mackie 1965, 1974; Skyrms 1980). Thus, if it is not only the case that, in field Z, the conjunction ABC is followed by Y but also that DEF is followed by Y and that GHI is followed by Y, we have a disjunction of conjunctions: "In Z, (ABC or DEF or GHI) is followed by Y" (Mackie 1974). In Z, (ABC or DEF or GHI) is a condition that is both necessary and sufficient for Y, whereas each conjunction alone is sufficient but not necessary for Y. Using the notation of Einhorn and Hogarth (1986), if we label the conjunction ABC in Z as α, that is, $\alpha = (A \cap B \cap C|Z)$, α is a minimally sufficient condition for Y if

$$p(Y|\alpha) = 1, \quad \text{but } p(\alpha|Y) \neq 0,1.$$

In other words, given α, Y always follows, but given Y, α may or may not occur—it is neither certain nor impossible. However, α would no longer be sufficient if any of its conjuncts, A or B or C, were not present. Thus, a single factor, such as A, is necessary but not sufficient for α:

$$p(A|\alpha) = 1, \quad \text{but } p(\alpha|A) \neq 0,1.$$

Since α is itself sufficient but not necessary for Y, A is neither necessary nor sufficient for Y but is what Mackie (1974, p. 62) has labeled an *inus* condition: "an *insufficient* but *nonredundant* part of an *unnecessary* but *sufficient* condition" (the term *inus* being derived from the first letters of the italicized words).

Since most effects result from multiple causation, what we typically refer to as a cause is an inus condition. For example, if experts investigating the cause of a house fire conclude that an electrical short circuit (X) caused the fire (Y), they are not saying that X was a necessary or sufficient condition for Y. They know that smoking in bed, the overturning of a lighted oil stove, or any one of a number of other events, if it had occurred, might have set the house on fire. They also know that it was the short circuit conjoined with a particular set of

conditions (e.g., flammable material near the short circuit, no sprinkler system, etc.) that actually led to the fire.[3] Thus, when we say that X caused Y, we rarely mean that X is either necessary or sufficient for Y. What we mean is that X, when conjoined with other factors, leads to Y. Specifically, we mean that "(a) X is an inus [condition] for Y; (b) X occurred; (c) the other conjuncts occurred; and (d) all minimally sufficient conditions for Y not having X in them were absent on the occasion in question" (Einhorn and Hogarth 1986, p. 7). For X to be a necessary cause of Y, it would have to be in all minimally sufficient conjunctions that produce Y. For X to be a sufficient cause of Y, it would always have to conjoin with its conjuncts. Because the traditional notions of necessity and sufficiency in causation pertain to complex scenarios, often involving a disjunction of conjunctions that we rarely, if ever, know fully, our elliptical understanding of these scenarios results in the observation of probabilistic regularities between identifiable "causes" and their effects.

In the social sciences our interest focuses on scientific theories pertaining to classes of events or things. We are therefore usually interested in the identification of what we might call a causal structure, as reflected in the disjunctive plurality of causes that may produce an effect. In discussing evolutionary theory, Sober (1984) suggests that these disjunctive properties be referred to as explanations rather than causes, since they are not themselves causally efficacious. Disjunctions identify a plurality of causes, any one of which may produce the effect, but do not pinpoint the actual cause that produced the effect in a particular case. Sometimes a concept representing a disjunctive plurality of causes plays an important role in formulating general patterns of explanation. For example, Sober (1984) argues that the concept of the overall fitness of an organism, which summarizes the chances of mortality due to all possible causes, including ones that will not actually be the cause of death, plays an explanatory but not a causal role in evolutionary theory.[4] The concept of overall fitness resembles the concept of life expectancy familiar to social scientists and epidemiologists. If an individual is identified as having high susceptibility to

[3] A version of this example was first cited by Mackie (1965), and other versions have been used widely since then.

[4] By comparison, the concept of selection for properties of organisms plays a causal role, since "selection for a given property means that having that property *causes* success in survival and reproduction" (Sober 1984, p. 100).

several causes of death and dies shortly thereafter, this information offers some explanation of why the individual died but does not single out the actual cause of death. It may be irrelevant to know which of several possible causes produces an effect if one is interested only in calculating the effects of the present state of an organism or system on its future.

Because complex regularities are seldom, if ever, fully known, we are usually in a position to formulate only incomplete propositions reflecting them, from which inferences can be made *with probability* about the relation between a cause, which is actually an inus condition, and its effect. Although such causes are not sufficient for their effects, the causing itself is not necessarily probabilistic. The statistical regularity we perceive may indicate that A is likely to be, in some particular case, necessary in the circumstances for Y, rather than that A is likely to necessitate Y (Mackie 1974). Thus, an incomplete causal generalization can sustain the counterfactual conditionals involved in a singular causal judgement. The generalization may be probabilistic when the relation in any particular case that fulfills it is one of necessity, and perhaps sufficiency, in the circumstances. Although even complex regularities involving disjunction of conjunctions may be necessary for Y and therefore deterministic rather than probabilistic, it is unknown whether strict determinism ever holds, since all laws have statistical features when framed in operational terms because of measurement error. One problem with *theories* of probabilistic causality is that although only statistical regularities are observed, causation on the ontic level may not be probabilistic (Salmon 1980).

Since the introduction of quantum mechanics in physics at the beginning of the twentieth century, the idea that at least some aspects of nature are, in fact, irreducibly probabilistic has gained wide acceptance (Sober 1984; Crutchfield et al. 1986; Kolata 1986). Causal laws are probabilistic or statistical if there are sequences of events that fulfill them and if the proportionality of the outcome—e.g., that A's produce Y's in x percent of cases—is observed repeatedly when the causal conditions occur. A statistical law involves the assignment of a measured probability or chance to each individual that falls under it—i.e., to each instance of the relevant assemblage of conditions. It is of the form, "Every A has an x percent chance of becoming (or producing) Y" (Mackie 1974). Statistical laws reflect constancies in nature, but what is constant is the proportionality of the outcome.

Regardless of whether causal laws on the ontic level are deterministic or probabilistic, we observe probabilistic regularities, involving covariation but not necessary connection between X and Y. Rather than indicating that if X occurs, Y must occur, empirical regularities indicate that if X occurs, the probability of Y occurring increases (or changes). As noted above, this fact has led to the formulation of probabilistic theories of causality, which actually define causal relations probabilistically. Thus far, these formal theories have been found to have serious flaws when applied to some empirical situations (Reichenbach 1956; Good 1961, 1962; Suppes 1970; Skyrms 1980; Salmon 1980).

According to philosophers, necessity in the circumstances, as reflected in empirically supported counterfactuals, is not the only distinguishing feature of a cause. Because causal relations are commonly recognized to reflect some genetic connection through a process, they are also asymmetrical, or directional (Bunge 1979; Mackie 1974). According to Mackie (1974, p. 180), this idea of *causal priority* is best captured by the notion of fixity: "An effect cannot be fixed at a time when its cause is not fixed." Or, as Bunge (1979, p. 63) states, "The causal principle requires that the cause be there if the effect is to occur." The notion of fixity bears a strong relationship to temporal succession, which Hume ([1739] 1896, [1740] 1938, [1748] 1900) regarded as an essential criterion for causation. However, causal priority is something more than temporal priority and can be variously directed with respect to time. Modern philosophers tend to agree that temporal priority is not required for causation, although it is usually consistent with a cause's being there before its effect occurs (Bunge 1979; Mackie 1974).

One reason temporal priority is not required for causation is that causation has been conceived to involve contemporaneous links; i.e., causes can occur simultaneously with their effects. For example, Kant considered the case of a leaden ball resting on a cushion and causing a hollow. This example shows how causal priority is established between simultaneous cause and effect when these are continuations of events between which there is a temporal sequence. Another example of simultaneous causes and effects is a Newtonian gravitational set-up in which the acceleration of a body at a point in time causally depends on the masses and distances of other bodies at the same point in time

(Mackie 1974). An example of simultaneously occurring cause and effect in the social sciences is the occurrence of poor mental functioning as a result of depression. Being in a state of depression at a point in time causes poor mental functioning at the same point in time. Backward causation, in which the effect precedes its cause in time, has also been proposed and has become a subject of extensive debate among philosophers (Brier 1974; Mackie 1974, p. 162). An example of backward causation is direct precognition. A literal foreseeing of future events would mean that the future event was affecting the precognizer. Although contemporaneous and backward causation are rare, if they exist at all, the concept of causal priority, or fixity, allows for these possibilities, whereas temporal priority does not. As a practical matter, causal priority usually involves forward causation, or temporal succession.

Skyrms (1980) observes that with the exception of some elementary particle interactions, none of the basic theories in the physical sciences is time-asymmetric. The familiar time asymmetries of physical macrophenomena are represented by the repeated operation of the basic theories of electrodynamics and statistical mechanics, which produce changing states in an ongoing process. For example, if one throws a pebble into a still pond, waves of water radiate coherently out to shore. The temporal inverse of this process is permitted by theory but is never observed. A similar situation may pertain in many social science settings in which basic theories (e.g., decision theories) are time-symmetric, but asymmetries in conditions produced by ongoing processes result in time-asymmetric behavior patterns. For example, after one enters the labor market, wages increase rather than decrease over the life course. The inverse of this process is permitted by basic theories that are time-symmetric, but age-related asymmetries in the actual conditions affecting the decisions of workers and employers make such an inverse unlikely.

In general, the language of causation is more likely to be used when causal laws are molar, or stated in terms of large or complex objects. These laws usually involve delayed causation, mediated via causal chains that operate through time. Molar causal laws are particularly likely to involve disjunctions of conjunctions of the type described above. As a result, the observation of molar relationships tends to be contingent upon many conditions. Until these conditions are more fully

known, molar causal laws will be highly fallible and, hence, probabilistic.

> It is probably the case that the more molar the causal assertion and the longer and more unspecified the assumed micromediational causal chain, the more fallible the causal law and the more probabilistic its supporting evidence. (Cook and Campbell 1979, p. 33)

In the social sciences causal generalizations are difficult to make and, when made, are highly probabilistic because most attempts at causal explanation have been of this molar type.

By comparison, ahistorical, more microlevel field theories that directly mediate delayed effects have been dominant in the physical sciences. These "micromediational" laws specify causal connections at a smaller level of particles and on a finer time scale that is often instantaneous. The focus in these theories is on variables in the present. Although delayed causes may exert effects through these immediate influences, the historical forces that impinge on the present are neglected. The mediational processes that operate through time are made explicit by focusing on influences at a given point in time that impinge directly on the persons or objects under study. As noted above, social science examples of this type of ahistorical, instantaneous, micromediational theory can be found in theories of purposive action, or decision making.

The idea of causal priority has also been argued to be captured by the notion of manipulability. A causal relation is seen as one in which Y is produced by the manipulation of X (Collingwood 1940; Gasking 1955; von Wright 1971). Agents bring about certain ends by manipulating means to those ends, or by creating circumstances that are conditions productive of effects. Thus, a causal relation depends on the concept of action; a causal connection can be distinguished from a noncausal regularity only if manipulation of one factor can bring about another. However, because there are some circumstances in which nothing can be manipulated, it is argued that it is enough that we *assume* that if we *could* manipulate X, we could bring about Y (von Wright 1971). Thus, thought experiments in which we manipulate putative causes that cannot be manipulated at will, if at all, play an

important role in the social sciences. In writing about economics, Haavelmo (1944, p. 6) states,

> When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved we have *in mind* [emphasis added] some actual *experiment*, or some *design of an experiment*, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes.

Pratt and Schlaifer (1984) have also noted that meteorologists subject even the Rocky Mountains to "conceptual manipulation" when they assert that there would be more snow in Denver if the Rocky Mountains were lower.

In addition to necessity in the circumstances, which may be reflected in complex regularities that, in their elliptical form, constitute part of what we know as causation in the empirical world, and causal priority, which is reflected in the ideas of fixity and manipulability, the existence of an underlying causal mechanism, or continuity of process, may distinguish causal sequences. A number of philosophers have argued that because basic laws are, in part, forms of persistence, causal sequences governed by laws are processes that have some qualitative or structural continuity (Russell 1948, 1959; Kneale 1949; Mackie 1965, 1974; Salmon 1984). Thus, it is suggested that to provide the *connection* between cause and effect which Hume called "the cement of the universe," we must analyze causal relations not only in terms of an event that constitutes the cause and an event that constitutes the effect but also in terms of a causal process that connects the two events and explains their relationship.

In describing what he termed a "causal line," Russell (1948, 1959) focused attention on the importance of space-time structure, "which often remains constant, or approximately constant, throughout a series of causally connected events" (1959, p. 198). He regarded a causal line as the persistence of something:

> Throughout a given causal line, there may be constancy of quality, constancy of structure, or a gradual

> change of either, but not sudden changes of any considerable magnitude. (Russell 1948, p. 459)

A causal line may describe the persistence of objects and self-maintaining processes that do not involve qualitative change, or it may describe the persistence of structure throughout changes in quality. Russell identified examples of causes and effects with similar structure that differed in intrinsic quality; for example, he noted that there are qualitative differences but structural similarities among what is printed in a book, the noises made by someone reading aloud from the book, the noises heard by someone else listening to the reader, and the words written down as dictation by the listener. Kneale (1949) argued that even processes that on the face of it exhibit no structural similarity can be assumed to involve some continuity and persistence in structure that cannot be observed but that are expressible in the language of mathematics. At the perceptual level, a cause may be followed by an utterly different effect, such as when a match is struck and a flame appears. But if this macroscopic picture were replaced by a detailed picture of molecular and atomic movements linked by an adequate physicochemical theory to the perceived process, it is argued that more continuity and persistence would be found (Mackie 1974).

Salmon (1984, p. 179) proposes that we take processes rather than events as the basic entities, viewing causal processes as "the means by which structure and order are *propagated* or transmitted from one space-time region of the universe to other times and places." This concept of spatio-temporal continuity is related to Hume's ([1739] 1896, [1740] 1938, [1748] 1900) claim that contiguity in space and time is a criterion of causation,[5] since contiguity makes qualitative or structural continuity possible. What distinguishes causal processes from other processes is that causal processes are capable of transmitting their own structure and, therefore, certain modifications in that structure (Salmon 1984). Because causal processes are self-determined and independent of what goes on elsewhere, they transmit uniformities of

[5] Hume, like modern philosophers, did not require that every cause be contiguous with its effect but that when this was not the case, cause and effect would be viewed as joined by a chain of intermediate factors, where each item was the effect of a prior item and the cause of a succeeding item and was contiguous with both (Mackie 1974). However, Hume hesitated to include even this broad view of contiguity as a criterion of causation.

structural and qualitative features. By transmitting their structure, they are "capable of propagating a causal influence from one space-time locale to another" (Salmon 1984, p. 155). An intervention at a particular point in the process transforms it in a way that persists from that point on. Evidence supporting laws that reflect this partial persistence justifies the use of the counterfactuals that seem to sustain them.

2.2. *Types of Causes*

Elaborating on scattered ideas of Plato, Aristotle provided the first codification of the word *cause* in his *Physics* (Book 2, chap. 3, 194b). He identified four causes that were responsible for the production of an effect: formal, material, efficient, and final. A formal cause contributed the essence or quality of a thing and could be thought of as that *into which* a thing was made. A material cause was the matter or substrate *out of which* the thing was made. An efficient cause was the motive force *which made* the thing, and a final cause was the purpose or goal *for which* the thing was made. This view of causes predominated until the Renaissance, when the birth of modern science focused attention exclusively on Aristotle's efficient cause (Bunge 1979; Kuhn 1977). Formal and final causes were abandoned because they were outside the bounds of experiment, and material causes were taken for granted in all natural phenomena. The concept of cause in science therefore came to be equated with *efficient cause*. Accordingly, in his analysis of causality in modern science, Bunge (1979) defines causation as "determination of the effect by the efficient (external) cause" (p. 17) and restricts the meaning of cause to "extrinsic motive agent, or external influence producing change" (p. 33). When the concept is used in this narrow sense, activity and productivity are inherent in causation. However, causal determination in this narrow sense is only one of a number of types of determination that figure prominently in science, and it is the broader principle of determinacy, or lawful production, that now occupies the place once held by the principle of efficient cause (Bunge 1979). In considering the concept of cause in physics, Kuhn (1977) describes how the concept of efficient cause predominated during the seventeenth and eighteenth centuries but gave way to a broader concept in the nineteenth and twentieth centuries. Now, use of the concept of cause focuses generally on explanation: "To describe the cause or causes of an event is to explain why it occurred" (Kuhn

1977, p. 23). However, despite this broadening of the concept, the narrow concept of efficient (external) cause is sometimes taken as fundamental or regarded as distinct. Thus, in his recent article on statistics and causal inference, Holland (1986) uses the term *cause* to refer only to extrinsic determination.

Recognition of the importance of intrinsic, or internal, determination and its causal role in the production of effects is evident in discussions of interaction between external influences and internal processes. When the concept of cause is restricted to external causes, extrinsic agents tend to be seen as molding a passive lump of clay. Now, it is more common to view extrinsic agents as interacting with inner processes, which reflect the influence of predispositions, preconditions, and immanence or heritability (Bunge 1979; Rothman 1976; Koopman 1977; Cox 1986). Thus, a synthesis of intrinsic and extrinsic determination has provided a more adequate picture of causal relations.

Although the concept of cause has been broadened to include internal as well as external causes, philosophers and scientists alike maintain that the concept has limited usefulness when applied to some types of internal processes. A failure to understand these limits has proved to be a serious problem in social science. One type of relation to which application of the concept of cause is usually not helpful is the relation between an earlier and a later state of a *continuous self-maintaining process*. In such a process there is an unfolding of states that differ from one another only in quantitative respects that can be characterized as self-determination. Although it is possible to argue that in the absence of any constraint or external force, the earlier state is necessary and sufficient for the later state, and is therefore a "cause" of the later state, we are unlikely to view the earlier state as anything more than an intermediate cause. When there is an ongoing process, it is of primary interest to look for an initiating or sustaining cause of the *process*. When additional determiners, such as constraints and external forces, are present, interest focuses primarily on these as causes of a change in the process, and the earlier state of the process is relegated to the "causal field." The earlier state of an ongoing process is not a satisfactory cause of the later state because such states (qualities, dispositions) have no productive capacity relative to one another. Although one state in the process can be an antecedent of another, it

cannot act upon another state.[6] A state in an ongoing process can only be an *outcome* of inner processes or external influences or both. Thus, an earlier state is more appropriately viewed as a boundary condition than as a cause of a later state. As Mackie (1974, p. 156) states,

> It is especially where we are inclined to ask why-questions that we will be ready to accept because-statements as answers to them, and we are particularly ready to call causes those items that can be described by clauses introduced by "because."

If we ask a why-question about the state of a process at t_1, we are unlikely to accept a description of the state of the process at t_0 as a satisfactory answer. For the same reason, we are also unlikely to accept the mere passage of time as a satisfactory answer.

Another type of determination that can be argued to fit within a broad conception of the meaning of *cause* but is unlikely to provide satisfactory answers to why-questions is *teleological determination*. Teleological determination, or what Aristotle would have called a final cause, is determination of the means by the ends, or goals. Since much human behavior is directed to the satisfaction of goals, intentions, or motives, conscious purposive human action falls into this category. Teleological determination therefore plays an important role in social science. An actor believes that an action, X, will bring about a result, Y, and this belief coupled with the actor's wanting or intending to bring about Y causes the actor to do X. As Mackie (1974) argues, teleological determination has the distinguishing features of a causal relation but is additionally characterized by a subjective or relative point of view. A goal or an intention or a motive may be considered an intermediate cause of behavior, but, again, we are unlikely to accept such a cause as a satisfactory answer to why the behavior occurred.

[6] Our statement that a state in an ongoing process is not a satisfactory cause differs from Holland's (1986) far more sweeping claim that "attributes," such as sex and race, cannot be causes. Like most of those commenting on Holland's article (Glymour 1986; Granger 1986; Rubin 1986), we see no reason to exclude attributes as causes.

3. BASES OF CAUSAL INFERENCE

Having considered the meaning of causality in general and as reflected in different types of causes, we now turn to the epistemological question of how we acquire causal knowledge. Since Hume ([1739] 1896, [1740] 1938, [1748] 1900) first argued that causation is an "idea" inferred from observed relations between empirical objects, it has been recognized that empirical evidence constitutes the basis on which inductive inferences of causation are made. The process of induction involves the use of empirical cues in the initial conception of causal hypotheses, laws, and theories and the cumulation of a body of evidence to confirm them, or justify their acceptance. Induction is an inferential process that expands knowledge in the face of uncertainty, involving reasoning from a part to a whole, from the particular to the general, or from the individual to the universal (Holland et al. 1986).

3.1. *Empirical Cues*

The first idea of a causal relation is derived from empirical cues. Observational criteria that follow from the meaning of causality lead us to consider the possibility that a causal relation exists. Empirical cues suggest that a causal inference may be justified, although none constitutes indisputable evidence for or against the causal hypothesis, and none is a *sine qua non*. A subjective element in judging the evidence from empirical cues is an unavoidable aspect of causality assessments.

In *A Treatise on Probability* (1921), Keynes considered the possibility of characterizing induction within a mathematical theory of probability, where probability was interpreted as "degree of belief." He was concerned with the question of when observed instances warrant acceptance of a generalization. One of his conclusions was that before examining the evidence for or against a suggested generalization, a nonzero probability (degree of belief) must be attached to the generalization on the basis of prior knowledge. As discussed by Russell (1959), this prior knowledge comes from empirical cues, which lead to inferences that are not logically demonstrative. "Nondemonstrative inference" differs from inference based on deductive logic in at least two ways. First, when the premises are true and the reasoning correct, the conclusion is only probable. Second, the premises are often uncer-

tain. Thus, empirical indications of causality lead to the formation of causal hypotheses, to which varying degrees of "credibility" or "doubtfulness" can be attached. In the social sciences, these empirical cues include covariation of various types, temporal plausibility, and spatio-temporal contiguity.

Covariation

Our natural conception of causality derives from a program in the brain that "allows the forecast that sequences of events that have been repeatedly connected in the past will (probably) be connected in the future" (Young 1978, p. 234). Such a conception of causality is derived relative to the degree of detailed knowledge available about the empirical phenomena under investigation and can be established with varying degrees of probability. The existence of covariation between X and Y is consistent with Hume's concept of causality and with several early methods of experimental inquiry identified by Mill (1843) for the study of causation. Covariation is also viewed as an important cue to causality by modern statisticians (Suppes 1970; Mosteller and Tukey 1977; Holland 1986), social scientists (e.g., Granger 1969, 1980; Gibbs 1982; Davis 1985), and epidemiologists (Evans 1976, 1978). Although covariation is not sufficient for causation, where covariation exists and it is possible to imagine a mechanism whereby X could cause Y, a causal hypothesis is likely to be entertained.

Covariation may be exhibited either cross-sectionally or longitudinally. For example, it may be observed that if any two members of a population differ in their values of Y, they also tend to differ in their values of X. Or it may be observed that if the value of Y has changed for any member of the population, the value of X for that same individual will tend to have changed. These two associations are not the same, and one can be true when the other is false. However, since not all X's can change, it may be impossible to observe the association of a change in X with a change in Y. For example, except in rare and special cases, attributes such as sex and race cannot change. For attributes of this type, change can be observed only at the population level. A change in the proportion of women in management positions, for example, may be associated with a change in the nature of interaction in the workplace. In general, because the concept of causal-

ity reflects ideas of agency and productivity, covariation between *changes* in the values of X and Y provides the most convincing evidence of causality.

Coherence. One can also look for coherence in the relationships among variables with which X and Y are associated. If X and Y covary, the relationships of these variables to other variables should be consistent with that covariation. For example, if labor force participation encourages an instrumental rather than expressive orientation toward interpersonal relationships, and if males spend more time in the labor force than females, males should be more instrumentally oriented in their relationships than females. Evidence of associations of this type may be particularly valuable if X and Y are difficult to measure directly or if direct measures are not available in the population under study.

Strength. Not only the existence of covariation but its *strength* has been argued to be an important cue to causality. For example, in evaluating the relationship between smoking and lung cancer, the U.S. Surgeon General's Advisory Committee identified a *high degree* of association as one of five criteria to be used in epidemiologic decisions about causality, and these criteria have now been widely accepted in that field (Feinstein 1986). It is viewed as particularly instructive when the hypothesized cause bears a higher degree of association to the effect than do other factors that can be viewed as alternative causes (Hill 1965).

Since a high degree of association between X and Y is suggestive of a causal link, its observation usually leads to efforts to rule out alternatives to the causal hypothesis. A low degree of association, or even the absence of an association, however, may not be a valid indication of a lack of causal connection. As described above, X is likely to be an inus condition of Y, i.e., one of several factors that conjoin to constitute a minimally sufficient condition for Y. The degree of association between X and Y is therefore affected not only by the boundaries of the reference population, which reflect the selection of a field against which the degree of association is assessed, but by the distribution of other variables in the population, including especially the joint distribution of X and its conjuncts for the production of Y and the joint distribution of factors constituting other minimally sufficient conditions for Y (i.e., alternative causes). As described by Einhorn and Hogarth (1986), if we think of a 2×2 table in which the

occurrence or nonoccurrence of X (x and $\bar{x}$, respectively) is crossed with the occurrence or nonoccurrence of Y (y and $\bar{y}$, respectively), x and y will occur jointly only when x and all its conjuncts are present. If x is present but one or more of its conjuncts is absent, x will not produce y, thereby causing x and $\bar{y}$ to occur jointly. This could result from either the absence of an enabling condition or the presence of a counteracting condition. The more x must conjoin with other factors to produce y and the more likely counteracting factors are to offset x, the weaker will be the association between X and Y. Similarly, if there are alternative causes of y in which x is not a factor, the conjunction of conditions that constitute these alternative causes will produce y in the absence of x. Thus, if the minimally sufficient condition for y of which x is a part is only one of a larger number of minimally sufficient conditions for y (i.e., if there are multiple causes of y), $\bar{x}$ and y will occur jointly. The strength of association between X and Y is therefore affected by the distributions of other causally relevant variables in the population. These other variables may act not only as dichotomous threshold triggers that condition the action of other variables but as more finely graded effect modifiers producing complex synergistic effects. There are a large number of possible ways in which variables may interact to affect the association of X and Y (Koopman 1977).

Congruity. In addition to the presence of covariation, one may observe congruity in an association. Congruity refers to some form of similarity between cause and effect. The type of congruity used most often as a cue to causality is similarity of the strength or duration of cause and effect. If an effect is large, one is inclined to expect the cause(s) of the effect to be of comparable size; strong causes produce strong effects, and weak causes produce weak effects. A good example of this type of congruity, or concomitant variation, is the "dose-response curve," in which a disease rate increases with the *amount* of exposure to the purported cause. In considering whether smoking is a cause of lung cancer, the fact that the death rate from lung cancer increases linearly with the number of cigarettes smoked per day is considered important evidence beyond that contained in the observation that cigarette smokers have a higher death rate than nonsmokers (Hill 1965). Of course, congruity between the strength or duration of cause and effect may not occur. When it does not, some sort of amplifying or dampening process must be postulated to connect cause and effect (Einhorn and Hogarth 1986). When small causes are hy-

pothesized to have big effects, a linkage process involving amplification must be postulated. When big causes are hypothesized to have small effects, a linkage process involving dampening must be postulated. A recent example of this type of incongruity, where small causes were hypothesized to have big effects via an amplifying process, is available in the work of Cole and Singer (1987) on gender differences in the productivity of scientists. An example outside the social sciences is Darwin's ([1859] 1964) formulation of evolution.

Another type of congruity that is sometimes used to make causal inferences is qualitative or structural resemblance between cause and effect. For example, physical similarity between offspring and adults helps to identify particular adults as causal agents of those offspring. A cause and effect may also be similar in space-time structure but different in intrinsic quality. For example, in broadcasting, where electromagnetic waves cause the sensations of hearers, there is a resemblence in structure between cause and effect. All visual and auditory perceptions involve the transmission of structure but not intrinsic quality. Similarly, the power and prestige hierarchy of a group produces a patterning in the interactions of group members, but the resemblance between cause and effect is in structure only. Complex structures can therefore be transmitted causally throughout changes of intrinsic quality.

Responsiveness. In situations in which it is possible to intervene and manipulate X, it is possible to obtain direct evidence of the responsiveness of Y to changes in X. Such responsiveness may be observed in either a natural or a controlled experimental setting, since one can introduce a treatment or withdraw a treatment and observe the response in either setting. Even taking preventative action with respect to X may provide an opportunity to observe the responsiveness of Y in a natural setting (Evans 1978; Cook and Campbell 1979). For example, one can observe whether those who stop smoking have lower rates of lung cancer than those who do not. Although randomization of treatment assignment or withdrawal is not a feature of all controlled experiments, in a controlled randomized experiment it is possible, at least in principle, to eliminate all alternative sources of influence and thereby increase the strength of the evidence that X causes Y. However, for studying many aspects of human behavior, controlled experimentation is either impossible or unwise. It is often impossible to undertake experimentation for practical and ethical reasons, and because the naturally occurring relationships of X to other variables are

frequently ignored in an experiment, experimentation introduces the real possibility of obtaining results that have no applicability in a natural setting (Fienberg, Singer, and Tanur 1985; Heckman and Hotz 1987).

Conjunctions

As noted above, what we typically refer to as a cause is what Mackie (1974) called an inus condition. X is usually one of a number of factors that, when conjoined, constitute one of a number of minimally sufficient conditions for Y. Because the conjuncts of X for the production of Y are often unknown, the strength and consistency of the relationship between X and Y may appear considerably weaker than would be the case if the relationship of X to its conjuncts were known and taken into consideration.

Although it is recognized that causation usually involves disjunctive plurality of causes, consideration of conjunctive plurality is rare in the social sciences. It seems likely that such consideration has been impaired by the popularity of statistical techniques for the estimation of linear, additive models and the relative inaccessibility of effective tools for identifying interactions. During the past 15 years, several new exploratory data-analytic methods have been put forth to deal directly with the problem of detecting conjunctions without imposing strong a priori linearity and normality assumptions. As might be anticipated, these methods are very computer intensive and still in need of further theoretical and practical development before they can be regarded as broadly applicable, "well-understood" techniques. We briefly discuss two such methods because of the insight they have already provided in a diversity of scientific problems in which the detection of relationships—conjunctions—among variables in a high-dimensional state space was of central importance. Much further methodological research and testing of these and analogous strategies in a broad array of social science problems will be necessary if a good technology for efficient detection of conjunctions as candidate causes is to be widely available.

Projection pursuit. The fundamental problem that any method for detecting conjunctions with many variables must confront is the fact that high-dimensional space is mostly empty. Projection pursuit techniques deal with this problem by selecting low-dimensional projections of high-dimensional point clouds, which an investigator then examines visually and attempts to interpret substantively.

The selection of projections has been carried out via the PRIM-9 program at Stanford Linear Accelerator Center, where the user may select any three variables at a time and have them displayed as a two-dimensional image of the projection of the points along any direction. By continuously moving the direction, the three-dimensional configuration of the points is revealed. To the best of our knowledge, the PRIM-9 system has not been used to detect relationships among variables in a social science setting; however, it has been used very effectively in medical contexts. (See, especially, Reaven and Miller [1979] for an application of PRIM-9 to a study of the etiology of diabetes.)

Automated machine selection of potentially interesting projections of a high-dimensional point cloud is carried out by numerical maximization of a "projection index," which, in its original version (Friedman and Tukey 1974), was the product of a robust measure of scale (trimmed standard deviation) and a measure of clumpiness (weighted count of the number of "close pairs" of data points). This kind of criterion is motivated by some limited experience gained watching scientists examine projections and attempt to characterize, in a general way, those projections that they regard as "interesting." Although *interesting* cannot be given a completely context-free definition, the common denominator of the judgments is that *non*normal projections tended to be interesting. This immediately suggests that maximizing some measure of clumpiness is important for any algorithm that is looking for interesting projections by searching a high-dimensional space. Further support for basing projection indices on clumpiness arises from a general mathematical result of Diaconis and Freedman (1984), who show that for *most* high-dimensional point clouds, *most* low-dimensional projections are approximately normal. Thus, the interesting and unusual projections should tend to be *non*normal.

Projection pursuit may, of course, be carried out on aggregations of variables (e.g., on linear or selected nonlinear combinations of them), thereby increasing the possible complexity of conjunctions that might be exposed when an investigator visually explores many projections. The reader interested in learning more about this technology and its scope and limitations as seen to date should consult Friedman and Tukey (1974), Friedman and Stuetzle (1982), and Huber (1985).

Grade-of-membership representations. The activity of exploring high-dimensional data for interpretable conjunctions may often be

viewed as an attempt to classify heterogeneous populations on the basis of a large number of characteristics, each of which is important for distinguishing some subgroup, however small, of the full population. Attempts at developing a small number of categories (conjunctions), defined in terms of the original large list of characteristics, such that each member of the population can defensibly be classified as a member of just one category frequently fail for a very fundamental reason. Many individuals are represented by multiple interrelated characteristics, no combination of which occurs with high frequency in the full population. Thus, we have a high-dimensional state space, most portions of which are either occupied by a few individuals or simply empty. Such data do not readily lend themselves to aggregation into a small number of *interpretable* categories, into only one of which any given individual should be classified.

Rather than strive for crisp classification with high-dimensional sparse data structures, we seek a *representation* of individual response vectors in terms of similarity, or grade-of-membership (GOM), scores relative to interpretable profiles (conjunctions) of conditions. If, for example, we construct—compatible with the data, in a sense to be made precise below—four "meaningful" sets of levels or categories extracted from an original long list of potentially important variables and for each individual assign GOM scores g_1, g_2, g_3, and g_4 such that $g_k \geq 0$ and $\sum_{k=1}^{4} g_k = 1$, then we identify an individual as having degree of similarity score g_1 relative to profile 1, degree of similarity score g_2 relative to profile 2, etc. If $g_k = 1$ for some profile k, then the individual is viewed as expressing only the characteristics of the kth profile. If $g_{k^*} = 0$ for some profile k^*, then the individual is interpreted as having no resemblance to profile k^*. The basic point is that for very heterogeneous populations in which individuals are initially described by high-dimensional vectors and in which no combination of responses occurs with very high frequency, it is often useful to *represent* individuals by GOM scores relative to a set of ideal (or pure-type) profiles.

To clarify the structure of GOM representations, let $\mathbf{X} = (X_1, \ldots, X_J)$ be a vector whose components are discrete variables, each of which can only assume a finite number of possible values. Thus, any continuous variables will be assumed to have been approximated by an ordinal categorical variable having a similar distribution. The distribution of $\mathbf{X}$ will be denoted by $\text{Prob}(\mathbf{X} = \mathbf{l})$, where $\mathbf{l} = (l_1, \ldots, l_J)$ is a vector whose coordinates are possible levels of the variables $X_1, \ldots, X_J$. The basic idea of a GOM representation for $\text{Prob}(\mathbf{X} = \mathbf{l})$ is the

transferring of the stochastic dependence among the original variables into GOM scores for individuals and then, conditional on these scores, assuming that the original variables are independent.

More formally, we associate with each individual a set of scores $\mathbf{g} = (g_1, \ldots, g_K)$ such that $g_k \geq 0$ and $\sum_{k=1}^{K} g_k = 1$. Then we label an individual's response vector as $\mathbf{X}^{(\mathbf{g})} = (X_1^{(\mathbf{g})}, \ldots, X_J^{(\mathbf{g})})$, and the distribution of $\mathbf{X}$ will, henceforth, be denoted by $\text{Prob}(\mathbf{X}^{(\mathbf{g})} = \mathbf{l})$. Now we *assume* that conditional on the values of $\mathbf{g}$, the original variables are independent. This leads immediately to the representation

$$\begin{aligned}\text{Prob}(\mathbf{X}^{(\mathbf{g})} = \mathbf{l}) &= \int_{S_K} \text{Prob}(\mathbf{X}^{(\mathbf{g})} = \mathbf{l} | \mathbf{g} = \boldsymbol{\gamma})\, d\mu(\boldsymbol{\gamma}) \\ &= \int_{S_K} \prod_{j=1}^{J} \text{Prob}\left(X_j^{(\mathbf{g})} = l_j | \mathbf{g} = \boldsymbol{\gamma} \right) d\mu(\boldsymbol{\gamma}), \qquad (1)\end{aligned}$$

where $\mu(\boldsymbol{\gamma})$ is the distribution of GOM scores and $S_K = \{\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)\colon\ \gamma_k \geq 0,\ \sum_{k=1}^{K} \gamma_k = 1\}$ is the unit simplex with K vertices. We further *assume* that the conditional probabilities in (1) are given by

$$\text{Prob}\left(X_j^{(\mathbf{g})} = l_j | \mathbf{g} = \boldsymbol{\gamma} \right) = \sum_{k=1}^{K} \gamma_k \lambda_{k,j,l_j}, \qquad (2)$$

where

$$\sum_{l_j \in L_j} \lambda_{k,j,l_j} = 1 \quad \text{for } 1 \leq k \leq K, \qquad 1 \leq j \leq J$$

and L_j = the set of possible levels of variable X_j.

The probability distributions $\{\lambda_{k,j,l_j}\}_{l_j \in L_j}$, $1 \leq j \leq J$, $1 \leq k \leq K$, have the interpretation λ_{k,j,l_j} = (probability in profile [pure-type] k of observing level l_j on variable X_j). They are the basis for defining the ideal, or pure-type, profiles. To this end, we introduce the following definition.

Definition 1. A family of probability distributions $\{\lambda_{k,j,l_j}\}_{l_j \in L_j}$, $1 \leq j \leq J$, $1 \leq k \leq K$, will be said to define a set of K *extreme admissible profiles* if the following are true:

a. There is at most one profile, call it k_0, such that $\lambda_{k_0,j,l_j} = 0$ for $1 \leq j \leq J$ and all "distinguished" levels $l_j \in L_j$.

Remark. One or more levels of each variable will be called *distinguished* if they represent a priori designated characteristics that

should be the basis for designating membership in a pure-type profile.

Thus, k_0 is identified as a profile in which no distinguished characteristic occurs.

b. For all profiles other than k_0, there is at least one variable for which $\lambda_{k,j,d_j} = 1$, where d_j = the set of distinguished levels on variable j. For all variables not satisfying this condition, $\lambda_{k,j,d_j} = 0$.
c. For each pair of profiles (k, k') not including k_0, there is at least one variable for which $\lambda_{k,j,d_j} = 1$ and $\lambda_{k',j,d_j} = 0$. This means that there is at least one condition that distinguishes each profile from all of the others.

Conditions (a)–(c) imply that the *joint* probability of occurrence of "distinguished" conditions associated with each profile, except k_0, is one. Thus, an individual for whom $g_k = 1$ is characterized as someone who must have *all* the distinguishing characteristics associated with profile k. The important interpretive feature of extreme admissible profiles is that they are described by logical AND statements (i.e., a conjunction) involving distinguished levels of subsets of the original variables $X_1, \ldots, X_J$. The sense in which (1) and (2) with $\{\lambda_{k,j,l_j}\}_{l_j \in L_j}$, $1 \le j \le J$, $1 \le k \le K$, satisfying (a)–(c) define a new *representation* of a data set $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(N)}$, where $N =$ number of individuals, is the fact that we define a point mapping from J-coordinate space $X_1 \otimes \cdots \otimes X_J$ (defined by the variables $X_1, \ldots, X_J$) to the simplex S_K whose vertices are identified with extreme admissible profiles. Each individual's GOM score $\mathbf{g}^{(m)} = (g_1^{(m)}, \ldots, g_K^{(m)})$, $1 \le m \le N$, defines a location for that individual in S_K.

For some applications, criteria (a)–(c) prove to be excessively stringent. Thus, it is important to have a somewhat more general definition of *admissible profile* whose logical interpretation is slightly more intricate than that of *extreme admissible profile*. To this end, we introduce the following definition.

Definition 2. A family of probability distributions $\{\lambda_{k,j,l_j}\}_{l_j \in L_j}$, $1 \le j \le J$, $1 \le k \le K$, will be said to define a set of K *admissible profiles* if the following are true:

a′. There is at most one profile, call it k_0, such that $\lambda_{k_0,j,d_j} < 2\ \text{Prob}(X_j \in d_j)$ for $1 \le j \le J$ and d_j are the distinguished levels for variable X_j.

b′. For all profiles other than k_0 there is at least one variable for which $\lambda_{k,j,d_j} > c\,\text{Prob}(X_j \in d_j)$, where c is chosen by the investigator subject to the constraint, $c \geq 2$.

c′. For each pair of profiles (k, k') not including k_0, there is at least one variable for which

$$\lambda_{k,j,d_j} > c\,\text{Prob}(X_j \in d_j)$$

and

$$\lambda_{k',j,d_j} < c\,\text{Prob}(X_j \in d_j).$$

A verbal interpretation of (a′)–(c′) is given by the following:

1. The *joint* probability—within profile k—of occurrence of distinguished levels is much greater than the joint frequency of these levels in the population as a whole.
2. The probability within each profile other than k_0 of the occurrence of at least one distinguished level is much greater than the frequency of occurrence of the same level(s) in the population as a whole.

The above specifications are restricted to static representations of heterogeneous populations. An extension of these ideas to vector stochastic processes is of great importance and is presented in Manton et al. (1987), where estimation strategies and computational issues are also discussed. For an insightful application of grade-of-membership representations in detecting conjunctions in psychiatry, see Swartz et al. (1986).

Our purpose in presenting these brief summaries of projection pursuit and grade-of-membership representations is to indicate by explicit example that there are flexible strategies currently available and under development which should facilitate the search for conjunctions.

Temporal Plausibility

Temporal succession is widely regarded as an important cue to causality because it is strongly tied to the criterion of asymmetry, or causal priority, reflected in the requirement that a cause must be there if its effect is to occur. Causes usually occur prior to their effects. Contemporaneous and backward causation are possible, as noted above, but genuine instances of them are rare, and it is a matter of debate whether they even exist. This is not to say that measurements are

always, or even often, refined enough to identify temporal succession but that temporal succession characterizes almost all causal relations in the empirical world. Thus, the requirement that a cause precede or be contemporaneous with its effect (i.e., that Y not occur before X) plays an important role in operational attempts to identify causal relations in all branches of social science (Granger 1969; Davis 1985; Einhorn and Hogarth 1986) and epidemiology (Hill 1965; Evans 1978; Feinstein 1986). In general, longitudinal data are required to assess temporal plausibility. Even when a cause is believed to be contemporaneous with its effect, tests of causal priority usually require that variations in the postulated cause be related to later variations in the postulated effect. Thus, one seeks to observe that changes in the cause have temporal precedence over changes in the effect prior to reaching a state of contemporaneous causation (Cook and Campbell 1979).

A major problem with use of the criterion of temporal plausibility in social science is that the temporal order in which behavior occurs, or in which events resulting from behavior occur, is often not a good indication of causal priority. Because human beings can anticipate and plan for the future, much human behavior follows from goals, intentions, and motives; i.e., it is teleologically determined. As a result, causal priority is established in the mind in a way that is not reflected in the temporal sequence of behavior or even in the temporal sequence of the formation of behavioral intentions.

For example, consider the relationship between women's educational attainment and the timing of entry into marriage, where the temporal sequencing of exit from school and entry into marriage does not provide an appropriate basis for distinguishing a hypothesized causal effect of educational attainment on the timing of marriage from a hypothesized causal effect of the timing of marriage on educational attainment. Among women who leave full-time schooling *prior to* entry into marriage, there are some who will leave school and then decide to get married and others who will decide to get married and then leave school in anticipation of the impending marriage. For example, among women who terminate their schooling with the completion of college and then marry are women who will finish college, find a desirable mate, and then marry and women who will find a desirable mate while in college, abandon plans for graduate school in anticipation of an impending marriage, and then marry. Both groups of women experience exit from school and entry into marriage in the same temporal

sequence, but the causal mechanisms at work are different. For the former group, educational attainment is causally prior to the timing of entry into marriage; for the latter group, the timing of entry into marriage is causally prior to educational attainment. That is, for the latter group the intention to marry causes women to truncate their schooling earlier than they would otherwise, lowering their educational attainment. Since the anticipation of entry into marriage may affect educational attainment prior to entry into marriage, whether education occurs before or after entry into marriage does not distinguish the effect of educational attainment on the timing of marriage from the effect of the timing of marriage on educational attainment. Moreover, this problem is not necessarily solved by considering the temporal sequence of the formation of behavioral intentions. In a situation such as this, in which there is perceived incompatibility between future activities, a single decision may jointly (i.e., simultaneously) produce a related set of behavioral intentions. To identify causal priority, therefore, one must search for evidence that is not reflected in the *temporal* sequence of behavior or even of thought. This evidence may be obtainable only by asking people about the causal processes at work. Although such information can be affected by distortions in perception, memory, and reporting, it is sometimes the only evidence with which to establish causal priority in the social sciences.

Contiguity

Evidence that events are contiguous in time and space is another empirical indication of causality. As noted by Hume ([1739] 1896, [1740] 1938, [1748] 1900), when contiguity is high—i.e., when there is little time or distance between the occurrence of X and Y—one is more likely to suspect a connection between X and Y and to postulate a causal mechanism by which X and Y are related. By comparison, when contiguity is low, one is less likely to suspect a causal connection, and it is only the identification of a causal mechanism linking the events that justifies the inference that causality is involved. For example, some knowledge of human biology and chemistry is needed to bridge the temporal gap between intercourse and birth in maintaining that there is a causal relation between these events (Einhorn and Hogarth 1986). It is impossible to establish specific, universally applicable time and space boundaries within which both X and Y must fall to satisfy a contiguity criterion of causation. High contiguity is an aid

to causal inference, but it need not be present if a mechanism linking X and Y can be identified. Contiguity facilitates the identification of relationships and thereby leads to speculation about causal mechanisms that may produce them.

3.2. *Theoretical Development*

Causal inference occurs not only through the "bottom-up" process of forming hypotheses on the basis of empirical observation but also through the "top-down" process of relating what is observed empirically to an existing body of relevant knowledge, including knowledge of the world gained through previous experience with similar empirical relations. In drawing on knowledge of similar empirical relations, use of analogies plays an important role. If X and Y are similar to other variables for which there is already a body of knowledge to suggest the existence of a cause-effect relationship, the similarity among the separate causes and among their separate effects will be suggestive of a cause-effect relationship between X and Y. For example, if there is reason to believe that gender stereotypes unconsciously affect assessments of the performance of women and men, it is reasonable to expect, by analogy, that racial stereotypes unconsciously affect assessments of the performance of blacks and whites.

A causal inference is strengthened if there is a carefully reasoned explanation (theory) that provides details of a mechanism by which the cause is related, often step by step, to the effect. As noted by Simon, the postulating of a causal mechanism occurs by a process of induction:

> Mechanisms and laws are theoretical constructs that are derived *inductively* from empirical evidence but are not derivable *deductively* from that evidence. We can never show that a particular mechanism did, in fact, cause certain phenomena; we can only show that a particular mechanism *could* have produced the phenomena—that, if the mechanism had been at work, the phenomena would have appeared. (Simon 1979, p. 71)

Because the possible causal mechanisms that are likely to be imagined are derived from the empirical indications of causality discussed above

and knowledge of the world gained through experience with similar empirical relations, they should be consistent with empirical evidence and the existing body of relevant knowledge. (See Simon [1986] for a discussion of the problems arising in economics from insufficient attention to empirical evidence in the development of theory.) In linking statements that X causes Y to a postulated causal mechanism reflecting a subject-matter theory, subjectivity and beliefs in the plausibility (or lack thereof) of particular theories will enter into causality assessments. The controversial nature of claims of substantive coherence (i.e., strong arguments on behalf of a particular theory about how X drives Y) is an integral part of the process of assembling evidence to support the claim that X causes Y.

As noted by Einhorn and Hogarth (1986), it has been suggested that attributions of physical causation involve the perception that causes and effects are linked by a "generative" force. This implies a mechanistic view in which there is a physical transfer of causal "energy" from X to Y. Events are seen as linked via a causal chain so that the force can be transmitted from one link to the next. This view makes it clear that if one cannot construct a causal chain to link X and Y, there is no basis on which to claim that X causes Y. Thus, the causal chain connecting X and Y is only as strong as its weakest link; if one link in the chain cannot be made, the explanation is incomplete.

In scientific contexts, we seek to make generalizations that denote lawful regularities. These are distinguished from "accidental" regularities by direct inductive confirmation and by their relationship to other laws. Although a strict separation between what is observable and what is not may be impossible, observational laws, which deal with observable things and processes, can be contrasted with theoretical laws, which postulate unobservables. The development and use of theories take on particular significance because theories explain and unify broad classes of facts, demonstrating that phenomena thought to be disparate are similar.[7] Theories differ from sets of laws in several respects (Holland et al. 1986). First, an observational law gives an account of a set of observations, whereas a theory often explains sets of laws. For example, a theory of power-dependence relations may explain why both increases in female earnings and improvement in household technology increase the probability of divorce. Second, theories are intended to unify phenomena in different domains. Thus, a

[7] This characteristic of a theory was termed consilience by Whewell (1967).

theory of power-dependence relations may explain changes in the relative position of a disadvantaged group as well as changes in the probability of divorce. Third, it is by postulating unobservable entities, such as power and dependence, that theories usually achieve their unifications.

3.3. *Confirmation*

After empirical observation and inductive reasoning have led to the formation of a causal hypothesis, which is often linked to a postulated causal mechanism, a body of evidence must be assembled to support the claim that X causes Y. There are no hard-and-fast rules of evidence that *must* be obeyed before a causal inference is justified. Causal inferences can never be made with absolute certainty but are made with varying degrees of confidence depending upon the evidence available. As noted by Russell (1959, p. 102), "'Knowledge' is not a precise conception, but merges into 'probable opinion'." Confirming the hypothesis that X causes Y involves demonstrating consistency in the association of X and Y, demonstrating that the association between X and Y is not attributable to alternative explanations, and identifying the mechanism by which X causes Y. This process of confirmation is usually accomplished by accumulating a body of evidence from multiple studies.

Consistency

Consistency refers to the need to replicate an association, i.e., to demonstrate that it is consistently observed. This is, of course, a well-recognized general criterion for increasing the evidence relevant to any hypothesis in all branches of science. In accumulating evidence to support a generalization in the social sciences, it is usually desirable to replicate the association in different localities and by different methods to establish that it exists under varying conditions. As noted by Keynes (1921),

> No one supposes that a good induction can be arrived at merely by counting cases. The business of strengthening the argument chiefly consists in determining whether the alleged association is stable, when the accompanying conditions are varied.

For hundreds of years, philosophers have attempted to characterize the process of induction whereby inferences lead from knowledge of particular observations to acceptance of a generalization. In the last several decades, this activity has focused largely on the question of when a general hypothesis can be considered confirmed by its instances. Because of the advances made by the use of formal logic in the study of deduction, in which the truth of an inference is guaranteed by the truth of the premises on which it is based, most recent work on induction has been governed by syntactic approaches. Beginning with Keynes (1921) and followed up by Nicod (1930), Hempel (1945), Carnap (1950), Carnap and Jeffrey (1971), and Salmon (1967), to name only some of the principals, we have a series of increasingly refined attempts to build rigorous, purely logical theories of induction and confirmation of evidence that would, in effect, do for induction what Frege ([1884] 1961) and Russell (1908) did for deduction. These theories use the formalism of logic and probability theory, most recently employing Bayesian probability theory and decision theory. However, this primarily formal and syntactic approach to inductive reasoning has been shown to give rise to numerous paradoxes; see, especially, Hempel (1945) and Goodman (1965).

Holland et al. (1986) have initiated a thorough and very different rethinking of induction at the interface of cognitive psychology, artificial intelligence, and philosophy. In their view, purely syntactic accounts are insufficient because they do not consider the kinds of things about which inferences are being made and the goals the inferences serve. Some time ago, Mill (1919, p. 206) noted that the number of instances required to warrant acceptance of a generalization is not constant. He asked, "Why is a single instance, in some cases, sufficient for a complete induction, while in others myriads of concurring instances, without a single exception known or presumed, go such a very little way towards establishing a universal proposition?" Holland et al. (1986) have argued that the number of instances is only one of two major components in evaluating the acceptability of a generalization. The other consists of knowledge of the statistical properties of the populations about which one wishes to generalize, including the distributions of events and the role of chance in producing them. If the events of interest are known to be the kinds of things that are highly invariant and not much subject to random fluctuation, generalization from a few instances will be considered legitimate.

However, if such things are highly variant or highly subject to randomness, generalization will require many confirming instances.

When people are able to code the uncertainty of events, their reasoning is affected because they invoke abstract inferential rules, or "pragmatic reasoning schemas," that embody statistical reasoning. As a result of differences across different content domains in the variability of events and the role of chance, the inferential rules used in different domains differ. In the social sciences, where variability is high and chance plays an important role, inductive reasoning is guided by inferential rules that embody statistical principles. The inferential rules employed are not domain-specific empirical rules tied to particular types of events but abstract inferential rules describing relations over general classes of objects, relationships between events, and problem goals. Thus, inductive reasoning involves knowledge structures at an intermediate level of abstraction. Pragmatic reasoning schemas potentially apply across a wide range of content domains but are constrained by certain broad types of goals and event relationships in a way that content-free logical rules are not. This revised perspective on induction, which is rooted in an understanding of the cognitive processes by which humans acquire, develop, and accept (as plausible) new knowledge, suggests that the body of evidence required to support the claim that X causes Y will be affected by the distributions of X and Y and the degree to which X and Y are subject to both random fluctuation and measurement error. Knowledge of this type helps to establish the "finite antecedent probability" described by Keynes (1921) as necessary to validate inductions based on a number of observed instances. To a degree, awareness of such probabilities is embedded in the norms that emerge within scientific specialties regarding the strength of evidence required to support the claim that X causes Y. Such norms arise from the nature of the phenomena under study, the goals (end uses) of causality assessments, the measurement technology in the substantive area of inquiry, and the richness of the proposed explanatory theories being investigated.

Because the association between X and Y is affected by the joint distribution of factors constituting various minimally sufficient conditions for Y, the association between X and Y is likely to vary across populations, even being absent in some. As noted by Gibbs (1982, p. 97), "It would be unrealistic for sociology and perhaps all observational sciences to require that each finding support a causal assertion."

Similarly, when Sober (1984) discusses population-level causation in the context of evolution, he suggests that a positive causal factor must raise the probability of the effect in *at least one* background context without lowering it in any. Although we believe it is impossible to establish a specific, universally applicable level of consistency that must be met to satisfy an associational criterion of causation, the greater the degree of consistency with which a high level of association is observed and the greater the ability to explain situations in which an association is not observed, the stronger the basis for causal inference.

In the social sciences two types of consistency are particularly important for establishing the validity of a causal assertion. Consistency across methods establishes "construct validity," and consistency across localities establishes what Cook and Campbell (1979) have called "external validity." Construct validity is established by demonstrating *convergence* across different measures of the same thing and *divergence* between measures of related but conceptually distinct things. External validity is established by demonstrating that a causal relationship is generalizable to and across populations of persons, settings, and times. Thus, one is interested in demonstrating that a relationship is limited to neither a particular idiosyncratic sample nor to a particular population but holds in different populations, across settings and times. It is often through attempts to assess the consistency of an association across settings and times that knowledge of the factors that constitute various minimally sufficient conditions for the production of an effect is acquired. Ultimately, variation in the association between X and Y across populations should be attributable to variation in these contingent conditions.

In attempting to assess the consistency of an association, one faces the problem of knowledge synthesis, namely, the need for defensible strategies for combining the evidence from multiple studies to support or refute claims that there is an association (response) relating X and Y. This problem has a long history, the rigorous development of which is traceable to the work of Fisher (1946), Pearson (1933), and Tippett (1931). Fisher's work, in particular, has had a major influence on subsequent developments. It focuses primarily on combining evidence from independent studies that purport to measure the same quantity and where measured effects in any single study are just at the borderline of being declared significant (i.e., weak association in a single study), while the "combined" evidence from all studies indicates

a very significant (in the sense of statistical significance) association. The difficulties involved in combining estimates of effects from multiple studies in which a similar, but *not the same*, quantity is being measured were treated in early papers of Cochran (1937) and Yates and Cochran (1938) and have received intermittent attention since then, as reviewed, for example, in Hedges and Olkin (1985). Two recent papers on this topic which deserve close reading are Mosteller and Tukey (1982, 1983). A fundamental issue in all of this literature, which requires further development, is the provision of a good rationale for the inevitably subjective weights that are used to combine (pool) the comparisons from studies of varying size and quality and involving the measurement of similar, but *not the same*, quantities.

The early literature we refer to on combining evidence usually involves consideration of evidence from controlled experiments or at least from studies of the same type; i.e., they are all controlled randomized experiments or all observational studies involving the same or very similar data-collection plans. In many settings, such as the evaluation of educational interventions (teaching materials, kinds of homework, classroom organization, discipline, etc.), the evidence about potential causal relationships comes from both experimental and observational studies. The assessment of effect sizes in the contemporary meta-analysis literature (Hedges and Olkin 1985; Wolf 1986) represents one approach for dealing with the knowledge-synthesis problem in this kind of heterogeneous setting. However, more in-depth, quantitative syntheses in a variety of substantive contexts are needed to develop broadly applicable strategies for supporting claims about the consistency of an association.

Ruling Out Alternatives to the Causal Hypothesis

The existence of empirical cues such as covariation, temporal plausibility, and contiguity raises the question of whether there is a causal relation. However, to justify acceptance of the causal hypothesis, it is necessary to establish that the relationship between X and Y is one of agency and productivity, i.e., that it is not attributable to common causes of X and Y or to a causal effect of Y on X. In ruling out the possibility that the X, Y relation is spurious, the task is not to rule out the existence of other causes of Y, since the possibility of multiple causation is fully acknowledged, but to rule out the possibility that one or more of these causes accounts for the relationship between X and Y.

Einhorn and Hogarth (1986) draw a distinction between the "gross strength" of the evidence that X causes Y, which is determined by the causal field and empirical indications of a causal relation between X and Y, and the "net strength" of the evidence that X causes Y, which is the gross strength of the X, Y relation discounted by the gross strengths of alternative explanations or causes. The net strength of a causal explanation increases as its gross strength increases but decreases as the gross strengths of its alternatives increase. Thus, the strength of the evidence in support of a causal claim is greatest when there are no plausible alternative explanations and is lower when there are such alternatives.

Consideration of alternative explanations has been an important emphasis in the work of Campbell and his colleagues on threats to "internal validity" (Campbell and Stanley 1963; Cook and Campbell 1979). In assessing internal validity, one asks what factors other than X could have produced Y. Consideration of *specific* alternatives is important because one can consider empirical indications of their plausibility and replace the current explanation with an alternative when the plausibility of the current explanation is reduced. To rule out alternative explanations, it is necessary to show

> that either there are no plausible common causes of X and Y or that the quantitative relationships between the plausible common causes and X and Y are inadequate to explain an observed clear and consistent association. (Mosteller and Tukey 1977, p. 261)

It is also desirable to show that Y cannot cause X.

Ruling out alternative explanations and thereby demonstrating that the X, Y relation is causal requires, in effect, a demonstration that the outcome with respect to Y would have been different in the absence of X. The usual approach to accomplishing this is to examine the relationship between X and Y under conditions in which alternative explanations are rendered inoperative. If X and Y covary under these conditions, there is greater reason to believe that X causes Y.

An experiment is sometimes the preferred method for gathering evidence to support the claim that X causes Y. Experimental control permits manipulation of the values of X, making it possible to observe the responsiveness of Y to changes in X rather than merely the

association of differing values of X and Y across population units or even the association of changes in the values of X and Y for individual population units. When experimental control involves randomization to treatments, a further attempt is made to make the distribution of X independent of the distributions of other variables that may affect Y, rendering alternative explanations of the association between X and Y inoperative. As noted above, experiments—including both controlled laboratory experiments and randomized experiments in field settings—are often not possible in social science and, even when possible, are not always the preferred method. Thus, although experiments can play an important role in some areas of research, most causal inferences will, of necessity, be based on other kinds of evidence. For more detailed discussion of the advantages and limitations of experiments, see Campbell and Stanley (1963), Cochran (1965), Cook and Campbell (1979), Rubin (1974), Fienberg et al. (1985), Singer (1986), Holland (1986), Heckman (1987), Heckman and Hotz (1987), and Berk (1988).

Quasi experiments, which have treatments, outcome measures, and experimental units but do not involve randomization to treatment, are another means of gathering evidence to support causal inferences in field settings. In quasi experiments comparisons to determine the effect of a treatment are based on nonequivalent groups, and the task in interpreting the results is to separate the effects of the treatment from those due to the initial noncomparability of treatment groups. Quasi-experimental designs are of two principle types: (a) nonequivalent group designs, in which responses of a treatment group and a comparison group are measured before and after a treatment, and (b) interrupted time-series designs, in which the effects of a treatment are examined by comparing measures taken at many time intervals before the treatment with measures taken at many time intervals after the treatment (Cook and Campbell 1979). These designs combine experimental manipulation with different data collection plans to provide a stronger basis for causal inference than would be possible without manipulation.

In nonexperimental, or "observational," research the role played by alternative causes must be examined without experimental manipulation. In some cases it is possible to collect data on what are, in effect, "natural experiments," in which X is distributed more independently of the hypothesized alternative causes of Y than is usually the case,

and the X, Y relation can be observed more apart from these influences. For example, in seeking to understand the effects of heredity and environment on behavior, scientists have sought to identify naturally occurring situations in which one of these influences is relatively constant while the other varies, such as when identical twins are separated at birth and reared in different environments. The modes of analysis used in quasi experiments can sometimes be employed when the "treatment" is a natural occurrence, such as a natural disaster, rather than a planned intervention. Analytic approaches appropriate for interrupted time-series designs can be used when the event being examined is abrupt and precisely dated and when it does not result from prior change in the level of the indicator. Analytic approaches appropriate for nonequivalent group designs can be used even when the "treatment" is a permanent institution. As in quasi experiments, the availability of measures of Y taken before and after a "treatment" can provide a sounder basis for causal inference.

When it is impossible to obtain observational data that approximate a natural experiment, statistical analysis alone is used to "partial out" alternative causes that covary with X under the assumption that the X, Y relation that remains after partialling is likely to be causal. In the past several decades it has become common practice to represent (model) hypothesized causal relations by systems of equations. Since the development of path analysis by Wright (1921, 1934, 1954), there has been extensive development and application of linear models, referred to variously as structural equation models, factor analytic models, path analysis models, or even regression models (see, e.g., Simon 1953, 1957, 1979; Blalock 1964, 1971; Fisher 1966; Goldberger and Duncan 1973; Aigner and Goldberger 1977; Bielby and Hauser 1977). In these models a set of causal relationships among variables is represented by a set of equations relating the variables of the model and a set of stochastic assumptions about the probability distributions of those variables, jointly and individually. The models usually assume linearity in both variables and parameters and an underlying multivariate normal distribution. (See Freedman's [1987] critique on the use of structural equation models and other related articles in the *Journal of Educational Statistics*, volume 12, number 2.) The use of stochastic process models to represent evolving processes has also occurred but has been less common (see, e.g., Bush and Mosteller 1955; Cootner 1965; White 1970; Bartholomew 1982; Malliaris and Brock 1982).

Although the technology for causal analysis has at times been subject to mindless application, it properly involves the formal representation of hypothesized causal relations. It is not a way of deducing causation but of quantifying already hypothesized relationships. As noted by Fisher (1946, p. 191), causality cannot be identified by statistical operations alone:

> If ... we choose a group of social phenomena with no antecedent knowledge of the causation or absence of causation among them, then the calculation of correlation coefficients, total or partial, will not advance us a step toward evaluating the importance of the causes at work.

All attempts to define causality within the context of a formal model require some imposed structure. Using available knowledge, including existing theory and the kinds of empirical indications of causality discussed above, a scheme of causal relations is postulated and then quantified. A priori assumptions reflecting postulated causal mechanisms make estimation of the coefficients of the model possible.

Because the coefficients estimated are conditioned on an assumed causal structure, knowing that they are statistically significant and can generate good predictions of the data does not prove the existence of causal relations. It indicates only that the data are consistent with the proposed causal hypothesis. Of course, if a causal hypothesis is first suggested by observed correlations or partial correlations in the data analyzed, estimation of a causal model using the same data cannot provide *independent* confirmation of the hypothesis.

To develop causal models that more accurately reflect empirical causal relations, the meaning of causality, as discussed above, needs to be kept in mind when models are formulated. One factor rarely considered now in the formulation of causal models is the likelihood that the "causes" identified are not additive. The "causes" studied are likely to be inus conditions, which conjoin with other "causes" to produce effects. Attention needs to be paid, therefore, to understanding the relationships among candidate causes. Absent from most "partialling" exercises is any awareness that a conjunction of factors may constitute a minimally sufficient cause of Y, that several such conjunctions may be multiple causes of Y, and that a single factor may operate in one or several of these conjunctions. Partialling out other

"causes" of an outcome in an effort to estimate the effect of a single cause, as is commonly done, may not be appropriate. With well-formulated models of the interrelationships among the independent variables in an equation, the basis for making causal inferences from observational data could be strengthened considerably beyond what it has been thus far.

A second factor requiring more attention when models are formulated is the use of temporal succession as an indication of causal priority. Temporal order has been used heavily as a guide in model formulation because causes do not follow their effects in time. However, as noted above, the order in which behaviorial events or behavioral intentions occur is not always indicative of causal priority when purposive human action is involved. Moreover, in the process of operationalizing cause-effect relationships, the reference point used in establishing temporal, and therefore causal, order is often the time at which a measurement is taken rather than the time at which an influence occurs, particularly if the measurement refers to an ongoing process rather than to a discrete event. Because the number of measurements taken is finite and because they may be spaced at wide intervals, ordering variables primarily on the basis of the time at which they are measured can grossly misrepresent the influence process. This problem is particularly serious when measures of mental states or attitudes are involved, since these reflect ongoing processes that are difficult to measure repeatedly.

A third factor warranting more attention in model formulation is the distinction among types of causes, particularly those involving extrinsic determination, self-determination, and teleological determination. In some branches of social science, it is quite common to treat the earlier state of an ongoing process as a "cause" of a later state in the same manner that causes extrinsic to the process are treated. As noted above, self-determination of this type does not have the same causal status as extrinsic determination. Similarly, measures of intentions are often treated in the same way as other causes. Although inclusion of these measures reflects an awareness of teleological determination, the distinct role played by teleological determination requires more attention than it has received thus far. (See Frydman and Phelps [1983] for an example of an attempt to deal with this issue in economics.)

Although causal models were developed to represent hypothesized causal relations, many models have not been well tuned to the

phenomena under investigation and the theories driving them. The absence of explicit subject-matter considerations, including theory, for example, has been a major criticism of the Wiener-Granger tests of causality applied in the analysis of economic time-series (Geweke 1984; Zellner 1979). Stochastic variables and temporal succession are central to the Wiener-Granger framework, and within that framework, a "cause" is identified on the basis of its ability to *predict* an effect. Thus, if time series $\{X_t, t \geq 0\}$ and time series $\{Y_t, t \geq 0\}$ are conditionally independent given time series $\{Z_t, t \geq 0\}, \{X_t, t \geq 0\}$ is not a cause of $\{Y_t, t \geq 0\}$. Rather than rely on purely statistical criteria of this type, we take it as essential that (a) a link be provided between the context-free class of models usually employed in Wiener-Granger causality analyses (e.g., ARIMA models) and a focused theory, (b) an assessment via multiple goodness-of-fit tests of empirical data to the autoregressive models be carried out, and (c) a clear statement be presented defending the time periods over which stationary processes are a reasonable approximation of the phenomena under observation.

The importance of explicit subject-matter considerations is emphasized in a recent paper by Basmann (1988), which considers the problem of observational equivalence in interdependent, or nontriangular, systems of equations. Previous discussion of the causal interpretation of parameters in simultaneous equations by econometricians, such as Strotz and Wold (1960), has tended to view the problem in terms of algebraic operations. Since a given probability distribution determines a unique reduced form, coefficients in triangular recursive systems have been considered to have a "causal interpretability." The role of nonstatistical subject-matter information has been minimized or eliminated in these earlier treatments, and causality has been viewed as testable by methods of statistical inference alone. However, in nontriangular systems a given probability distribution determines an infinite class of observationally equivalent simultaneous equations representations. These representations, each of which corresponds to a set of identifying restrictions, assert different and contradictory hypotheses about causal relations but have the same significance probability and power when tested on the data.

In such a situation, in which statistical tests cannot support one choice against the others, information external to the model is needed to warrant the use of one specific representation as truly "structural." The information must come from the existing body of knowledge

relevant to the domain under consideration. If this information is insufficient to rule out all but one of the observationally equivalent structural representations, a unique causal relation cannot be singled out. The problem of observational equivalence is not addressed by solving the problem of identifiability, since alternative observationally equivalent representations are identified under different identifying restrictions (Koopmans and Reiersol 1950).

In some instances models are not devoid of subject-matter considerations, but the importance of their being able to predict an empirical reality external to the model is not recognized. For example, Simon (1953, 1957) viewed causality as a deductive logical property of models and was not concerned with the ability of the model to predict empirically observable outcomes. As Simon stated in the first chapter of *Models of Man* (1957, p. 12),

> The aim of this chapter is to provide a clear and rigorous basis for determining when a causal ordering can be said to hold between two variables or groups of variables in a model.... The concepts to be defined all refer to a model—a system of equations—and not to the "real" world the model purports to describe.

Simon's notion of causal order can be viewed as part of a larger program of establishing causality provided (a) the models are a formalization of mechanistic theories of the manner in which X could have produced Y, and (b) the models have been tested against data, where context-specific multiple goodness-of-fit criteria are set up as a standard of performance.

The relevance of causal models to empirical phenomena is often open to question because assumptions made for the purpose of model identification are arbitrary or patently false. The models take on an importance of their own, and convenience or elegance in the model building overrides faithfulness to the phenomena. Whether assumptions made in order to estimate a model impair the usefulness of the model for confirming or disconfirming causal hypotheses is a question warranting more attention. Important aids in addressing this question can be additional research to test the plausibility of the assumptions and the estimation of alternative models to examine the sensitivity of estimates to differing assumptions (Leamer 1978).

When comparisons are made between what happens under the action of a candidate cause and what would have happened without it, statements of the magnitude of a comparison that will be regarded as "strong" (or "large") are both subjective and context-dependent. An important set of unresolved statistical research problems are associated with the comparison activity, particularly in the context of stochastic processes. The issue is one of delineating what empirical regularities, or what details of an evolving process, a model will be required to reproduce with some fidelity before it is regarded as an adequate descriptive model of the underlying phenomena. Examples of the kinds of features we have in mind are

1. differences in rates of change of an outcome variable in a population exposed to a candidate cause vs. the corresponding quantity in an unexposed population,
2. the ratio of the number of occurrences of a given event in an exposed (to the candidate cause) population to the same count for an unexposed population,
3. the difference between the "delay" in response to a secondary stimulus (not the candidate cause) in an exposed (to the candidate cause) population relative to the response delay in an unexposed population.

Multiple criteria such as (1)–(3) lead to vector optimality criteria (Tukey 1987) for fitting models to data and *multiple* goodness-of-fit tests for assessing the descriptive adequacy of models. An informal assessment (comparison) of multiple models subject to multiple criteria is a routine part of much empirical analysis (see, e.g., Bush and Mosteller [1959] for a comparative study of eight stochastic learning models subject to multiple goodness-of-fit criteria), but a deep understanding of the properties of such tests lies in the future. Comparative analysis and multiple tests to assess "large" differences ("strong" associations) are presented in a linear models context in Glymour et al. (1987), in a log-linear context in Goodman (1973), and in a counting process context in Heckman and Walker (1987). Much remains to be done in this direction to put the comparative analysis component of causality assessments on a firm foundation.

Identifying the Causal Mechanism

In addition to seeking evidence that X causes Y by attempting to rule out the possibilities that the X, Y association is spurious and that Y causes X, one is interested, at least ultimately, in understanding *why* X causes Y. In other words, one seeks to identify the mechanism by which X causes Y. To date, most empirical work in the social sciences has sought to identify causal relationships at a molar level, providing evidence of empirical regularities, or observational laws. Attempts to "explain" a causal relation usually involve the identification of other observed variables measured at the same level that are part of a causal chain linking X and Y. This approach to elaborating a causal chain has produced an increasingly detailed picture of empirical relationships, often varying across time and place. Although knowledge of such relationships is essential for the development of theories that are tuned to the phenomena, there has as yet been little attempt to use these observations as a basis for the development of theories aimed at unifying broad classes of facts by postulating unobservables. In economics considerable theoretical work has been carried out, but theory development is largely a deductive exercise, often done with little attention to its relationship to empirical phenomena. In other fields, such as sociology, the importance of theory in developing a cumulative body of knowledge has gone largely unrecognized, and the documentation of empirical regularities is an end goal in itself. Once theories have been developed to account for what appears, on the basis of prior observation and research, to be a causal relation between X and Y, research designed to test these theories will be needed. A confirmed theory identifying the mechanism by which X causes Y greatly strengthens the basis for causal inference.

Cumulating a Body of Evidence

Few would question that the use of "causal" models has improved our knowledge of causes and is likely to do so increasingly as the models are refined and become more attuned to the phenomena under investigation. However, causal models have limitations that call for the use of other operational approaches in building a body of evidence to support causal inferences. One limitation noted by Cordray (1986, p. 17) is that causal models are "mechanistic and inflexible" because "the logic of testing rival explanations is buried in the statistical machinery." Since there are limits to the kinds of alternative

models that can be considered and compared within the framework of a single study, evidence from multiple studies that examine a variety of consequences of a causal explanation and test the effectiveness of the operationalization can greatly increase the confidence with which causal inferences are made. Because it is usually possible to imagine different consequences of the truth of a causal hypothesis, multiple studies can be used to determine whether each of these consequences holds. Often, if-then conjectures about the nature of the influence process and its observable results require the specification of conditional relationships: If this influence process occurs, we expect that outcome; or, if that result occurs, it could have been produced this way or that way (Cordray 1986). When there are two (or more) rival explanations, conjectures of this type can identify situations in which the explanations would lead to different consequences, thereby permitting a "critical test" of their relative effectiveness. In short, reasoning about the implications of causal explanations can be accompanied by reasoning about appropriate methodological strategies. Since there are numerous ways to gather evidence, multiple research designs, as well as multiple measures and multiple analyses, provide a means of broadening the evidential base.

4. EXAMPLE: REHABILITATION OF HEROIN ADDICTS

To illustrate some key aspects of the process of causal inference, we finally consider an example that focuses on attempts to identify an effective treatment for the rehabilitation of heroin addicts. This example illustrates the importance of considering the causal field and the possibility that a cause may be a disjunction of conjunctions. The process by which evidence can be cumulated from multiple studies, most of which are observational, is also illustrated. In this case the evidence is used to test competing theories of addiction that call for different modes of treatment.

4.1. *Statement of the Problem*

Heroin abuse occurs in widely disparate environments and in varying degrees. The incidence of heroin abuse fluctuates dramatically with geographical location and macrolevel historical events, such as the

end of World War II, major recessions, and the Vietnam War protest movement. In urban ghetto neighborhoods, it is frequently associated with grossly inadequate parental support in early childhood, unemployment in the teen and young adult years, and strong peer influence supporting heroin usage. The natural history of heroin addiction seems to be invariant no matter what the geographical setting. Once heroin abuse has begun, most users go through periods of temporary abstinence but, after six to eight years of heroin use, tend to persist as chronic, daily drug-seeking users. With increasing age, the death rate among chronic users rises rapidly and is substantially higher than the age-specific mortality rates for heroin nonusers (Vaillant 1966*b*).

Individual addiction histories, free of any formal treatment program, involve transitions back and forth between periods of active addiction and temporary abstinence. Persons who start abusing heroin between ages 16 and 22 and who abuse for less than one or at most two years tend to revert to permanent abstinence with no treatment, even if they have been daily users for spells of several months or even one year. Persons who continue with sustained, essentially daily heroin use for more than two years (chronic users) rarely revert voluntarily to even temporary abstinence. In a 25-year follow-up study of 50 heroin addicts from New York City who had used heroin for an *average* of 3 1/2 years at the start of the study, roughly 40 percent achieved stable abstinence by age 42 (Vaillant 1966*a*, *b*, *c*, *d*, 1973). Of addicts who became permanently abstinent, as measured by abstinence for at least three years and without known subsequent relapse, two thirds found stable employment, supported themselves and a family, did not engage in serious alcohol, barbiturate, or tranquilizer abuse, and refrained from the criminal activity that was necessary to support heroin addiction. In the course of an addiction "career," the average addict in the study spent only six years actively addicted. Roughly five more years were spent in jail and one year in a hospital. The average addict was known to have been withdrawn from drugs either in jail or in a hospital a total of nine times and, thus, relapsed a minimum of eight times. From the vantage point of social disability, the average addict was neither in jail nor actually addicted for 13 years out of the 25-year period of the follow-up study. Nevertheless, he was abstinent and fully employed for less than four years out of 25. This characterization of the natural history of heroin addiction forms the baseline against which treatment programs are to be evaluated.

We focus attention here on Methadone Maintenance Treatment (MMT) programs whose aim is *rehabilitation*, defined as termination of and sustained abstinence from heroin use, stable employment, and stable family life (Dole and Nyswander 1965; Dole, Nyswander, and Kreek 1966). We seek to determine whether there is a basis for inferring a causal relationship between X and Y, where $X = MMT$ and $Y =$ rehabilitation. MMT programs vary with respect to at least five identifiable components:

A_1 = proper drug (methadone) dosage,
A_2 = compliance in schedule of administration (daily),
A_3 = presence in the program of sensitive psychological counseling personnel,
A_4 = availability in the program of supportive personnel to assist in job search,
A_5 = supportive home base or peers.

Which of these five components constitutes the true cause of Y has been a source of disagreement. Proponents of a metabolic theory of addiction claim that prolonged daily injections of heroin trigger a metabolic disorder so that no matter how hard an addict may try to stop the use of heroin, the addict is simply overwhelmed by an inner drug-seeking drive and persists in the addiction. MMT is argued to be necessary for rehabilitation, and a high-quality MMT program is believed to be one that combines drug therapy with social and psychological support. The cause, X, of Y is therefore seen as a disjunction of conjunctions, where $X = [A_1 \cap A_2 \cap A_5 \cap (A_3 \cup A_4)]$ and each of the components A_i, $1 \leq i \leq 5$, can be interpreted as an inus condition. Proponents of a psychologically based theory of addiction oppose MMT programs. They point to the effects of peer influence, unstable home life during early childhood, and a social environment accepting of drug use on initiation of heroin abuse and claim that willpower and desire by a motivated addict can generate Y in the absence of drug therapy. Thus, they argue that $X' = [(A_3 \cap A_5) \cup (A_3 \cap A_5 \cap A_4)]$ causes Y and that the amendment of $(A_1 \cap A_2)$, while possibly helpful, is not necessary to achieve Y. A psychologically based theory leading to X' has provided the rationale for both drug-free therapeutic communities and government resistance in many countries to the proliferation of MMT programs. (See Dole and Nyswander [1968] for a lucid statement of both metabolic and psychological theories of addiction.)

4.2. *The Body of Evidence*

The following associational evidence suggesting that X causes Y has been obtained from a diversity of studies in the U.S., Canada, several West European countries, and Hong Kong:

1. Prior to the discovery of methadone as a narcotic blockade, chronic daily users of heroin had a progressively degenerating life, very frequently leading to early death. Except for exceedingly rare isolated cases, Y virtually never occurred in the population of chronic users who persisted in daily use for more than two years.
2. In early MMT programs in New York and later in Sweden, Y occurred with very high frequency when X was operative. $(A_1 \cap A_2)$ alone was only rarely associated with Y. A_1 alone or A_2 alone was never associated with Y in any program anywhere.
3. In Hong Kong, $(A_1 \cap A_2 \cap A_4 \cap A_5)$ was strongly associated with Y.
4. In Hong Kong and Sweden, when methadone was withdrawn or withheld from a control group selected by randomization, $[A_5 \cap (A_3 \cup A_4)]$ alone was not associated with Y.
5. In Vietnam veterans who experienced at most one year of daily use of heroin in Vietnam and returned immediately to the U.S., A_5 and Y were strongly associated (Robins, Helzer, and Davis 1975). At first glance, this study appears to lend strong support to X'. However, careful examination of the entry criteria for MMT programs reveals that no one in the Vietnam veterans study would qualify for MMT. Two years of daily use of heroin was a minimal admission requirement for MMT clinics everywhere. An important feature of this study is that together with (1)–(4), it suggests that two different causal claims are warranted, depending on the causal background. "X' causes Y" is supported for the population of heroin users with at most 1–1 1/4 years of daily use, whereas "X causes Y" is supported for the population of heroin users with two or more years of daily use.

Nearly all associations in (1)–(5) were determined from observational studies. The only controlled, randomized experimentation was in studies of $(A_1 \cap A_2)$ vs. placebo in Hong Kong (Newman and Whitehill 1979) and vs. no treatment in Sweden (Gunne and Gronbladh 1981) to ascertain whether $(A_1 \cap A_2)$ was related to retention in the MMT program of addicts who had already self-selected into the

program and met the admission criteria. The setting for assessing whether X caused Y in the context of MMT is prototypical for the evaluation of a wide range of complex programs in which (a) voluntary self-selection into a treatment program is a condition of entry; (b) it is unknown whether the decision criteria for entry used by volunteers are the same or different from those used by nonvolunteers; (c) the program has a multiplicity of components constituting the treatment; and (d) a control group from the target population cannot be assembled. Conditions (a)–(d) are also present in evaluating rehabilitation programs for chronic alcoholics, family planning programs in developing countries, and manpower training programs, to name only a few instances.

In the context of MMT programs, the effect of $(A_1 \cap A_2)$ is recognizable almost immediately. At a daily dose (usually from 50 to 100 mg), the concentration of methadone in the addict's blood is kept at all times above the threshold for withdrawal symptoms and well below the threshold for narcotic effects (Dole 1980). When the MMT patient is thus stabilized, he/she is functionally normal, protected from narcotic effects by pharmacological tolerance of the drug *and* from withdrawal symptoms by the constant presence of methadone in the blood stream. Thus, there is no question about contiguity of $(A_1 \cap A_2)$ and the initial signs that Y might occur. Since, operationally, rehabilitation means abstinence from heroin use for at least 1 1/2–2 years *and* stable employment and family life, contiguity of X and Y requires that this time period be judged to be adequate for seeing the influence of A_3, A_4, and A_5. In fact, contiguity can clearly be claimed, since prior to MMT, most daily users of heroin are unemployed, heavily engaged in street crime, and have no stable family life.

Although the associational evidence accumulated to date supports the metabolic theory of addiction, several types of additional evidence would enhance the support for this theory. First, it would be desirable to demonstrate that treatment programs designed to provide only social and psychological support for chronic addicts are ineffective in accomplishing their rehabilitation goal. Unfortunately, long-term follow-up data comparable to that collected on MMT programs are not available for drug-free therapeutic communities, which detoxify addicts and assist them in developing the inner resources to adopt alternative, more effective life-styles. Although the withdrawal of methadone from control groups in Hong Kong after initial stabiliza-

tion in an MMT program indicated that the social and psychological support available to program participants was insufficient for rehabilitation in the absence of methadone treatment, it would be desirable to demonstrate that a treatment program designed specifically to provide only social and psychological support was ineffective in rehabilitating chronic addicts. Second, it would be desirable to have an analogue of the Vietnam veterans study for two-or-more-years daily users of heroin. That study forms the only natural experiment of a psychological theory of addiction to date, and it supports the psychological theory for up-to-one-year daily users of heroin. If there had been a population of two-or-more-years daily users of heroin in Vietnam who achieved Y upon returning to the U.S. without $(A_1 \cap A_2)$, the metabolic theory would be in serious doubt. An experiment analogous to the Vietnam veterans experiment for two-or-more-years daily users of heroin would involve the physical transfer of a population of persons who satisfy MMT admission criteria to a new environment where $[A_5 \cup A_5 \cap (A_3 \cup A_4)]$ is available. For example, it would be interesting to know how chronic heroin addicts in New York City who meet the MMT program admission criteria would behave if they were simply transferred to a highly favorable environment several thousand miles away. If Y did not occur under such circumstances, there would be further evidence to rule out X' as a spurious cause of Y for persons eligible for admission to MMT programs. Unfortunately, establishing such an experiment is highly impractical. Third, it would be desirable to identify a biological mechanism triggering a clearly defined metabolic disorder in persons with two or more years of daily heroin use. To date, no biological mechanism that can be claimed to generate a metabolic disorder has been identified.

5. CONCLUSIONS

We have tried to present a global picture of causality in the social sciences, starting from basic philosophical and conceptual issues and proceeding through questions of operational strategies for confirming or refuting claims that X causes Y. A broad unifying discussion in a paper of limited size carries with it the price of lack of thoroughness, which would arise in a sustained in-depth analysis of a particular problem. This, of necessity, must be the focus of another paper. Our aim here has been to isolate the central features of the concept of

causality *and* their relationship to operational strategies for cumulating a body of evidence, an integration that seems to be lacking in the extant literature and is critical for any assessment of causality.

Our basic position with regard to the nature of causes in most social science settings—which stands in sharp distinction to the way most analyses that purport to be causal are currently carried out—is that the plausible candidates for causes are usually disjunctions of conjunctions. This has major implications for the formal modeling and rigorous testing of explanatory theories in that coupled nonlinear dynamical systems will, of necessity, be the natural mathematical language within which to describe conjunctions and their evolution. The variables currently incorporated in the wide-spread applications of path analysis are usually best thought of as inus conditions and thus represent only individual components of causes.

The causes operating in social science settings also represent different types of determination, including extrinsic determination, intrinsic determination, self-determination, and teleological determination. In our view, restricting the concept of cause to extrinsic determination, as recommended by Holland (1986), is artificial and disfunctional, since external influences often interact with internal processes in producing effects. Moreover, mental processes are a major focus of study in the social sciences because they mediate most human action. What requires far more attention is the special nature of self-determination and teleological determination. In many instances the concept of cause has limited usefulness when applied to these types of determination. For example, it is usually not helpful to refer to the relation between an earlier and a later state of a continuous self-maintaining process as causal, since the earlier state has no productive capacity relative to the later state, and it is of primary interest to look for an initiating or sustaining cause of the *process*. Similarly, it is usually not helpful to treat goals, intentions, and motives in the same way that other causes are treated, since they play a distinct role in social science theories.

One implication of the importance of teleological determination in the social sciences is that the temporal ordering of behavior or even of the formation of behavioral intentions is often not a valid indication of causal priority. In the absence of other information, there has been a tendency to make unwarranted causal inferences on the basis of the temporal sequencing of behavioral measures. To make valid causal

inferences about the actions of individuals will require far more direct questioning of individuals and the mounting of longitudinal studies with successive waves of data collection spaced at short intervals.

Criteria for making causal inferences do not seem to us to be specifiable across the full range of scientific inquiry, or even the full range of inquiry in the social sciences. Criteria employed in a diversity of problems are dependent on the nature of the phenomena under study, the end uses of the causality assessments, ever-evolving measurement technology, and the degree of fine-grained detail in proposed explanatory theories. These points are clearly illustrated by Evans's (1976) historical discussion of the evolution of causality criteria in epidemiology, starting in 1890 with the Henle-Koch postulates and changing, of necessity, with the development of new measurement technology and the discovery of new phenomena (e.g., slow viruses), which were associated with different explanatory theories. The criteria put forth by Evans are focused primarily on a setting where scientific explanation is the end use of the causality assessment. These criteria stand in contrast to the modified causality criteria, also in epidemiology, which were put forth by the U.S. Surgeon General's Advisory Committee on smoking and lung cancer, where regulation (at least of cigarette advertising) in a policy setting was the end use of the causality assessment.

In this paper we have outlined the kinds of empirical cues and inductive reasoning that lead to the formation of causal hypotheses and the kinds of evidence needed to confirm them. The stronger the demonstrated consistency of an association under conditions that rule out alternative hypotheses and the stronger the evidence regarding a mechanism that can explain the observed association, the more likely we are to accept the causal hypothesis. Usually the evidence required to confirm a causal hypothesis is cumulated across multiple studies, many of which are, of necessity, observational. Although a wide variety of research designs and analytic techniques are available to assist in gathering evidence to support a causal inference, they are helpful only to the extent that their use is guided and constrained by appropriate subject-matter considerations. No method or set of methods defines causality.

The integrated philosophical and operational framework we have presented represents a set of flexible guideposts that now require systematic incorporation in a diversity of social science investigations.

It is our opinion that this will lead to a systematic evolution of confirmations of causality and a more organized cumulative development of new knowledge in the social sciences.

REFERENCES

Aigner, D. J., and Arthur S. Goldberger, eds. 1977. *Latent Variables in Socio-Economic Models*. Amsterdam: North-Holland.

Aristotle. 1930. *Physics*. In *The Works of Aristotle*, vol. 2, edited by W. D. Ross. Oxford: Clarendon Press.

Ayer, A. J. 1956. "What is a Law of Nature?" *Revue Internationale de Philosophie* 10:144–65.

Bartholomew, D. J. 1982. *Stochastic Models for Social Processes*. 3d ed. London: Wiley.

Basmann, R. L. 1988. "Causality Tests and Observationally Equivalent Representations of Econometric Models." *Journal of Econometrics*, in press.

Beauchamp, Tom L., ed. 1974. *Philosophical Problems of Causation*. Encino, CA: Dickenson.

Berk, Richard A. 1988. "Causal Inference for Sociological Data." In *Handbook of Sociology*, edited by N. Smelser. Beverly Hills: Sage.

Bernert, Christopher. 1983. "The Career of Causal Analysis in American Sociology." *British Journal of Sociology* 34:230–54.

Bielby, William T., and Robert M. Hauser. 1977. "Structural Equation Models." Pp. 137–61 in *Annual Review of Sociology*, vol. 3, edited by A. Inkeles, J. Coleman, and N. Smelser. Palo Alto: Annual Reviews.

Blalock, Hubert M., Jr. 1964. *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

_____, ed. 1971. *Causal Models in the Social Sciences*. New York: Aldine/Atherton.

Brier, Robert. 1974. *Precognition and the Philosophy of Science: An Essay on Backward Causation*. New York: Humanities Press.

Bunge, Mario. 1979. *Causality and Modern Science*. 3d rev. ed. New York: Dover.

Bush, Robert R., and Frederick Mosteller. 1955. *Stochastic Models for Learning*. New York: Wiley.

_____. 1959. "A Comparison of Eight Models." Pp. 293–307 in *Studies in Mathematical Learning Theory*, edited by R. R. Bush and W. K. Estes. Stanford: Stanford University Press.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Carnap, Rudolf, and Richard C. Jeffrey, eds. 1971. *Studies in Inductive Logic and Probability*. Berkeley: University of California Press.

Cochran, William G. 1937. "Problems Arising in the Analysis of a Series of Similar Experiments." *Journal of the Royal Statistical Society*, suppl., 4:102–118.

_____. 1965. "The Planning of Observational Studies of Human Populations." *Journal of the Royal Statistical Society*, ser. A, 182:234–55.

Cole, Jonathan, and Burton Singer. 1987. "A Theory of Limited Differences: Explaining the Productivity Puzzle in Science." Technical Report. New York: Columbia University, Center for the Social Sciences.

Collingwood, R. G. 1940. *An Essay on Metaphysics*. Oxford: Clarendon Press.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Cootner, Paul, ed. 1965. *The Random Character of Stock Market Prices*. Cambridge, MA: MIT Press.

Cordray, David S. 1986. "Quasi-Experimental Analysis: A Mixture of Methods and Judgement." Pp. 9–27 in *Advances in Quasi-Experimental Design and Analysis*, edited by W. M. K. Trochim. New Directions for Program Evaluation, No. 31. San Francisco: Jossey-Bass.

Cox, D. R. 1986. "Comment on 'Statistics and Causal Inference'." *Journal of the American Statistical Association* 81:963.

Crutchfield, James P., J. Doyne Farmer, Norman H. Packard, and Robert S. Shaw. 1986. "Chaos." *Scientific American* 254:46–57.

Darwin, Charles. (1859) 1964. *On the Origin of Species*. Cambridge, MA: Harvard University Press.

Davis, James A. 1985. *The Logic of Causal Order*. Beverly Hills: Sage.

Diaconis, Persi, and David Freedman. 1984. "Asymptotics of Graphical Projection Pursuit." *Annals of Statistics* 12:793–815.

Dole, Vincent P. 1980. "Addictive Behavior." *Scientific American* 243:138–54.

Dole, Vincent P., and Marie E. Nyswander. 1965. "A Medical Treatment for Diacetylmorphine (Heroin) Addiction." *Journal of the American Medical Association* 193:646–50.

_____. 1968. "Methadone Maintenance and Its Implications for Theories of Heroin Addiction." Pp. 359–66 in *The Addictive States*. Association for Research in Nervous and Mental Disease, Volume 66.

Dole, Vincent P., Marie E. Nyswander, and Mary-Jeanne Kreek. 1966. "Narcotic Blockade." *Archives of Internal Medicine* 118:304–9.

Ducasse, C. J. 1966. "Critique of Hume's Conception of Causality." *The Journal of Philosophy* 63:141–48.

Einhorn, Hillel J., and Robin M. Hogarth. 1986. "Judging Probable Cause." *Psychological Bulletin* 99:3–19.

Emanuel, Steven. 1979. *Criminal Law*. New Rochelle: Emanuel Law Outlines.

Evans, Alfred S. 1976. "Causation and Disease: The Henle-Koch Postulates Revisited." *Yale Journal of Biology and Medicine* 49:175–95.

Evans, Alfred S. 1978. "Causation and Disease: A Chronological Journey." *American Journal of Epidemiology* 108:249–58.

Feinstein, Alvan. 1986. "Decisions About Causality." Lecture notes in Quantitative Clinical Epidemiology, Yale University.

Fienberg, Stephen E., Burton Singer, and Judith M. Tanur. 1985. "Large-Scale Social Experimentation in the United States." Pp. 287–326 in *A Celebration of Statistics*, edited by A. C. Atkinson and S. E. Fienberg. New York: Springer-Verlag.

Fisher, Franklin M. 1966. *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Fisher, Ronald A. 1946. *Statistical Methods for Research Workers*. 10th ed. Edinburgh: Oliver and Boyd.

Freedman, David A. 1987. "As Others See Us: A Case Study in Path Analysis." *Journal of Educational Statistics* 12:101–28.

Frege, Gotlob. (1884) 1961. *Die Grundlagen der Arithmetic Eine logischmathematische Untersuchung über de begriff der Zahl*. Breslau: Hildesheim, G. Olms Verlag.

Friedman, J. H., and W. H. Stuetzle. 1982. "Projection Pursuit Methods for Data Analysis." In *Modern Data Analysis*, edited by R. Launer and A. F. Siegel. New York: Academic Press.

Friedman, J. H., and J. W. Tukey. 1974. "A Projection Pursuit Algorithm for Exploratory Data Analysis." *IEEE Transactions on Computers* C-23:881–89.

Frydman, Roman, and Edmund S. Phelps. 1983. *Individual Forecasting and Aggregate Outcomes*. New York: Cambridge University Press.

Gasking, Douglas. 1955. "Causation and Recipes." *Mind* 64:479–87.

Geweke, John. 1984. "Inference and Causality in Economic Time Series Models." Pp. 1101–44 in *Handbook of Econometrics*, vol. 2, edited by Z. Griliches and M. D. Intriligator. New York: Elsevier.

Gibbs, Jack P. 1982. "Evidence of Causation." Pp. 93–127 in *Current Perspectives in Social Theory*, vol. 3. Greenwich, CT: JAI Press.

Glymour, Clark. 1986. "Comment: Statistics and Metaphysics." *Journal of the American Statistical Association* 81:964–66.

Glymour, Clark, Richard Scheines, Peter Spirtes, and Kevin Kelly. 1987. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science and Statistical Modelling*. New York: Academic Press.

Goldberger, Arthur S., and Otis Dudley Duncan, eds. 1973. *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

Good, I. J. 1961. "A Causal Calculus I." *British Journal of the Philosophy of Science* 11:305–18.

_____. 1962. "A Causal Calculus II." *British Journal of the Philosophy of Science* 12:43–51.

Goodman, Leo A. 1973. "Guided and Unguided Methods for the Selection of Models for a Set of T Multidimensional Contingency Tables." *Journal of the American Statistical Association* 68:165–75.

Goodman, Nelson. 1965. *Fact, Fiction, and Forecast*. 2d ed. Indianapolis: Bobbs-Merrill.

Granger, Clive W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37:424–38.

_____. 1980. "Testing for Causality: A Personal Viewpoint." *Journal of Economic Dynamics and Control* 2:329–52.

_____. 1986. "Comment on 'Statistics and Causal Inference'." *Journal of the American Statistical Association* 81:967–68.

Gunne, Lars-M., and Leif Gronbladh. 1981. "A Swedish Methadone Maintenance Program: A Controlled Study." *Drug and Alcohol Dependence* 7:249–56.

Haavelmo, Trygve. 1944. "The Probability Approach in Econometrics." *Econometrica* 12:S1–118.

Hart, H. L. A., and A. M. Honoré. 1984. *Causation in the Law*. 2d ed. Oxford: Oxford University Press.

Heckman, James J. 1987. "Selection Bias and Self Selection." Pp. 287–97 in *The New Palgrave*, vol. 4, edited by J. Eatwell, M. Milgate, and P. Newman. New York: Stockton.

Heckman, James J., and V. Joseph Hotz. 1987. "Do We Need Experimental Data To Evaluate the Impact of Training on Earnings." *Evaluation Review*, in press.

Heckman, James J., and James R. Walker. 1987. "Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite Data." Pp. 247–307 in *Sociological Methodology* 1987, edited by C. C. Clogg. Washington, DC: American Sociological Association.

Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. Orlando: Academic Press.

Hempel, Carl G. 1942. "The Function of General Laws in History." *Journal of Philosophy* 39:35–48.

_____. 1945. "Studies in the Logic of Confirmation." *Mind* 54:1–26, 97–121.

Hill, Austin Bradford. 1965. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58:295–300.

Holland, John H., Keith J. Holyoak, Richard E. Nisbett, and Paul R. Thagard. 1986. *Induction*. Cambridge, MA: MIT Press.

Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945–60.

Huber, Peter. 1985. "Projection Pursuit." *Annals of Statistics* 13:435–525.

Hume, David. (1739) 1896. *A Treatise of Human Nature*. Oxford: Clarendon Press.

_____. (1740) 1938. *An Abstract of a Treatise of Human Nature*. Cambridge: Cambridge University Press.

Hume, David (1748) 1900. *An Enquiry Concerning Human Understanding*. Chicago: Open Court.

Keynes, John Maynard. 1921. *A Treatise on Probability*. London: Macmillan.

Kneale, William C. 1949. *Probability and Induction*. Oxford: Oxford University Press.

_____. 1950. "Natural Laws and Contrary-to-Fact Conditionals." *Analysis* 10:121–25.

_____. 1961. "Universality and Necessity." *British Journal for the Philosophy of Science* 12:89–102.

Kolata, Gina. 1986. "What Does It Mean to be Random?" *Science* 231:1068–70.

Koopman, James S. 1977. "Causal Models and Sources of Interaction." *American Journal of Epidemiology* 106:439–44.

Koopmans, Tjalling C., and O. Reiersol. 1950. "The Identification of Structural Characteristics." *Annals of Mathematical Statistics* 21:165–81.

Kuhn, Thomas S. 1977. *The Essential Tension*. Chicago: University of Chicago Press.

Leamer, Edward. 1978. *Specification Searches*. New York: Wiley.

Lewis, David. 1973. "Causation." *Journal of Philosophy* 70:556–72.

Mackie, J. L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2:245–64.

_____ _____. 1966. "Counterfactuals and Causal Laws." Pp. 65–80 in *Analytical Philosophy: First Series*, edited by R. J. Butler. Oxford: Basil Blackwell.

_____. 1974. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.

Malliaris, A. G., and William A. Brock. 1982. *Stochastic Methods in Economics and Finance*. Amsterdam: North-Holland.

Manton, Kenneth G., Eric Stallard, Max A. Woodbury, H. Dennis Tolley, and Anatoli I. Yashin. 1987. "Grade-of-Membership Techniques for Studying Complex Event History Processes with Unobserved Covariates." Pp. 309–46 in *Sociological Methodology* 1987, edited by C. C. Clogg. Washington, DC: American Sociological Association.

Mill, John Stuart. 1919. *A System of Logic Ratiocinactive and Inductive*. 8th ed. London: Longmans, Green, Reader and Dyer.

Mosteller, Frederick, and John W. Tukey. 1977. *Data Analysis and Regression*. Reading, MA: Addison-Wesley.

_____. 1982. "Combination of Results of Stated Precision: I. The Optimistic Case." *Utilities Mathematica* 21A:155–78.

_____. 1983. "Combination of Results of Stated Precision: II. A More Realistic Case." Unpublished manuscript, Bell Laboratories.

Nagel, Ernest. 1961. *The Structure of Science*. New York: Harcourt Brace Jovanovich.

Newman, R. G., and W. B. Whitehill. 1979. "Double-Blind Comparison of Methadone and Placebo Maintenance Treatments of Narcotics Addicts in Hong Kong." *Lancet* 2:485–88.

Nicod, Jean. 1930. *Foundations of Geometry and Induction*. New York: Harcourt, Brace.

Pearson, Karl. 1933. "On a Method of Determining Whether a Sample of Given Size *N* Supposed to Have Been Drawn From a Parent Population Having a Known Probability Integral Has Probably Been Drawn at Random." *Biometrika* 25:379–410.

Popper, Karl. (1959) 1972. *The Logic of Scientific Discovery*. London: Hutchinson.

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

Reaven, G. M., and R. G. Miller. 1979. "An Attempt to Define the Nature of Chemical Diabetes Using a Multidimensional Analysis." *Diabetologia* 16:17–24.

Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley: University of California Press.

Robins, L. N., J. E. Helzer, and D. H. Davis. 1975. "Narcotic Use in Southeast Asia and Afterward." *Archives of General Psychiatry* 32:955–61.

Rothman, Kenneth J. 1976. "Causes." *American Journal of Epidemiology* 104:587–92.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688–701.

_____. 1986. "Comment: Which Ifs Have Causal Answers?" *Journal of the American Statistical Association* 81:961–62.

Russell, Bertrand. 1908. "Mathematical Logic As Based on the Theory of Types." *American Journal of Mathematics* 30:222–62.

_____. 1948. *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.

_____. 1959. *My Philosophical Development*. New York: Simon and Schuster.

Salmon, Wesley C. 1967. *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

_____. 1980. "Probabilistic Causality." *Pacific Philosophical Quarterly* 61:50–74.

_____. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Simon, Herbert A. 1953. "Causal Ordering and Identifiability." Pp. 49–74 in *Studies in Econometric Method*, edited by W. C. Hood and T. C. Koopmans. New Haven: Yale University Press.

_____. 1957. *Models of Man: Social and Rational*. New York: Wiley.

_____. 1979. "The Meaning of Causal Ordering." Pp. 65–81 in *Qualitative and Quantitative Social Research*, edited by R. K. Merton, J. S. Coleman, and P. H. Rossi. New York: Free Press.

_____. 1986. "The Failure of Armchair Economics." *Challenge* (November–December):18–25.

Singer, Burton. 1986. "Self-Selection and Performance-Based Ratings: A Case Study in Program Evaluation." In *Drawing Inferences from Self-Selected Populations*, edited by H. Wainer. New York: Springer-Verlag.

Skyrms, Brian. 1980. *Causal Necessity*. New Haven: Yale University Press.

Sober, Elliot. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press.

Strotz, Robert H., and H. O. A. Wold. 1960. "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis." *Econometrica* 28:417–27.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Swartz, Marvin, Dan Blazer, Max Woodbury, Linda George, and Richard Landermare. 1986. "Somatization Disorder in a U.S. Southern Community: Use of a New Procedure for Analysis of Medical Classification." *Psychological Medicine* 16:595–609.

Tippett, L. H. C. 1931. *The Method of Statistics*. London: Williams and Norgate.

Tukey, John. 1987. "Configural Polysampling." *SIAM Review* 29:1–20.

Vaillant, George E. 1966*a*. "A 12-Year Follow-up of New York Narcotic Addicts: I. The Relation of Treatment to Outcome." *American Journal of Psychiatry* 122:727–37.

_____. 1966*b*. "A 12-Year Follow-up of New York Narcotic Addicts: II. The Natural History of a Chronic Disease." *New England Journal of Medicine* 275:1282–88.

_____. 1966*c*. "A 12-Year Follow-up of New York Narcotic Addicts: III. Some Social and Psychiatric Characteristics." *Archives of General Psychiatry* 15:599–609.

_____. 1966*d*. "A 12-Year Follow-up of New York Narcotic Addicts: IV. Some Determinants and Characteristics of Abstinence." *American Journal of Psychiatry* 123:573–84.

_____. 1973. "A 20-Year Follow-up of New York Narcotic Addicts." *Archives of General Psychiatry* 29:237–41.

von Wright, Georg Henrik. 1971. *Explanation and Understanding*. Ithaca: Cornell University Press.

Whewell, William. 1967. *The Philosophy of the Inductive Sciences*. 2d ed. London: Frank Cass.

White, Harrison C. 1970. *Chains of Opportunity*. Cambridge, MA: Harvard University Press.

Wolf, F. M. 1986. *Meta-Analysis: Quantitative Methods for Research Synthesis*. Beverly Hills: Sage.

Wright, Sewall. 1921. "Correlation and Causation." *Journal of Agricultural Research* 20:557–85.

_____. 1934. "The Method of Path Coefficients." *Annals of Mathematical Statistics* 5:161–215.

_____. 1954. "The Interpretation of Multivariate Systems." Pp. 11–33 in *Statistics and Mathematics in Biology*, edited by O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush. Ames, IA: Iowa State College Press.

Yates, F., and W. G. Cochran. 1938. "The Analysis of Groups of Experiments." *Journal of Agricultural Science* 28:556–80.

Young, J. Z. 1978. *Programs of the Brain*. Oxford: Oxford University Press.

Zellner, Arnold. 1979. "Causality and Econometrics." Pp. 9–54 in *Three Aspects of Policy and Policymaking*, edited by K. Brunner and A. H. Meltzer. Amsterdam: North-Holland.