Introduction

Causation and its Asymmetries

Causation apparently has several different asymmetrical features. In this book I shall say what these features are and how they are related to one another. Here is a list of many of these purported asymmetries:

Time order: Effects do not come before their causes (chapter 3).
Probabilistic Independence: Causes of a given effect are probabilistically independent of one another, while effects of a given cause are probabilistically dependent on one another (chapters 4, 12).
Agency or manipulability: Causes can be used to manipulate their effects, but effects cannot be used to manipulate their causes, and effects of a common cause cannot be used to manipulate one another (chapters 5, 7).
Counterfactual dependence: Effects counterfactually depend on their causes, while causes do not counterfactually depend on their effects and effects of a common cause do not counterfactually depend on one another (chapters 6, 7).
Overdetermination: Effects overdetermine their causes, while causes rarely overdetermine their effects (chapter 6).
Explanation: Causes can be cited to explain their effects, but effects cannot be cited to explain their causes and effects of a common cause cannot be cited to explain one another (chapter 8).
Invariance: If the dependent variables in an equation system are effects of the independent variables, then if one intervenes and changes the value of an independent variable and substitutes the new value in the equations, one has the best prediction of new values for the dependent variables. If on the other hand the independent variables causally depend on the dependent variables and one substitutes new values for the independent variables, then the values one calculates for the dependent variables will be incorrect (chapters 8, 11).
Screening-off: Causes screen off their effects – i.e., controlling for causes makes the probabilistic dependence among effects disappear – while effects do not screen off their causes and effects of a common cause do not screen off one another (chapter 10).
Robustness: The relationship between cause and effect is invariant with respect to the frequency of the cause or with respect to how the cause comes about but not with respect to the frequency of the effect or with respect to how the effect comes about (chapter 11).
Fixity: Causes are "fixed" no later than their effects (chapter 7).
Connection dependence: If one were to break the connection between cause and effect, only the effect would be affected (chapter 6).

Many of these claims are, at best, approximate, yet even those that break down help one to understand causation. One understands causal asymmetry

only when one knows whether these claims are true and how they are related to one another. I shall place particular emphasis on an asymmetry of *causal independence*, which is not in the list, but which is linked to the asymmetry of probabilistic independence. The asymmetry of causal independence is implicated in most of the others, and it has a special role in linking them. I shall argue that causal independence is the central thread in causal asymmetry, even though it is not the whole cloth.

The first two chapters set the stage. Chapter 1 clarifies my presuppositions, and chapter 2 says what the causal relation relates. The remaining chapters, except for the last, discuss the asymmetries listed above and explore the relations among them. The last chapter takes up complications concerning event fusions, overdetermination, and preemption, and in it I state my conclusions. Whenever possible, I prove the assertions I make about relations among the various asymmetries. Unfortunately, these proofs and the precise statements of conditions they depend on are tedious. To make the book readable I have accordingly confined digression, subsidiary arguments, and most of the proofs to separate chapters with asterisks following their numbers.[1] Readers who are not interested in the by-roads and the proofs should skip the starred chapters. Along the way (though mainly in the starred chapters) I formulate explicit conditions and propositions, which I label with capital boldfaced letters. For easy reference, these are listed in alphabetical order in Appendix A. When I prove a proposition that is of interest or to which I shall want to refer, I call it a "theorem" and number it, beginning with the chapter number. For example, theorem 4.1 is the first theorem in chapter 4*. For convenient reference I list the theorems in numerical order in Appendix B. "Theorem" is a grandiose name for these humble results, but I needed some way to make convenient references to the propositions I prove.

Although most of this book explores the precise relations among these asymmetries, some general conclusions emerge: Human beings single out some lawful relations as causal, and they distinguish causes from effects. The reason is ultimately practical. When one factor can be independently manipulated without breaking its nomological links to others, then the factor that can be manipulated can be used to control whatever continues to be linked to it. The possibility of manipulation and control and the related possibility of giving a specifically causal explanation obtain when there is a certain pattern of independence within nomological relations. There are causal relations exactly when these patterns of independence hold. The asymmetries of independence are "objective," but their significance depends

[1] I am borrowing this expository technique from Amartya Sen's masterful *Collective Choice and Social Welfare* (1970).

on human interests. Their centrality to causation and explanation are explained by human interests and are manifest in the intricate relations between the asymmetry of independence and the other asymmetries listed above.

As the first chapter explains, I began with quite a different picture in mind, and these conclusions only emerged painfully out of the detailed arguments in this book. What these sketchy remarks mean and why one should believe them cannot be explained here in the introduction. The book as a whole is devoted to that task. I call this book *Causal Asymmetries*, because it deals with many asymmetric aspects of the causal relation. Articulating them and clarifying the relations among them tells one a great deal about causation.

1

Metaphysical Pictures and Wishes

This chapter tells you where I started, explains how my initial hunches led to difficulties, and lays bare a fundamental ambivalence that nags the discussions in this book. It exposes unclarities in what I was looking for in a theory of causation.

1.1 Metaphysical Theories

In this century, metaphysics has been in ill repute. The word "metaphysics" conjures up an image of a philosopher envisioning "the essential features of reality" and then confidently dictating to artists, moralists, and scientists. Does metaphysics have any genuine content? Are there metaphysical questions that do not collapse into either empirical questions subject to scientific inquiry or semantic questions subject to conceptual analysis?

I sympathize with this "positivistic" skepticism concerning the possibility of conjuring substantive knowledge out of pure contemplation, and there is, I hope, no conjuring in this book. Although questions such as "What are the differences between causes and effects?" are more general than the questions scientists ask, they are not of a different kind. Metaphysical questions concerning causation are continuous with scientific questions. Although more abstract than theories in geology or sociology, the account of causation I defend in this book is intended as an empirical theory (jointly of nature and of human explanatory practices). This theory, like scientific theories generally, is acceptable only if it helps our beliefs and practices to fit together coherently.

1.2 The Question

One might say, "The explosion was caused by the foreman's striking a match," or "Margaret's hitting the tomato with a hammer smashed it." The second claim does not use the word "cause," but smashing is a kind of causing. To say, "The hammer blow smashed the tomato," appears to state physical relations between the hammer and the tomato and causal relations between the hammer blow and the smashing. Causation appears to be a *rela-*

tion between *event tokens*. This appearance may be misleading. Philosophers have questioned whether causation is a *relation*, whether if it is a relation, its relata are *tokens*, and whether if it relates tokens, these tokens are *events*. I shall grapple with these questions in chapter 2, which concludes that appearances are not misleading: Causation is a relation among token events. In this book I shall ask what this relation is. In particular I want to understand its asymmetry.

1.3 How to Begin

More needs to be said about how to proceed. Bear with me in these discussions, for they are abstract and have little immediate payoff. Beginnings are the hardest part (though, if truth be told, there aren't any easy parts!).

In any inquiry one needs criteria upon which to judge the adequacy of one's claims, lest one stab in the dark at one-knows-not-what. Often these criteria are settled. It is obvious what counts as a good answer to the question, "Did Cassius Clay change his name to 'Muhammed Ali'?" and it is obvious how one goes about finding the answer. It is less obvious how one assesses answers to the question, "Is Cassius Clay the same person as Muhammed Ali?" And I think it is less obvious still how one assesses theories of causation. On what basis can one judge one theory of causation superior to another?

Rather than listing a refined set of standards, which I couldn't have enunciated when I began and which have been shaped by the theorizing they are supposed to evaluate, let me begin by describing what I sought, even though I could not find it. I reached some of my conclusions reluctantly ("kicking and screaming" would be more like the truth), and I would like you to understand both why I wound up where I did and why I wanted to arrive somewhere else. In that way you may understand better some of the particular formulations, and you may be able to share the frustration, struggle, and excitement out of which emerged the propositions and theorems listed in the appendices. Stripped of the motivation, the false turns, and the unfulfilled hopes, my conclusions are frozen images of themselves. I shall explain how arguments drove me from "initial" hunches to articulated conclusions.[1]

Criteria to judge answers to metaphysical questions are like descriptions of ideal lovers. They are vague, conflicting, and subject to rewriting as one explores real alternatives. My commitment to the following picture is now

[1] What were my "initial" predispositions? Did I have them when I first thought about causality at age six, or when I studied science in high school and college, or when I first seriously began to wonder about causation, when I was finishing a dissertation in philosophy of economics? My "initial" predispositions are unavoidably my current reconstruction of the hunches that have kept me struggling with this difficult material.

cautious, qualified, and conflicted, but I still feel its pull, and my conclusions may be more intelligible in relation to this picture.

1.4 The Initial Picture

Here is the picture of causation I began with:

Causation is "objective." It is a relation "in the objects." "Out there" are causal relations among events. Most of these relations would obtain if there were no one to think about them. Without humans to notice them, large meteorites would still make craters in the Earth's surface, though meteorites would not be called "meteorites," and the results of their collisions would not be described as "craters." Nonhuman cognizers might describe meteorites and collisions in radically different ways. But substances such as meteorites, events such as collisions, and (*pace* Earman 1976) causal relations such as that holding between a collision and the creation of a crater do not depend on minds, whether they be human or Martian. Idealism and phenomenalism, which deny that anything is independent of mind, are false. There are substances and events that are independent of mind. If this general realism is untenable, it remains the case that events and causal relations are no *more* mind dependent than are substances.

Events resemble substances. Like substances, the same event may be picked out by many different descriptions. World War I is the war that began with the assassination of Archduke Ferdinand. The properties of events and substances are not exhausted by the descriptions we offer of them. Just as there are many different kinds of substances – blocks, stomachs, beetles, and principles – so there are many different kinds of events – lives, performances, conquests, and conversions. Some substances, such as cats and paper clips, have natural boundaries, while others, such as the aluminum comprising the bottom half of a beer can, are separated from other substances only by our descriptions. Similarly, some events, such as the great Chicago fire, have natural boundaries, while other events, such as the third hour of the D-Day landing, do not. Events enter into many different relations with one another. Some of these are the same as relations among substances. Events may, like substances, be part of one another, and events typically have spatial relations. Like Chicago itself, the great Chicago fire was located in Illinois. Events may be larger or smaller than other events. Like substances, events may have aesthetic qualities, and they may be the objects of psychological states.

But events are not substances, and they enter into relations that substances do not enter into. One event can be a temporal part of another (Mellor 1995, pp. 122–3). Pickett's charge is a temporal as well as a spatial part of the Battle of Gettysburg. The second day of the battle is a purely

temporal part of it. A substance, on the other hand, cannot be a temporal part of another substance. Brian's-cat-during-the-third-day-of-its-life is either not a substance, or it is identical with Brian's cat. Substances, unlike events, are wholly present in single instants. Although some substances may *last* longer than others, substances cannot *be* temporally longer or shorter than one another. (To talk about how long a substance is in existence is arguably to talk about the length of an event.) Events, unlike substances, have temporal *dimensions,* not just temporal locations. When some people live longer than others, it is their lives, not the people themselves that have different temporal "sizes."

Finally, and crucially, events enter into causal relations with one another. These relations are apparently irreflexive, transitive, and asymmetric. Causes apparently never occur after their effects, and perhaps always precede them. Effects seem to be counterfactually dependent on their causes. Causation seems like a "glue" attaching events and like a "force" making things happen. Causation is reflected in regularities of at least a probabilistic form. Causation seems connected to intervention and manipulation: One can use causes to "wiggle" their effects. Causal knowledge seems crucial in decision making.

There are reasons to feel queasy about this sketch, because it says nothing about the close connection between causation and explanation. To cite a cause of an event is to explain the event. Since explanation is a human activity linked to human interests, this intimate bond between causation and explanation threatens the objectivity of causation. The picture of mind-independent causal relations among concrete events nevertheless motivates my work. This sketch is not particularly idiosyncratic, though not everyone finds it attractive. I no longer think it is tenable, and the detailed portrait of causality this book draws differs considerably from the initial sketch. But without the picture I wouldn't have had any idea what questions to ask or what sorts of answers to look for. The arguments I make and the strategies I employ should be understood against the background of this initial picture.

1.5 Wishes

The picture gave shape to my questions, but why bother asking these questions, and how can one judge purported answers? Hopes or wishes concerning what a satisfactory theory of causation ought to achieve also drive this inquiry. In a moment I will codify these aspirations as criteria of adequacy, but these are misleading, since I didn't know precisely what I was looking for when I began and I am still unsure about what a good theory of causation ought to do. One of the most difficult parts of metaphysics is to determine what metaphysical theories ought to achieve.

As I conceive it, my task is not to *analyze* causality or to define the term "causes." My job is instead to formulate general truths concerning the causal relation and what it has to do with other relations among events. A theory of causality ought to clarify what role causality plays in human practices such as explaining or making decisions and to explain why people (including scientists) hold the beliefs about causality that they do. One would like a theory of causality to make sense of the methods people use to determine what causes what and perhaps even to improve them. I sought a theory of causality that would link together beliefs about causality and practices of causal inference, and I hoped that a theory of causality would help answer questions such as why people know so much more about the past than the future. A theory of causality that analyzed causal relations in terms of simpler notions would be an exciting achievement, but I never hoped for such a reduction. Since causal claims are so fundamental and pervasive, such a reduction has always seemed unlikely.

According to John Mackie, theories of causality determine what causal claims *mean,* and they specify what causality *is* "in the objects" (1980, p. 1). One can go about the first task in two radically different ways. First, like an anthropologist or linguist, one might ask what people in fact mean when they make causal claims. One would regard a question such as, "Are causes regarded as necessary conditions for their effects?" as an empirical question about what people believe. When these beliefs are part of general linguistic competence, such questions can be answered by consulting linguistic intuitions. But they are still empirical questions about usage and everyday belief. Mackie argues, incidentally, that the meaning of causal claims is partly counterfactual: When we say that *a* is a cause of *b*, we mean that *a* and *b* occurred and, in the circumstances, if *a* had not occurred then *b* would not have occurred. No theory of causality can avoid the task of asking about the meaning of causal claims, because there are limits to how far a theory of causality can diverge from what people take causation to be. No acceptable theory of causation can make the meaning of causal claims bizarre and inexplicable. Causal language may be full of mistakes, but if a theory of causation finds only mistakes or if it makes the mistakes it finds inexplicable, then it undercuts its claim to be a theory of *causation.*

Analyzing the meaning of causal claims need not be understood as linguistic description. It can instead be an exercise in applied logic or conceptual analysis. If one begins with constraints that adequate definitions of concepts must satisfy, like those imposed by empiricists such as Hume or Mach, one can offer analytical reconstructions that revise and correct the faulty definitions people actually accept. In this way Bridgman proposed substituting operational definitions of scientific concepts for unsatisfactory nonoperational definitions – regardless of how generally accepted everyday

nonoperational definitions might be (1938). Rather than discovering what beliefs are implicit in causal claims, one might explore the "logic" of causal claims and of their relations to other sorts of claims. A conceptual analysis of causation, just like a theory of what causation is in the objects, must explain the relevant linguistic phenomena, but it is not limited to providing a description of them.

I think it is futile and misleading to draw a line between stating the *meaning* of causal concepts and using those concepts to make substantive assertions about the world. Attempts to provide a correct *analysis* of causality have implications for what causality is "in the objects," and theories of what causality is have implications for the analysis of causal concepts. My skepticism about the value of separating claims about meaning and claims about the world originally derived from Quine's critique of the distinction between analytical claims that are true by virtue of the meanings of the terms they contain and synthetic claims that are true by virtue of the way the world is (1953, 1960). But it is ultimately independent of concerns about semantics. Like Putnam (1962), my point is methodological. I find it futile rather than impossible to demarcate statements that give the meaning of theoretical notions from statements that use those notions to make synthetic claims. Whether or not one is persuaded by Putnam's general argument against classifying claims as analytic or synthetic, there are special grounds for skepticism about distinguishing the analysis of causation from theories of what the causal relation is. Causation is both so abstract and so basic to human thought that claims about causation, whether intended as analytic or synthetic, will have fundamental implications for our beliefs and meanings. Considering the meaning of causal claims involves a consideration of the beliefs of those who make them. Such consideration is an unavoidable part of this book, because those beliefs may be true, and they roughly pick out the relation that I seek to understand. The aim is to provide true assertions about the causal relation.

I have boiled down the various desiderata into the following five criteria for evaluating theories of causality:

1. *Intuitive fit*: A theory of causation should fit our "intuitions."
2. *Empirical adequacy*: A good theory of causation should "fit the facts" and permit a coherent construal of human practices, including especially scientific practices.
3. *Epistemic access*: A good theory of causation should explain how to find out what causes what.
4. *Superseding competitors*: A good theory of causation should be better than competing theories and help explain why they succeed and fail.
5. *Metaphysical fecundity*: A good theory of causation should clarify the links between causation and other relations, such as temporal relations, and it should help one to answer other metaphysical questions.

1. Intuitive fit: A theory of causation should imply that paradigm causal claims are true. If a theory of causation implied that Gertrude's drinking poison did not cause her to die, that the Titanic's striking an iceberg did not cause its sinking, and that Darwin's publication of *The Origin of Species* did not cause people to reject the theory of divine creation, then that theory of causation would probably be mistaken. Causal language and causal beliefs are *data*. One might be willing to entertain a theory of causation that counted the above claims as false, if it offered a satisfactory explanation for how people could be so persistently in error, but even then causal beliefs remain crucial data.

A theory of causation should also make possible explanations of how people come to use causal language as they do, how they come to have the causal beliefs that they have, and why their causal claims mean what they do. If causation turns out to be a highly esoteric notion, then one must explain how, nevertheless, people acquire causal concepts and make the causal claims that they do. A theory could not be both correct and a theory of *causation* otherwise. A theory of causation fits our intuitions if it permits one to explain why we believe what we do about causal relations.

2. Empirical adequacy: This is a tricky requirement. One cannot test theories of causality in the laboratory. They are too abstract for that, and they are theories of human practices as well as of relations among events. "Testing" will inevitably be indirect and controversial. Nevertheless some theories of causality make for a neater, more flexible, more fruitful, more usable body of theories than others. A theory of causation should also illuminate the role of causal notions in other theories. The account of causation developed here explains and justifies features of experimental, explanatory, and inferential practices.

The discovery of causes has traditionally been thought to be a central task for the sciences, and satisfactory accounts of features of science such as explanation or confirmation will bear upon and be influenced by accounts of causation. The study of causation belongs to philosophy of science as well as to metaphysics. Causal language is as prevalent in the laboratory as in the kitchen. The criterion of empirical adequacy is accordingly analogous to the criterion of intuitive fit. Just as intuitive fit demands congruence with everyday causal beliefs, so empirical adequacy requires that an adequate theory of causation explain why commitment to particular scientific theories leads scientists to their causal beliefs. The task is more complicated in the case of the sciences, because scientists read philosophy and attempt to regiment their language to fit their philosophical commitments. Since these commitments are in many cases inconsistent with scientific practice, theories of causation will have to call either for a reform of scientific practices or for a revision in the philosophical positions many scientists adopt. In my