9.1 Introduction

The standard counterfactual definition of causation (i.e., that E would not have occurred were it not for C) captures the notion of "necessary cause." Competing notions such as "sufficient cause" and "necessary and sufficient cause" are of interest in a number of applications, and these, too, can be given concise mathematical definitions in structural model semantics (Section 7.1). Although the distinction between necessary and sufficient causes goes back to J. S. Mill (1843), it has received semiformal explications only in the 1960s—via conditional probabilities (Good 1961) and logical implications (Mackie 1965). These explications suffer from basic semantical difficulties,[1] and they do not yield effective procedures for computing probabilities of causes as those provided by the structural account (Sections 7.1.3 and 8.3).

In this chapter we explore the counterfactual interpretation of necessary and sufficient causes, illustrate the application of structural model semantics to the problem of identifying probabilities of causes, and present, by way of examples, new ways of estimating probabilities of causes from statistical data. Additionally, we argue that necessity and sufficiency are two distinct facets of causation and that both facets should take part in the construction of causal explanations.

Our results have applications in epidemiology, legal reasoning, artificial intelligence (AI), and psychology. Epidemiologists have long been concerned with estimating the probability that a certain case of disease is "attributable" to a particular exposure, which is normally interpreted counterfactually as "the probability that disease would not have occurred in the absence of exposure, given that disease and exposure did in fact occur." This counterfactual notion, which Robins and Greenland (1989) called the "probability of causation," measures how *necessary* the cause is for the production of the effect.[2] It is used

[1] The limitations of the probabilistic account are discussed in Section 7.5; those of the logical account will be discussed in Section 10.1.4.

[2] Greenland and Robins (1988) further distinguish between two ways of measuring probabilities of causation: the first (called "excess fraction") concerns only *whether* the effect (e.g. disease) occurs by a particular time; the second (called "etiological fraction") requires consideration of *when* the effect occurs. We will confine our discussion here to events occurring within a specified time period, or to "all or

frequently in lawsuits, where legal responsibility is at the center of contention (see e.g. Section 8.3). We shall denote this notion by the symbol PN, an acronym for probability of necessity.

A parallel notion of causation, capturing how *sufficient* a cause is for the production of the effect, finds applications in policy analysis, AI, and psychology. A policy maker may well be interested in the dangers that a certain exposure may present to the healthy population (Khoury et al. 1989). Counterfactually, this notion can be expressed as the "probability that a healthy unexposed individual would have contracted the disease had he or she been exposed," and it will be denoted by PS (probability of sufficiency). A natural extension would be to inquire for the probability of necessary and sufficient causation (PNS)—that is, how likely a given individual is to be affected both ways.

As the examples illustrate, PS assesses the presence of an active causal process capable of producing the effect, while PN emphasizes the absence of alternative processes—not involving the cause in question—that is still capable of explaining the effect. In legal settings, where the occurrence of the cause (x) and the effect (y) are fairly well established, PN is the measure that draws most attention, and the plaintiff must prove that y would not have occurred *but for* x (Robertson 1997). Still, lack of sufficiency may weaken arguments based on PN (Good 1993; Michie in press).

It is known that PN is in general nonidentifiable, that is, it cannot be estimated from frequency data involving exposures and disease cases (Greenland and Robins 1988; Robins and Greenland 1989). The identification is hindered by two factors.

1. *Confounding*—Exposed and unexposed subjects may differ in several relevant factors or, more generally, the cause and the effect may both be influenced by a third factor. In this case we say that the cause is not *exogenous* relative to the effect (see Section 7.4.5).

2. *Sensitivity to the generative process*—Even in the absence of con-

none" outcomes (such as birth defects) for which the probability of occurrence but not the time to occurrence is important.

founding, probabilities of certain counterfactual relationships cannot be identified from frequency information unless we specify the functional relationships that connect causes and effects. Functional specification is needed whenever the facts at hand (e.g. disease) might be affected by the counterfactual antecedent (e.g. exposure) (see the examples in Sections 1.4, 7.5, and 8.3).

Although PN is not identifiable in the general case, several formulas have nevertheless been proposed to estimate attributions of various kinds in terms of frequencies obtained in epidemiological studies (Breslow and Day 1980; Hennekens and Buring 1987; Cole 1997). Naturally, any such formula must be predicated upon certain implicit assumptions about the data-generating process. Section 9.2 explicates some of those assumptions and explores conditions under which they can be relaxed.[3] It offers new formulas for PN and PS in cases where causes are confounded (with outcomes) but their effects can nevertheless be estimated (e.g., from clinical trials or from auxiliary measurements). Section 9.3 exemplifies the use of these formulas in legal and epidemiological settings, while Section 9.4 provides a general condition for the identifiability of PN and PS when functional relationships are only partially known.

The distinction between necessary and sufficient causes has important implications in AI, especially in systems that generate verbal explanations automatically (see Section 7.2.3). As can be seen from the epidemiological examples, necessary causation is a concept tailored to a specific event under consideration (singular causation), whereas sufficient causation is based on the general tendency of certain event *types* to produce other event types. Adequate explanations should respect both aspects. If we base explanations solely on generic tendencies (i.e., sufficient causation) then we lose important specific information. For instance, aiming a gun at and shooting a person from 1,000 meters away will not qualify as an explanation for that person's death, owing to the very low tendency of shots fired from such long distances to

[3] A set of sufficient conditions for the identification of etiological fractions are given in Robins and Greenland (1989). These conditions, however, are too restrictive for the identification of PN, which is oblivious to the temporal aspects associated with etiological fractions.

hit their marks. This stands contrary to common sense, for when the shot does hit its mark on that singular day, regardless of the reason, the shooter is an obvious culprit for the consequence. If, on the other hand, we base explanations solely on singular-event considerations (i.e., necessary causation), then various background factors that are normally present in the world would awkwardly qualify as explanations. For example, the presence of oxygen in the room would qualify as an explanation for the fire that broke out, simply because the fire would not have occurred were it not for the oxygen. That we judge the match struck, not the oxygen, to be the actual cause of the fire indicates that we go beyond the singular event at hand (where each factor alone is both necessary and sufficient) and consider situations of the same general type—where oxygen alone is obviously insufficient to start a fire. Clearly, some balance must be made between the necessary and the sufficient components of causal explanation, and the present chapter illuminates this balance by formally explicating the basic relationships between the two components.