1994

Propensities, Chance, Causation, And Contrastive Explanation

Christopher S. McCurdy

Propensities, Chance, Causation, and Contrastive Explanation

by

Christopher S. I. McCurdy

Department of Philosophy

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
April 1994

ISBN 0-315-93200-7

Abstract

Propensities, Chance, Causation, and Contrastive Explanation

A pragmatic account of scientific understanding is used both to examine and to unify fundamental questions concerning the propensity interpretation of probability and theories of chance, causation, and explanation. One of the most important problems to be addressed is the problem of defining homogeneous reference classes in theories of chance, causation, and explanation. The consistency of the propensity interpretation is defended against traditional criticisms such as "Humphreys's paradox." It is demonstrated that the application of this interpretation to theories of chance and probabilistic causation provides insights into problems common to both theories. Various approaches to causation are examined, including those based on identifying sufficient causal factors, necessary causal factors, and contrastive causes. These insights are applied to quantum mechanics and are presented in terms of a set of controlled experiments. The study of quantum mechanics focuses on the paradox of the two slit experiment and quantum logical and quantum probabilistic attempts to resolve this paradox. Finally, the analysis of chance and causation provides the basis for a version of the contrastive theory of explanation. This theory of explanation provides a unique understanding of the nature of explanation, and lessens the impact of the problems of homogeneity and of explanatory ambiguity.

Acknowledgements

I would like to express my sincere gratitude to the following people. Of course, their influence has been greater than can be described here. First, I would like to thank my parents. Without their encouragement and support I certainly would have never even attended university. Second, I am deeply indebted to Professors William L. Harper, John M. Nicholas, and Kathleen Okruhlik for providing continued advice, encouragement, and support. My interest in scientific explanation and my motivation for understanding the mechanism of contrastive explanation originated in discussions with John Nicholas. My supervisor, William Harper, provided the inspiration for my interest in probabilistic causation, especially in issues related to Humphreys's paradox and the interpretation of chance. Kathleen Okruhlik became my advisor when I first entered the Ph.D. program and, I dare to say that, with every problem and decision that I faced, she both listened to my perspective and provided indispensable advice. Finally, my wife has not only provided the patience and understanding necessary to live with a person writing a dissertation, but she has also proofread numerous drafts of the dissertation.

Research for part of this dissertation was supported by a fellowship from the Social Sciences and Humanities Research Council of Canada (award number 452-90-2513).

Table of Contents

List of Figures

List of Appendices

Introduction

The topics and issues discussed in this dissertation revolve around the problem of defining the "broadest homogeneous reference class," referred to hereafter as the problem of homogeneity. I was first introduced to this problem as it arises in theories of scientific explanation, and I will begin by defining the problem as it is commonly presented in those theories. One major presupposition in theories of explanation is that events are explained in terms of, or relative to, a reference class. Furthermore, most theories of explanation require that two conditions are placed on the definition of the reference class. First, the reference class must be homogeneous; that is, the reference class must be partitioned into only those events that are statistically or causally relevant. Second, the reference class must be as broad as possible; that is, the reference class must contain as many events as possible. Typically, there is no problem in meeting each of these conditions individually. However, a tension is created when these two conditions are imposed together: imposing the condition of homogeneity has a tendency to narrow the reference class, and imposing the condition of broadness has a tendency to create inhomogeneity. In general terms, the problem of homogeneity is the problem of defining a reference class that is at once homogeneous and as broad as possible. The overall aim of this dissertation is to demonstrate the manner in which this problem arises in the areas of explanation, causation, and chance.

From the time that I was introduced to the problem of homogeneity, I have been insistent that this problem was not a problem that had to be solved, or could be solved, by any account of scientific explanation. I was convinced that theories of explanation may have to find a way to live with, or avoid, the problem of homogeneity, but finding *the solution* to the problem of homogeneity was a task for accounts of causation. I have now come to the conclusion, however, that many approaches to causation are in a situation that is similar to that of accounts of explanation: the best they can do is avoid the problem of homogeneity--they have no hope of solving it. The reason that many accounts of causation cannot hope to solve the problem of homogeneity is that this problem is produced by the notion of chance, and any account of causation that presupposes chance-based criteria for causation, or employs "chance-like thinking," is doomed to inherit the problem of homogeneity. If this dissertation could be presented from the last chapter to the first chapter, it would tell the story of my search for the roots of the problem of defining the broadest homogeneous reference class. The aim of this dissertation is to demonstrate the usefulness of the propensity interpretation of probability in understanding and solving many of the traditional problems in the interpretations of chance, causation, and explanation that are associated with the problem of homogeneity.

The core of this dissertation is the propensity interpretation of the probability calculus. The reason for utilizing the propensity interpretation is to provide a formal connection between a probability distribution and the experimental setup or system responsible for the events over which a probability function is defined. Consequently, the propensity interpretation replaces the classical two-place probability function (defined over the unit interval and an event space) with a three-place relation (defined over the unit interval, an event space, and a system): the propensity function assigns *values* in the unit interval to the *members* of a Boolean σ-algebra for a *system*. Alternatively, propensities provide a formal representation of an experiment by providing the relationship between the experimental setup, the possible results produced by that setup, and the probabilities that the setup will produce the particular results.

Many propensity interpretations are deeply metaphysical, especially the theories of Karl Popper (1957, 1959a, and 1959b) and Ronald Giere (1973 and 1979). This dissertation has no metaphysical motivation for accepting the propensity interpretation; as outlined above, the reasons for accepting a propensity interpretation are largely formal. Consequently, the account of "propensities" given here attempts to remain neutral with respect to the metaphysical status of propensities. Propensities may be real, or they may not; they may correspond to physical dispositions, or they may not.

In fact, the propensity interpretation given here is compatible with both subjectivist and frequentist interpretations of probability in the following manner. As it is presented here, the distinguishing feature of the propensity interpretation is that it represents the probability function as a three-place relation rather than a two-place relation. However, the method for determining propensity values may be subjective or objective, and the interpretation of "a propensity'' may be subjective or objective. There are, of course, many frequentist versions of the propensity interpretation of probability, most notably Ian Hacking 1965, David Miller 1991, and Popper 1957, 1959a, and 1959b.

Also, there are many discussions of the connections between traditional frequentist interpretations of probability, such as Hans Reichenbach 1949 and Richard von Mises 1964, and the propensity interpretation; examples include James Fetzer 1974c, Ian Hacking 1973, Ronald Giere 1973. Popper himself is a frequentist turned propensity theorist. Many subjectivist interpretations can also be interpreted as providing a propensity interpretation of probability. Notable examples of subjectivist interpretations that employ propensity-like reasoning include I. J. Good 1983; David Lewis 1980; and Brian Skyrms 1980, 1984, and 1988.

Some authors, specifically Paul Humphreys (1985), Peter Milne (1986 and 1987), and Wesley Salmon (1979, 1984, 1988 and 1989), have argued that the propensity interpretation is inconsistent with the classical axioms and theorems of the probability calculus. Specifically, it is argued that the ''dispositional'' nature of propensities, created by the asymmetrical relationship between the system and its events, is inconsistent with the inversion theorems of the probability calculus. Chapters 1 and 2 provide a general semantics for the propensity interpretation and defend the foundations of the propensity interpretation against the attack of these authors. In chapter 2 it is argued that there is neither motivation nor justification for the robust causal interpretation of conditional propensities that is presupposed by these authors.

This analysis demonstrates, among other things, that individual propensity functions provide only statements of stochastic independence, and that causal independence can be determined only by comparing results of different propensity functions.

The propensity interpretation also provides the basis for an examination and an interpretation of chance. Chance and conditional chance are both represented as selection functions on a family of propensity functions. The most important consequences of this representation of chance is that it demonstrates that, on the one hand, chance-like thinking plays a large part in our understanding of experimentation, while on the other hand, the problem of homogeneity originates in chance-like thinking. Specifically, the idealized nature of chance provides the basis for our understanding of ideal and generalized experiments, and this understanding is essential to scientific and experimental method. But the idealized nature of chance also calls into question the physical meaning of statements concerning chance because chance is often defined over inhomogeneous situations.

This account of chance and experimentation provides the basis for an investigation into the relations between two distinct approaches to probabilistic causation. One approach, the sufficiency approach, bases its analysis of causation on the identification of sufficient positive causal factors. The other approach, the necessity approach, bases its analysis of causation on the identification of necessary positive causal factors. It is demonstrated that, in both approaches, the central criterion for identifying these causal factors makes an appeal to chance. Hans Reichenbach's ([1956] 1991) frequentist attempt to define causal relevance in terms of statistical independence and his "screening off" condition is presented as an example of the sufficiency approach to causation. Also, I. J. Good's (1983) development of a "causal calculus" in terms of a subjectivist's formulation of the weight of evidence is presented an example of the necessity approach to causation.

One result of this investigation into these two approaches to causation is the determination that the appeal to chance that is made by both approaches to causation causes these them to "inherit" the problem of homogeneity. Consequently, both Reichenbach's and Good's theories inherit the problem of homogeneity. Furthermore, the study of causation in terms of ideal experiments produces questions concerning the physical meaning of the results. To avoid the problem of homogeneity, accounts of causation must avoid appeals to "chance," and to the idealized results of chance-like thinking. Two strategies for avoiding the problem of homogeneity are introduced. One method of avoiding these kinds of appeals is to conduct controlled experiments. Chapter 3 provides an analysis of the two slit experiment (performed with bullets, water waves, and electrons) in terms of a group of controlled experiments. The empirical results of these experiments are represented in terms of two principles: the principle of strict summation, which provides a statement of the relationship between three different experiments; and the principle of strict composition, which provides a statement of the relations between the components of one (more complex) experiment.

There are two conclusions drawn from the analysis of the two slit experiments. First, as a general methodological point, it is demonstrated that the analysis of causation should proceed in terms of relations between propensity functions for the same event but different experiments, rather than in terms of relations between conditional propensities for different conditioning events and the same experiment. The second point is concerned with the interpretation of quantum mechanics and the two slit experiments performed with electrons. It is demonstrated that the "paradox" of the two slit experiment results from the assumption that there is a single consolidated causal explanation of the manner in which the principle of strict summation is violated and the principle of strict composition holds. Consequently, the paradox of the two slit experiment is not based on exclusively formal considerations. Finally, it is argued that strictly formal attempts to solve the paradox, such as the quantum logical arguments of Hilary Putnam (1979) and the quantum probabilistic arguments of Luigi Accardi (1984) fail to solve the paradox. Given this analysis of the two slit experiment, a brief characterization of current interpretations of quantum mechanics is provided.

Finally, the analysis of propensities, chance, causation, and experimentation are applied to the domain of scientific explanation. In particular, contrastive explanation, a theory of explanation based on answering contrastive why-questions of the form "why P *rather than* Q," is examined. In the first step of this examination, an account of contrastive explanation is developed in terms of Peter Lipton's (1991a and 1991b) notion of a "corresponding cause" and a principle based on John Stuart Mill's (1904) "difference condition." This account expands upon Lipton's account by providing a detailed account of the role of the foil (Q above) in contrastive explanation. This is accomplished by providing an understanding of the "rather than" relation in contrastive explanation. It is argued that this "rather than" relation confers an important advantage on the contrastive approach to explanation. Specifically, it is argued that contrastive explanation is better equipped to avoid the problem of homogeneity and that non-contrastive explanation, much like chance and causation, will encounter the problem of homogeneity.

In conclusion, the development of the theories of chance, causation, and contrastive explanation in terms of the propensity interpretation enables one to identify the sources of the traditional problem of homogeneity. Ultimately the source of this problem is the consideration of ideal circumstances because these circumstances typically include inhomogeneous causal factors. By considering specific situations--situations characterized by homogeneous causal factors--the problem of homogeneity can be avoided. This realization is not new. But the propensity interpretation provides a unique method for identifying the source of this problem in all three areas, as well as a method for avoiding this problem in all three areas.

1

Propensities and Chance

1.1 Introduction

The aim of this chapter is to provide an outline of both the propensity interpretation of the probability calculus and the propensity interpretation's relation to chance. The primary feature of the propensity interpretation is that it presupposes and reinforces a distinction between "systems" and "events." The basis for this distinction is the asymmetrical relationship between experimental setups and the results of experiments. Furthermore, this distinction provides the basis for developing the notion of chance in terms of chance in a situation. This chapter initiates a program to emphasize the temporal asymmetry involved in the propensity interpretation, and to deemphasize, or remove, the traditional emphasis on causal or "dispositional" asymmetry.

The first stage in developing the propensity interpretation consists in examining the definition of, and relationships between, "systems" and "events." This is accomplished by providing a detailed analysis of the two slit experiment, performed with electrons. In this analysis, systems are depicted as representations of experimental setups, described in terms of those factors or variables that are statistically relevant to the production of certain outcomes. It is the notion of "production"--rather than some stronger notion of causal dependence--that is considered to be central in characterizing the dispositional, or "forward looking," nature of propensities. Ultimately, the propensity function is demonstrated to be a useful tool in representing the relationships between experimental setups and their outcomes. For the sake of these discussions, the concept of a "version" of an experiment is introduced informally here; it is formally developed later. The analysis of experiments, as well as the relationship between propensity functions and the notion of statistical relevance, is discussed and developed throughout this chapter and the rest of the dissertation.

The second stage in developing the propensity interpretation of probability involves a continuation of the discussion of the two slit experiment, but is primarily concerned with the notion of chance. It becomes obvious that the propensity interpretation is particularly well suited to the formalization of intuitions about chance and conditional chance. To begin, an account of chance in a particular situation is presented. Then the account of chance in a situation is extended to provide an account of chance in set of situations, where that set of situations is described by a partition. Finally, the extended account of chance in a situation is further extended to produce a description of chance (itself) as a selection function on a partition of situations that is described by the statistically relevant systems, causal factors, or variables. This development of chance provides insights into the activity of experimentation, and supplies the basis for a formal definition of a version of an experiment (as introduced in the first stage). Also, a connection is established between systems and causal histories of events. Ultimately, the analysis of chance, experimentation, and causal histories, enhances the presentation of the semantics for the propensity interpretation itself.

The third stage extends the analysis of chance to consider conditional chance. In the end, conditional chance is represented as a selection function on conditional propensities. Through this analysis, it is demonstrated that chance must be defined over (causally) inhomogeneous situations. As a result, it is argued that, although the notion of chance is useful, the notion of chance cannot be physically meaningful in all situations. Finally, it is maintained that the notion of a single generalized two slit experiment should be understood as the result of chance-like thinking and, although ideas of generalized experiments are useful, they fail to be physically meaningful in some situations.

1.2 Propensities and experimental setups

The aim of this section is to provide a semantics for the propensity interpretation of probability. First, the notation that is adopted throughout the dissertation, along with many of the conventions involved with the notation, is introduced and discussed. Then a brief and formal consideration of the semantics for the propensity interpretation is provided. However, most of the semantics is revealed through an informal consideration of one version of the two slit experiment.

Propensity statements are of the form '**Pr**(A:S) = p,' and are read as "the propensity for a system S to produce an event A is p." The symbols occurring within the parentheses and before the colon denote propositions concerning the occurrence of events. Frequently, reference is made to the events themselves rather than to the propositions. Symbols occurring within the parentheses, but after the colon denote propositions comprising a set of background conditions that describes an experimental setup or system. Again, reference is often made directly to the background conditions, experimental setups, or systems. For clarity, the symbols '**Pr**' denote propensity functions, '**P**' denote probability functions, and '**Ch**' denote chance functions.

A propensity function is a numerical measure defined over a Boolean σ-algebra $\mathcal{B}$ on a set of events Ω, where each member of Ω is a possible outcome of some experimental setup or system. The propensity function is a probability measure: **Pr** assigns a propensity value--a number in the unit interval [0, 1]--to each member of $\mathcal{B}$ according to the axioms and theorems of mathematical probability. Symbols such as p, q, and r are used to denote propensity values. It is important to note that the propensity function takes both members of the σ-algebra $\mathcal{B}$ (defined on a set of possible outcomes Ω) and the system S as its domain: a propensity function is defined *over* a Boolean σ-algebra on a set of possible outcomes and *for* a system. The distinction between, and definition of, the set of outcomes in Ω (and consequently among members of the σ-algebra $\mathcal{B}$) and the system S depends primarily on a consideration of the elements and properties of an experiment.

A system is, quite simply, the experimental setup as it is arranged prior to the running of the experiment. A system can be represented as the statistically relevant events that *actually occur* at some time *before* the experiment begins. The notion of statistical relevance used here is discussed and developed throughout this chapter. The set of possible outcomes of an experiment are composed of measurable events that *either occur or do not occur* at some time *after* the experiment begins. Alternatively, the occurrence of the system-events is intentionally brought about (or allowed) by the experimenter in order to initiate the experimental process, whereas the occurrence of the outcome-events is spontaneous, once the experiment is initiated. Given the temporal nature of this distinction and the assumption that systems are described only by relevant events, once the experiment is over, the sequence of "relevant" events leading to the outcomes of the experiment are given in large part by the event description of the system. The only relevant events not given by the event description of the system are those events that occur between the time that the experiment is started and the time that the outcome occurs.

In order to better understand the analysis of experiments and propensities in terms of systems and possible outcomes, consider the application of the propensity interpretation of probability to a particular experiment. The particular experiment is the two slit experiment, originally proposed by Thomas Young in 1803. The experimental setup is described as follows. A source of particles (or waves) is placed in front of two parallel barriers. The first barrier, the diaphragm, contains two narrow slits that can be opened and closed. The two slits are placed at equal distances from the origin line O-L which runs perpendicular to the two barriers and through the source. These slits are just wide enough for the emissions to pass through, and unless the emissions negotiate the slits they are either absorbed or reflected by the diaphragm. If the emissions do pass through the diaphragm they are absorbed by the second barrier, the screen. For the sake of analyzing the results, those emissions that are absorbed by the screen must be recorded (in some fashion) as "hits" on the screen.

A device that is capable of detecting the arrival of a source emission at some designated region R is placed in front of the screen. The detector is movable along the face of the screen. Thus, the position of the detector can be expressed as a function of the distance x from the origin line. This system is shown diagrammatically in figure 1.1.

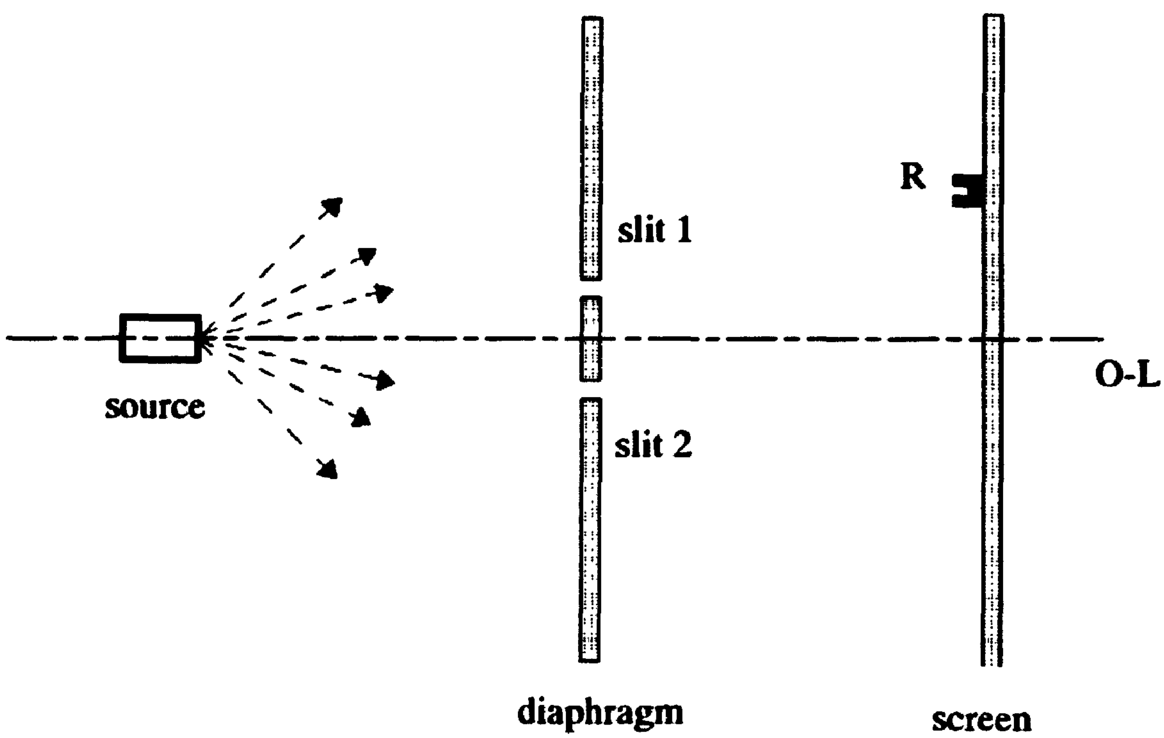

Figure 1.1. The basic two slit experiment.

The experiment is carried out in the following manner. At some time t_j, the source is allowed to emit a particle (or a wave) in the direction of the diaphragm. Only those trials in which an emission reaches the screen are counted as statistically relevant. Consequently, at some time t_k later than t_j the emission hits the screen and is recorded as either a hit by the detector in the region R, or a hit on the screen. Given the system and the experimental procedure described above, the event space consists of the events R and ~R, where R = "the emission hit region R at some time t_k." Finally, propensity values can be calculated from the number of hits in region R and the number of total hits.

There are many factors that can affect the results of running an experiment on the system described above. These factors include the separation between the source and the diaphragm, the separation between the slits, the size of the slits, the separation between the barriers, the area of region R, the distance x between the region R and the origin line O-L, and the symmetry of the emission pattern. The most important factors in the present discussion are the type of emission (particle or wave) and the status of the slits (both open, slit one open and slit two closed, or slit one closed and slit two open).

In order to run the two slit experiment, the state of each of these factors must be fixed, and in order to define a propensity for this experiment, the state of each of these factors must be determined or described. Each of these factors is evident before the start of the experiment and can be described as an event. Examples of events are "slit one is 5mm from the origin line at time t," "slit one is 2mm wide at time t," "slit one is open at time t," and "the source emits electrons." Thus, the system S denotes a conjunctive event, wherein each conjunct provides a description of the status of a factor that may influence the results of the experiment. The events representing the system must occur at some time t_i before the start of the experiment at time t_j. Consequently, the system S itself (as an event) is said to "occur" at some time t_i before the start of the experiment.

Consider performing the two slit experiment with a source that emits electrons. The source is an "electron gun" that is capable of firing electrons of approximately equal energy at the diaphragm. The diaphragm is simply a thin metal plate with two narrow slits. The screen is some suitable electron absorbing plate that is also capable of recording, or at least detecting, the impact of electrons. The screen could be an electron-sensitive film with a pre-marked region R, or an array of geiger counters with one designated as the detector for R. Thus, the experimenter can record the number of hits in region R and the number of hits on the screen. Although this experiment is incapable of being done in the idealized manner in which it is presented here, the results can be discussed (as they should arise) based on results of similar experiments with other particles. For example, a similar experiment (performed with neutrons) is reported in A. J. Leggett 1986.

Suppose that all the variables discussed above are fixed in some appropriate manner, except for the status of the slits as open or closed. The space of possible outcomes is expressed in terms of the event R = "the electron hits region R"; that is, $\Omega = \{R, \sim R\}$. Considering the measuring system described above, "the strength of the propensity for this system to produce the event R" can be calculated by dividing the number of hits in region R by the total number of hits.

Informally, any number of experimental setups are considered to be *versions* of one another if they have the same set of possible outcomes; that is, if two systems are capable of producing the same measurable events, then those two systems are considered to be versions of one another. A more precise account of a "version" of an experiment is given below. There are, of course, many versions of the two slit experiment, and several versions are discussed throughout this chapter and chapter 3. Consider the three non-trivial versions of the experiment that can be performed with a source that emits electrons. In the first version, called system E_1, the experiment is run with slit one open and slit two closed. In the second version, called system E_2, only slit two is open, and in the third version, called system E_3, both slits are open. The version in which both slits are closed is not considered here.

A complete description of the systems E_1, E_2, and E_3 would include a consideration of the events concerned with the status of each of the factors that influence the results of the experiment. If these three systems are exactly the same with respect to those factors other than the status of the slits, then, at some time t_i (before the experiments are to start), the only difference between the three systems is the occurrence or nonoccurrence of the events O_1 and O_2, where O_1 = "slit one is open at time t_i" and O_2 = "slit two is open at t_i." Ideally, then, for some residual event X (a conjunction of all the events representing the fixed variables), $E_1 = X\&O_1\&{\sim}O_2$, $E_2 = X\&{\sim}O_1\&O_2$, and $E_3 = X\&O_1\&O_2$.

Also, if experiments are carried out on all three systems at some time t_j

later than t_i, then (pairwise) each system may or may not produce the same outcomes at some later time t_k. But, the set of possible outcomes is the same for all three systems: for all three systems, $\Omega = \{R, \sim R\}$. For any pair of systems, if $\mathbf{Pr}(R{:}E_i) \neq \mathbf{Pr}(R{:}E_j)$, where $i, j \in \{1, 2, 3\}$, then the difference between the those two systems is statistically significant. In other words, systems E_i and E_j are statistically different "versions" of the two slit experiment. Because it is assumed that the only difference between the systems is the occurrence of the events O_1 and O_2, these events are considered to be statistically relevant to the system (or to the description of the system).

Thus, the statistically relevant features of an experiment are described and determined by the relations of three elements: the system, represented as a set of events actually occurring at some time t_i; the start of the experiment, occurring at some time t_j later than t_i; and the possible outcomes of the experiment, namely those events that occur at some time t_k later than t_j (if at all). Recall that a propensity function is defined *over* a σ-algebra defined on a set of outcomes Ω and *for* a system S. Furthermore, the members of Ω are considered as measurable events that are capable of being produced by the system S. Given these points, by defining a propensity function at the time that the experiment starts t_j, the propensity function provides a "dispositional" relationship between the system and the outcomes. According to one description of the account of the propensity interpretation, the semantics provides an analysis of experiments in terms of a system S, a "start" at some time t_i, and a set of possible outcomes Ω, such that the propensity function (defined at some time t_i, for some system S, and on some set of outcomes Ω) provides a "synthesis" of that experiment.

A complete notation is given as '$\mathbf{Pr}_{tj}(A_{tk}{:}S_{ti})$' and is read as "the propensity (at time t_j) for the system S_{ti} to produce the event A_{tk}," where t_i is earlier than t_j which is earlier than t_k. Unless the temporal properties of events are at issue, temporal indices are often suppressed. The dispositional nature of propensities reflects the fact that propensity values are "forward looking" and asymmetrical. In so far as the propensity function is defined *for a system* (and that system must actually occur by or at some time t_i) and is defined *over a σ-algebra* (and the events in that σ-algebra must occur at some time after t_i), the propensity function represents a relationship that can be described as "forward looking." The primary aim in describing this "forward lookingness" as dispositional is to stress the asymmetry involved in the relationship between the system and the members of the σ-algebra. Chapters 2 and 3 provide a further demonstration that one must be cautious in extending the dispositional nature of propensities to include or describe more robust causal relations.

1.3 Propensities and chance

One of the most important merits of the propensity interpretation of probability is the ease with which it provides an interpretation of chance. This section continues, and expands upon, the discussion of the two slit experiment in order to provide an outline of chance in terms of the propensity interpretation of probability. The analysis begins with a consideration of chance in a situation and extends this treatment to consider chance itself. Ultimately, it is demonstrated that chance can be represented as a selection function on a partition of statistically relevant systems, factors, or variables. Finally, this section provides a formal definition of a version of an experiment.

Two of the most important similarities between propensities and chances are already evident. First, as discussed above, propensities are "forward looking." Both chance and propensity apply only to future events; there is no propensity or chance concerning the occurrence of past events.[1] Second, in principle, propensity values represent objective physical tendencies and can be applied to single-case phenomena. Just as there are two similarities that are quite evident at this point, there is also one major difference between chance and propensity. On the surface, at least, chance is a function on events while propensity is a function on events and systems. The key to developing a theory of chance in terms of propensity is the relationship between events and systems. The basis for the development of the notion of chance depends on the explication of "chance" in terms of "chance in a situation,"[2] and the reduction of "chance in a situation" to "propensity for a system."

Suppose that some version of the two slit experiment is to be performed; that is, there is a situation in which either E_1, E_2, or E_3 is to be used. Recall that

---

[1]Some commentators have claimed that the inversion theorems of the probability calculus demand that propensities be defined over "past" events. This criticism is discussed in chapter 2.

[2]This explication proceeds in a manner that is similar to Brian Skyrms 1984 and 1988.

E_1 denotes the version of the experiment that is run with slit one open and slit two closed, E_2 denotes a version that is run with only slit two open, and E_3 denotes a version that is run with both slits open. In a situation in which system E_1 is to be used, the "chance" of R occurring is based upon the propensity for E_1 to produce R; that is, **Ch**(R) = **Pr**(R:E_1). Similarly, in a situation where system E_2 or E_3 is used, **Ch**(R) = **Pr**(R:E_2) and **Pr**(R:E_3) respectively. The propensity interpretation of probability gives a method of determining the value of *chance in a particular situation*: **Ch**(A) = p if and only if, for some system S describing that particular situation, A is a possible outcome of S and **Pr**(A:S) = p. Thus, the value of **Ch**(A) can be determined in any particular situation provided that the situation is one in which an appropriate propensity function (and system) can be properly defined.

The method of extending this interpretation of "chance in a particular situation" to "chance" is based on generalizing the situations for which chance is defined. Recall that systems E_1, E_2, and E_3 deno'e many events, each fixing the value of a number of factors capable of influencing the ability of the system to produce the event R. Consider a new situation: suppose that the status of the slits is the only factor that influences the results. That is, only the events O_1 and O_2 are statistically relevant to the descriptions of the systems capable of producing the event R. In other words, the events described by the residual state X are irrelevant to the descriptions of $E_{1'}$, $E_{2'}$, and $E_{3'}$. In this situation, $E_{1'} = O_1 \& {\sim}O_2$, $E_{2'} = {\sim}O_1 \& O_2$, and $E_{3'} = O_1 \& O_2$.

The systems constituting this situation are not the same as the originally defined systems. In the original situation, the systems E_1, E_2, and E_3 are *not* *exhaustive* of the systems producing the outcome R. In the present situation, however, the new systems form a *partition* of the systems producing R; that is, $E_{1'}$, $E_{2'}$, and $E_{3'}$ are mutually exclusive and exhaustive. In this situation, **Ch**(R) is not interpreted as a probability function on the possible outcomes and on a particular version of the experiment ($E_{1'}$ for example). Instead, **Ch**(R) is interpreted as a function on the possible outcomes and on a partition $\pi = \{E_{1'}, E_{2'}, E_{3'}\}$. In this situation, **Ch**(R) is defined as follows: **Ch**(R) = **Pr**(R‖π), where **Pr**(R‖π) = **Pr**(R:$E_{1'}$) if it is true that system $E_{1'}$ is at work; **Pr**(R‖π) = **Pr**(R:$E_{2'}$) if it is true that system $E_{2'}$ is at work; and **Pr**(R‖π) = **Pr**(R:$E_{3'}$) if it is true that system $E_{3'}$ is at work. The notation '**Pr**(R‖π)' can be read as "the propensity of R *on* π." Under this interpretation, **Ch**(R) can be viewed as a *selection function*. For any situation ω, **Ch**(R) selects some system $E_{i'} \in \pi$ (where i = 1, 2, 3) that describes the statistically relevant aspects of ω and assigns the value of **Pr**(R:$E_{i'}$) to R.

For the general case, define a partition $\pi = \{S_1, S_2, S_3, \ldots, S_n\}$, where each $S_i \in \pi$ is a system that is capable of producing some member of a set of outcomes Ω. Then, in a situation ω and for the event A in a Boolean σ-algebra defined on some Ω, the function **Ch**(A) in ω is defined as follows: **Ch**(A) = **Pr**(A‖π), where **Pr**(A‖π) = **Pr**(A:S_i) and the system S_i is the member of π containing ω. This is still a definition of chance in a highly restricted type of situation. Specifically, it is a definition of **Ch**(A) in a situation where all the systems capable of producing the event A form a partition.

Returning to the original experiment, although systems E_1, E_2, and E_3 are exclusive, they are not exhaustive of the systems capable of producing the event R. Ignoring the trivial case where both slits are closed, systems E_1, E_2, and E_3 exhaust only those versions of the two slit experiment in which all the factors other than the status of the slits are held fixed. There are many other versions of the two slit experiment based on altering the values of the other variables.

How, then, is one to interpret Ch(R) in the original situation? The answer is obtained by extending the previous analysis as follows. Suppose it is determined that the results of the original two slit experiment are affected by the values of some fixed number of variables, and each variable contains an arbitrary number of possible values. Let the number of variables be 'n' and the number of values for any variable be 'm.' For each variable, define a partition $\pi_i = \{F_{i1}, F_{i2}, F_{i3}, \ldots, F_{ij} \ldots, F_{im}\}$ where each F_{ij} (for $1<i<n$ and $1<j<m$) is the event "variable 'i' has the value 'j' at some time t." If the variables are exhaustive, then the family of partitions, $\Pi = \{\pi_1, \pi_2, \pi_3, \ldots, \pi_n\}$, contains every variable whose values are statistically relevant to the occurrence of the event R. Π is a family of variables such that each variable in Π is statistically relevant to the occurrence of R. Alternatively, the set of all "system-events" that are statistically relevant to the occurrence of R can be easily defined as the set $\varepsilon = \bigcup_{i=1}^{m} \pi_i$, where and $\pi_i \in \Pi$.

Concerning "complete" systems, any two slit system S that is statistically relevant to the occurrence of R can be defined as a conjunction of events: $S = F_{1w} \& F_{2x} \& \ldots \& F_{nz}$ where $F_{ij} \in \pi_i \in \Pi$, such that 'i' identifies the event with some variable and 'j' identifies the event with some value of that variable. Furthermore, the set of all two slit systems that are statistically relevant to the occurrence of R can be obtained by forming the set of all n-length conjunctions where each conjunct is a member of one and only one partition in Π. Call this set of all systems Σ. That is, Σ is a partition of all statistically relevant systems, much like the partition π discussed in the previous situation. Note that it is not necessarily the case that all members of Σ are capable of producing R. For example, a system in which the distance between the slits is relatively large while both the distance of the source from the barrier and the angle of emission are quite small may be incapable of producing an emission that passes through the barrier. Hence, this system is incapable of producing an emission that hits region R. Consequently, the analysis of chance in a situation has been generalized from situations in which only systems capable of producing the event R are considered to situations in which there may be (statistically relevant) systems that are incapable of producing emissions that hit region R.

The result of this analysis of the two slit experiment is that there are three manners of representing the statistically relevant events in the "history" of the event R. The family of partitions Π represents the events as statistically relevant values of variables; Π is the family of statistically relevant variables, each represented by some π_i. The indexed union ε represents a partition of the statistically relevant events themselves. The partition of systems, Σ, combines the events to form a partition of the statistically relevant and distinct systems. Those systems that are "capable of producing R" are a subset of Σ.

From this analysis of variables, events, and systems, **Ch**(R) for the two slit experiment is defined as follows: in a situation ω, **Ch**(R) = **Pr**(R‖Π), such that **Pr**(R‖Π) = **Pr**(R:S), where $S = F_{1w}\&F_{2x}\& \ldots \&F_{nz}$ for appropriate w, x, . . ., and z, where $F_{ij} \in \pi_i \in \Pi$ and S describes the statistically relevant features of ω. One interpretation of **Ch**(R) is as a "two step" function. In the first step, **Ch**(R) selects a value for each variable π_i in Π. This produces a complete description of the statistically relevant features of some system S in such a manner that S describes the situation ω. In the second step, **Ch**(R) assigns a number to R according to the propensity for ω to produce R; that is, **Ch**(R) = **Pr**(R:S), where S is determined in the first step. More simply, **Ch**(R) can be described as a selection function on the partition of all systems Σ that are statistically relevant to R: **Ch**(R) = **Pr**(R‖Σ). In any situation ω, **Ch**(R) selects a system S ∈ Σ such that S describes the statistically relevant aspects of ω to R, and **Ch**(R) assigns the same value to R as **Pr**(R:S).

This account of chance as a selection function has two important consequences. First, in situations that have the same (relevant) descriptions, or have the same "propensity" to produce the event R, the chance of R, **Ch**(R), has the same value. Second, this account allows for easily generalized definitions of experimental setups. If it is the case that "the two slit experiment" is the only mechanism relevant to the occurrence of R, then **Ch**(R) = **Pr**(R‖Π), where Π is defined as above. In the case that one introduces more experimental setups that can produce the same event, one can introduce a new partition or add to the partitions already contained in Π. For example, if the case in which both slits are closed is considered to be a system that is statistically relevant to the event R, then the event $\sim O_1\&\sim O_2$ is added to the "status of slits" partition. On the other hand, a "number of slits" partition could be added to accommodate one slit experiments, three slit experiments, four slit experiments, and so on, if these are considered (or found) to be statistically relevant.

As is to be expected, in the final analysis, the general definition of **Ch**(A) in ω presupposes the identification of those events that are statistically relevant to A in that situation. Thus, in general, **Ch**(A) = **Pr**(A‖Π), where the union of the partitions in Π (namely, ε) contains every event that is relevant to A and **Pr**(A‖Π) is defined as above. Then, to determine the value of **Ch**(A) is to select the appropriate event from each partition in Π, and assign a value to **Pr**(A:S), where S is the conjunction of the relevant events that are selected. Alternatively, **Ch**(A) presupposes the identification or selection of the appropriate system that is relevant to A. In this case, **Ch**(A) = **Pr**(A‖Σ), where Σ contains every statistically relevant system, and **Pr**(A‖Σ) = **Pr**(A:S), where S is the selected system. On this account, then, if the job of science is to determine "chances," then science is involved in (1) determining the "appropriateness" of certain systems as descriptions of certain situations, and (2) determining the values of propensity functions or the propensity for certain systems to produce certain events.

Given this account of chance, it is possible to properly characterize a "version" of an experimental setup. Reconsider the two slit experiment. The notion of a generalized two slit experiment is the product of "chance-like" thinking: "the" two slit experiment is represented by a partition of statistically different systems; that is, Σ. Consequently, there is a "version" of an experimental setup for every $S \in \Sigma$. The system denoted by E_1, for example, represents a particular version of the two slit experiment: every relevant factor is fixed, and E_1 is described by the conjunction of an event from each $\pi_i \in \Pi$. E_1 may be repeated time and time again, perhaps on different sets of apparatus, and at different times. The propensity function is a function on specific versions of experimental setups such as E_1. Chance is a function on experimental setups in the most general sense: chance is a function on objects represented by *partitions* of particular versions.

As discussed above, on the propensity account, given an event description of a particular system and the fact that an experiment on that system has ended with the occurrence of an outcome A, the event description of the system provides a statistically relevant history of the event A, prior to the start of the experiment. The propensity function **Pr**(A:S) is defined on the event A and a *specific* version of an experimental setup S, such that the function **Pr** relates the occurrence of the event A to the occurrence of the events describing S. The chance function **Ch**(A), as a function on an event A and a *partition* of systems Σ, relates the occurrence of the event A to the occurrence of the events describing each unique "version" of the setup in Σ.

Finally, a brief word needs to be said considering an important presupposition that was imposed in the account of chance. Inherent in the account of chance given above is the assumption that it is possible to identify an exhaustive set of statistically relevant variables and partition them into mutually exclusive and exhaustive values. The satisfaction of this assumption is a physical (applied) matter; it is the job of science and is discussed to some extent in later chapters.

1.4 Propensities and conditional chance

This section extends the analysis of chance to provide an account of conditional chance. As in the account of chance, the account of conditional chance begins with the notion of conditional chance in a particular situation, and propensities defined for systems that entirely describe that particular situation. The account proceeds to generalize the situations for which conditional chance can be defined. Considering the similarity of the development of conditional chance and the development of chance, intermediate stages of the development are skipped and a general account of conditional chance is given immediately following the account for conditional chance in a particular situation. The development of conditional chance is first presented in terms of the mathematical definition of conditional probability as a quotient of absolute probabilities. Then, it is presented in terms of conditional propensities. Finally, this interpretation of chance reveals that chance cannot be a physically meaningful concept in all situations.

Recall the definition of system E_3 above: E_3 is the system described by a particular set of events (including the fact that both slits are open) representing all the events that are statistically relevant to the occurrence of the outcomes R and ~R. That is, the set of outcomes for E_3 is {R, ~R}. Suppose that E_3 is altered by adding two detectors that can determine whether the electron passes through slit one or through slit two. Given that electrons scatter light, one could place an extremely strong light source behind the diaphragm and between the two slits. Consequently, if an electron passes through slit one, then a "flash" would be observed or detected within the vicinity of slit one, and if an electron passes through slit two, then a flash would be observed within the vicinity of slit two. Thus, as shown in figure 1.2, "flash detectors" (appropriately calibrated light sensors or trained observers), placed above slit one and below slit two, could focus on the light, in order to detect flashes.

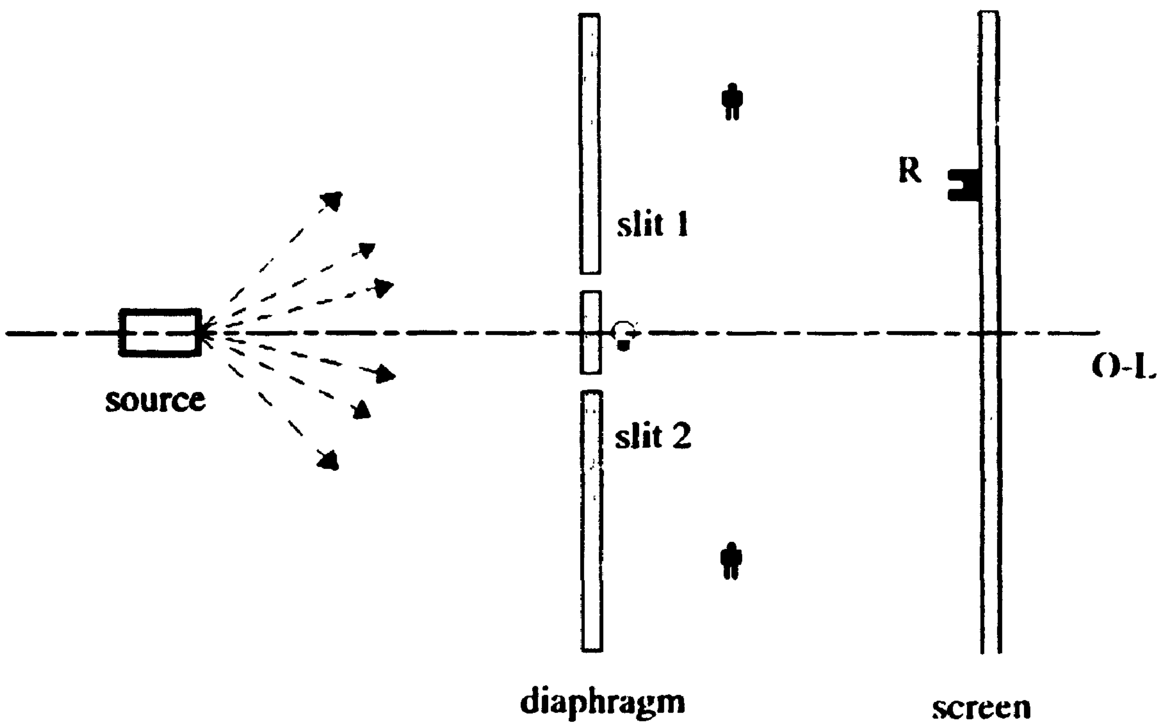

Figure 1.2. The two slit experiment with detectors.

Call this new system E'_3, and assume that the brightness of the light is such that electrons are detected at a given slit when and only when electrons pass through that slit. Consequently, define the events P_1 and P_2 as follows: P_1 = "the electron passed through slit one at some time between t_j (the start of the experiment) and t_k (the time that the electron hits the screen)," and

P_2 = "the electron passed through slit two at some time between t_i and t_k." Note that $\sim P_1 \equiv P_2$ since only those electrons that reach the screen are statistically relevant: electrons that are absorbed by the diaphragm are not counted as statistically relevant. Then the set of (measurable) outcomes for E'_3 is $\{R, \sim R\} \times \{P_1, P_2\}$. By the definition of conditional probability, $\mathbf{Ch}(R|P_1) = \mathbf{Ch}(R\&P_1)/\mathbf{Ch}(R)$. This definition of conditional chance can be extended to a definition of conditional chance in a situation as follows: for some situation ω $[\mathbf{Ch}(R|P_1) \text{ in } \omega] = [\mathbf{Ch}(R\&P_1) \text{ in } \omega]/[\mathbf{Ch}(R) \text{ in } \omega]$. Thus, according to the account of chance above, in a situation where system E'_3 is to be used, $\mathbf{Ch}(R|P_1) = \mathbf{Pr}(R\&P_1{:}E'_3)/\mathbf{Pr}(R{:}E'_3)$. That is, the chance of "R given P_1" is equal to the propensity for system E'_3 to produce $R\&P_1$ weighted by the propensity for system E'_3 to produce R. Alternatively, by the definition of conditional probability, define conditional propensities as follows: $\mathbf{Pr}(R|P_1{:}E'_3) = \mathbf{Pr}(R\&P_1{:}E'_3)/\mathbf{Pr}(R{:}E'_3)$. Conditional propensities are interpreted as follows: $\mathbf{Pr}(R|P_1{:}E'_3)$ = "the propensity for the system E'_3 to produce the event R given that the system E'_3 produces the event P_1." Consequently, the account of conditional chance can proceed more directly in terms of conditional propensities. In a situation where system E'_3 is to be used, $\mathbf{Ch}(R|P_1) = \mathbf{Pr}(R|P_1{:}E'_3)$.

Generalizing these results, the explicata of "$\mathbf{Ch}(A|B)$ in situation ω" includes $\mathbf{Pr}(A\&B\|\Sigma')$ and $\mathbf{Pr}(B\|\Sigma'')$, where Σ' is the partition of systems that are statistically relevant to A&B, and Σ'' is the partition of systems that are statistically relevant to B. In particular, $\mathbf{Ch}(A|B)$ in $\omega = \mathbf{Pr}(A\&B\|\Sigma')/\mathbf{Pr}(B\|\Sigma'')$, where $\mathbf{Pr}(A\&B\|\Sigma')/\mathbf{Pr}(B\|\Sigma'') = \mathbf{Pr}(A\&B{:}S)/\mathbf{Pr}(B{:}S)$ and S, which contains the situation ω, is a member of both Σ′ and Σ″. Alternatively, "**Ch**(A|B) in situation ω" = **Pr**(A|B||Σ), where Σ is the partition of systems that are statistically relevant to both outcomes A&B and B such that **Pr**(A|B||Σ) = **Pr**(A|B:S), where S is the member of Σ that contains the situation ω. That is, conditional chance is the corresponding conditional propensity for the true situation.

Finally, the propensity account of chance and conditional chance raises a very basic concern about the nature of chance: are "chances" a natural and measurable feature of the world? Recall that propensity functions are defined for particular experiments and represent the relations between particular experimental setups and the events that they produce. Once one properly identifies the statistically relevant features of an experimental setup and settles on an interesting set of outcomes, the propensity function can be defined by running repeated experiments while holding the statistically relevant features fixed, and measuring the occurrence or nonoccurrence of the outcome events.

Determining the value of a "chance in a situation" function is quite similar: it is a matter of defining and reproducing "the situation." The task of defining statistically relevant features of systems, or of defining a situation, is no easy matter (it may be impossible), but it is a matter of experimentation. The determination of propensities, and of chances in a situation, is a matter of controlling variables and recording the occurrence (or nonoccurrence) of events. When described in this manner, the determination of propensities is, in fact, the activities of scientists in every field. Chance, however, was represented as a selection function on a family of systems and their propensity functions. The determination of the values associated with **Ch**(A) and **Ch**(A|B) is not a matter of running a single controlled experiment. Instead, it involves the running of a diverse and large number of controlled experiments. The determination of chances cannot (for us) be a totally empirical matter.

Furthermore, given that the value of "**Ch**(A)" is based on the study of systems that are composed of diverse and possibly inhomogeneous causal factors, one is led to question whether this value is *physically* meaningful. For example, the value of **Ch**(John is late for school) could depend on economic, social, psychological, and physical systems. In one situation John is late because he must work at night to afford school; in another situation he is late because his child had to be driven to daycare; in still another he is late because he wanted to upset his instructor; and finally, in another situation, he is late because his car broke down. It is difficult to imagine how all the factors involved in these situations can contribute to a single value representing **Ch**(John is late for school).

There are two qualifying remarks that must be made. First, some may argue that the task of determining the values of propensities, or chances in a situation, is no more likely to be successful than is the task of determining chances. Certainly, especially for complicated systems, this point is well taken. But, the fact remains that the act of determining propensities, or chance in a situation, more greatly resembles the empirical study of causal and statistical systems.

The second point is made evident by the fact that the notion of the two slit experiment was regarded as a result of "chance-like" rather than "propensity-like" thinking. Chance certainly adds to the understanding and analysis of the empirical world in a way that single propensity functions do not. Chance must be regarded as providing a more general analysis of phenomena, and the propensity account captures this fact. There is a manner in which Ch(A) represents a property of A that transcends any particular occurrence of A. It may be said that the relationship between chance and propensity is analogous to the relationship between law and regularity. If chance is to represent the physical world but be something more than propensity, then there is an open question as to whether the notion of chance is or can be well-defined in all physical situations.

1.5 Concluding remarks

The overriding concern of this chapter has been to distinguish between systems and events, and to demonstrate the manner in which this distinction is important to the propensity interpretation of probability. This distinction is the sole basis for designating the interpretation as a "propensity" interpretation of probability. Systems have a "propensity" to produce events; events have no propensity to produce systems. The term 'propensity' is used to indicate the asymmetry of the relationship between systems and events. Furthermore, a causal interpretation of the term propensity must be used with caution, as the asymmetrical relationship is merely "forward looking." It is this feature of propensity functions that enables them to provide the basis for constructing the causal history of events. Chapters 2 and 3 take up the task of clarifying the degree to which the propensity interpretation of probability can be considered a "causal" interpretation of probability.

A propensity function is defined *for* a system and *over* a Boolean σ-algebra on a set of possible outcomes. For brevity, it is often stated that a propensity function is defined *for* a system and *on* an event space. Propensity functions are defined in order to represent the relations between a specific experimental setup and the results of running that experiment (the possible outcomes of that experiment). Consequently, propensity functions are particularly well suited for providing an interpretation of "chance in a situation" and an analysis of "chance" itself. Chance was represented as a selection function on the statistically relevant systems, factors, and variables. Furthermore, the propensity analysis of chance revealed that, in some situations, "chance" is defined over causally inhomogeneous situations. Thus, chance is not a physically meaningful concept when applied to certain situations.

The propensity analysis of chance also revealed that the notion of a single generalized two slit experiment should be thought of as the result of chance-like thinking: "the" two slit experiment is not a particular physical system and cannot be represented by a single propensity function. Propensity functions describe relations between aspects of the same physical experiment. Propensity functions describe only particular "versions" of the two slit experiment.

2

Conditional Propensities and Causation

2.1 Introduction

Chapter 1 provided a semantics for propensities, and demonstrated that conditional chance can be represented in terms of conditional propensities. Chapter 1 did not, however, give a detailed interpretation of conditional propensities. The aim of this chapter is to provide a detailed examination of conditional propensities and the issues surrounding them. The first issue involves the controversy over the interpretation of inverse conditional propensities--conditional propensities in which the conditioned event occurs before the conditioning event. The second issue is the consistency of the dispositional nature of the propensity interpretation and the inversion theorems of the probability calculus, where an inversion theorem is any theorem of probability that makes explicit (or implicit) appeal to an inverse conditional probability. The third issue concerns the relationship between the notion of stochastic independence which is supported by the propensity interpretation, and various notions of causal independence. Finally, this chapter examines the relationship between the propensity interpretation of probability and various theories of probabilistic causation.

Section 2.2 provides some background to the controversy--generally referred to as Humphreys's paradox-- concerning inverse conditional propensities and the inversion theorems of the probability calculus. Section 2.3 demonstrates that the identification of conditional propensities with conditional probabilities does not create problems for the propensity interpretation. By applying the propensity interpretation to an experiment involving a photon's transmission through, or reflection by, a half-silvered mirror, it is demonstrated that inverse conditional propensities can be assigned non-trivial values; that is, there are inverse conditional propensities with values not equal to 1 or 0. Ultimately, it is established that conditional propensities, and their inverses, are formally symmetric in the same manner as conditional probabilities, and their inverses. Furthermore, a procedure for updating propensities over time is developed in order to establish the criterion for a conditional propensity to be well-defined. This criterion is simply that both the conditioned event and the conditioning event occur after the time that the system and propensity function are defined. Building on the fact that inverse conditional propensities can be non-trivial and well-defined, section 2.4 demonstrates that there is no inconsistency between the dispositional nature of the propensity interpretation and the inversion theorems of the probability calculus. The basis for the demonstration that there is no inconsistency is the fact that conditional propensities do not provide a direct measure of the causal dependence of the conditioned event on the conditioning event.

Section 2.5 reinforces the analysis given in section 2.4 by establishing that the propensity interpretation yields a notion of independence which corresponds to stochastic independence, rather than to causal independence.

Furthermore, it is shown that the aforementioned attacks on the propensity interpretation are based on the misconception that the propensity interpretation should, or does, yield a formulation of causal independence in terms of (simple) conditional propensities. The most common form of this misconception is that conditional propensities provide a direct measure of the degree to which the singular conditioning event ''causes'' the singular conditioned event. Finally, it is argued that talk of event-event causation--talk of a singular event causing another singular event--is particularly susceptible to being formulated in terms of conditional propensities, and that these types of formulations involve the same sort of misconception described above.

The last section of this chapter utilizes the propensity interpretation of conditional propensities and chance (developed in chapter 1) to distinguish between, and to evaluate, two approaches to probabilistic causation: the sufficiency view, advocated by Hans Reichenbach ([1956] 1991); and the necessity view, advocated by I. J. Good (1983). By virtue of the propensity analysis of chance, the propensity interpretation is capable of demonstrating the manner in which both approaches to probabilistic causation encounter the problem of homogeneous reference classes. Furthermore, it is argued that any theory of probabilistic causation utilizing criteria expressed in terms of chance will, and must, encounter this problem. Finally, it is argued that the analysis of probabilistic causation in terms of propensities and controlled experiments is capable of minimizing, and possibly avoiding, the problem of homogeneity.

2.2 Humphreys's paradox

The aim of this section is to provide some background on the controversy over the interpretation of inverse conditional propensities and what has come to be called "Humphreys's paradox." Humphreys's paradox is a term used by James H. Fetzer (1981, 283) to describe the basis for one of the most fundamental criticisms of the propensity interpretation of probability. The paradox rests on the apparent inconsistency between the dispositional nature of propensities and the interpretation of "inverse" conditional probabilities. Wesley C. Salmon (1979) provides a particularly stark example of this inconsistency:

> Suppose we are given a set of probabilities from which we can deduce that the probability that a certain person died as a result of being shot through the head is 3/4. It would be strange, under these circumstances, to say that this corpse has a propensity (tendency?) of 3/4 to have had its skull perforated by a bullet. (213-14)

In a more formal characterization of this inconsistency, Paul Humphreys (1985) has presented an argument to show that the identification of conditional propensities with conditional probabilities is inconsistent with Bayes's theorem and other inversion theorems of the probability calculus.

Critics of the propensity interpretation, such as Salmon (1979)[1] and Peter Milne (1986; 1987), have used this apparent inconsistency to argue that, despite the intuitive appeal of using propensities for interpreting conditional

---

[1]See especially Salmon 1979, 213-14. Salmon also presents similar criticisms, in terms of alternative examples, in: 1984, 204-5; 1988, 14, and 1989, 87-89.

probabilities in certain situations, we should abandon the notion that we can develop a complete interpretation of the probability calculus in terms of propensities. Humphreys (1985, 557), on the other hand, contends that the inconsistency is a reason to reject the probability calculus as the correct interpretation of chance.

Despite the growing diversity and number of propensity interpretations of probability, Humphreys's paradox has received little attention, especially from advocates of "physical" propensities. There has, however, been some attention. Fetzer (1981, 283-86) has argued that the paradox can be avoided by considering the calculus of single-case propensities as a nonstandard interpretation of the probability calculus, much like non-Euclidian geometry was considered to be a nonstandard interpretation of Euclidean geometry. On this view, propensities are a part of a family of closely related "probabilistic" systems, and propensities are probabilistic in a manner that is broader than the notion of "probabilistic" as characterized by the "classical" set of probability axioms. Ilkka Niiniluoto (1988, 103, n. 16) alludes to an epistemic solution to an example of the paradox given by Salmon (1984, 205).

The most successful attempt to resolve the paradox is contained in section three of David Miller 1991. Although Miller is concerned with a frequentist interpretation of propensities, his defense against Humphreys's paradox is quite general and strikes at the core of the issues involved. In fact, as a proponent of the propensity view, I find his treatment quite successful. As a general resolution, however, Miller's treatment may be all too brief and may pass over too many issues for most critics of the propensity interpretation. Two of the aims of this chapter are to present a *generalized* resolution of Humphreys's paradox, and to outline the issues surrounding the paradox by demonstrating that the asymmetrical nature of *physical* propensities is not inconsistent with the interpretation of "inverse" probabilities.

2.3 Inverse conditional propensities and symmetry

This section provides a detailed examination of the interpretation of conditional propensities and their inverses. The main issue that is addressed here is raised in an objection to the propensity interpretation, presented by Milne (1986):

> a realist single-case [propensity] interpretation of probability is useful only in an indeterministic universe because otherwise the probabilities are all trivial [either equal to 1 or 0]. In such universes the future is "open" with respect to the present and past. Non-trivial conditional probabilities are only possible when the conditioned event occurs later than the conditioning event. (130-31)

This section demonstrates that, in fact, non-trivial conditional propensities are possible when the conditioned event occurs before (or at the same time as) the conditioning event. In other words, it is demonstrated that conditional propensities, like conditional probabilities, are formally symmetric. The demonstration includes the development of an update semantics for the propensity interpretation of probability.

The discussion of Milne's objection is presented in terms of Humphreys's (1985) example involving the "transmission and reflection of photons from a half-silvered mirror" (561). Following Humphreys (1985, 560-61), the notation reflects the temporal relationships between propensity functions and events: the propensity '$\mathbf{Pr}_{ti}(A_{tj}|B_{tk})$' is generally interpreted as "the propensity at t_i for A to occur at t_j, conditional upon B occurring at t_k." In the notation of chapter 1, the propensity would be represented as '$\mathbf{Pr}_{ti}(A_{tj}{:}B_{tk})$.'

Humphreys's example involves an experimental arrangement that is described as follows:

> A source of spontaneously emitted photons allows the particles to impinge upon the mirror, but the system is so arranged that not all the photons emitted from the source hit the mirror, and it is sufficiently isolated that only the factors explicitly mentioned here are relevant. Let I_{t2} be the event of a photon impinging upon the mirror at time t_2, and let T_{t3} be the event of a photon being transmitted through the mirror at time t_3 later than t_2. Now consider the single-case conditional propensity $\mathbf{Pr}_{t1}(.|.)$ where t_1 is earlier than t_2, and take these assignments of propensity values:
>
> i) $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1}) = p > 0$
>
> ii) $1 > \mathbf{Pr}_{t1}(I_{t2}|B_{t1}) = q > 0$
>
> iii) $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$
>
> where, to avoid concerns about maximal specificity, each propensity is conditioned on a complete set of background conditions B_{t1} which include the fact that a photon was emitted from the source at t_0, which is no later than t_1. (561)

In terms of this example, Milne's objection can be presented as follows: the propensities $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ and $\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1})$ have no realist single-case interpretation; once the event T_{t3} or $\sim T_{t3}$ has been realized, there is no indeterminacy about the occurrence of the event I_{t2}--it has either occurred or not (see Milne 1986, 131). But the propensity function is defined *at a specific time*, and propensities evolve in time along the lines suggested by David Lewis's "A Subjectivist's Guide to Objective Chance" (1980). It is argued that proper attention to this feature of the propensity account is sufficient to answer Milne's objections and to provide a general solution to Humphreys's paradox.

The propensity function $\mathbf{Pr}_{t1}$ is defined *over* a Boolean σ-algebra on the event space. The event space consists of those events, such as I_{t2} and T_{t3}, that may occur after t_1. The propensity function is defined *for* the specific arrangement described above; this is reflected by the conditionalization of all propensities on the set of background conditions B_{t1}. The background conditions typically include statements concerning the occurrence of certain events prior to t_1 such as "a photon was emitted from the source at t_0." Properly speaking, the background conditions, B_{t1}, and any events that are described in the background conditions as occurring before t_1, are not members of the event space or the σ-algebra.

In general, "propensities" are attributed to, or are considered as dispositional properties of, a system that satisfies the background conditions prior to time t_1. The "absolute" propensity $\mathbf{Pr}_{t1}(I_{t2}|B_{t1})$--$\mathbf{Pr}_{t1}(I_{t2}:B_{t1})$ in the notation of chapter 1--represents the propensity at time t_1, for a system that satisfies the background conditions B_{t1}, to produce a photon that impinges upon the mirror at time t_2. Similarly, the propensity $\mathbf{Pr}_{t1}(T_{t3}|B_{t1})$ represents the propensity at t_1, for a system that satisfies conditions B_{t1}, to produce a photon that is transmitted at t_3. Thus, a particular propensity (for a system that satisfies B_{t1} at t_1) is identified as a propensity to produce a particular event at some time after t_1. In this manner, propensities are properties of systems to produce "future" events.

At this point it is important to specify the level of generality 'hat is assumed in the use of the term 'propensity.' Most certainly, no claims concerning realism are made: propensities--whatever they are--may or may not be real. Furthermore, a formal definition of "a propensity" is not presented here. Claims concerning the nature of "propensities" are limited to issues surrounding their role in the interpretation of the probability calculus. To this end, the use of the notion of "a propensity" in the interpretation of the probability calculus is limited to providing the basis for the distinction between--as well as the relationship between--the members of the event space (the possible outcomes) on the one hand, and the background conditions (the events comprising the experimental setup) on the other hand.

As described above, this distinction rests on the notion that the background conditions describe a set of events that have a "propensity," or "disposition," to *produce* the events in the event space. That is, the events described by the background conditions are in some manner responsible for the production of outcomes. The notion of "production"--as opposed to the notions of "propensity," "disposition," or "cause"--is the central notion in the propensity interpretation. In a manner similar to Salmon (1984), the intention here is to indicate that, in making the distinction between the system and the events in the event space, one is only isolating a natural instance of "production" and is not making claims about the causal properties of particular events or systems. The events described by the background conditions are responsible for the assignment of particular probability values to the members of the (previously established) event space, because these events are *responsible for the production of* the events in the event space. As a result, an asymmetrical relationship between the background conditions and the event space is created, and this relationship is not formally represented in other interpretations of the probability calculus.

To return to the interpretation of propensity statements, the ''conditional'' propensity $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1})$--$\mathbf{Pr}_{t1}(T_{t3}|I_{t2}{:}B_{t1})$ in the notation of chapter 1--is interpreted as the propensity at t_1 (for a system satisfying conditions B_{t1}) to produce a photon that is transmitted at t_3 conditional upon its producing a photon that impinges upon the mirror at t_2. On this account, a conditional propensity such as $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1})$ is interpreted as the propensity at t_1 for the system to produce the event T_{t3}, given that the event I_{t2} is also produced. Similarly, the inverse conditional propensity, $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$, is interpreted as the propensity at t_1 for the system to produce the event I_{t2}, given that the event T_{t3} is also produced. The fact that the conditioning event occurs earlier than the conditioned event is inconsequential with respect to whether the inverse conditional propensities are physically meaningful or not. For any conditional propensity, inverses included, both the conditioning event and the conditioned event occur after the time at which the propensity function and the system are defined. The relationship between a conditional propensity and its inverse is symmetrical in the same manner as the relationship between conditional probabilities. This account of the relation between systems and events, and between the conditioned event and the conditioning event, provides a preliminary answer to Milne's objection.

In order to fully understand the manner in which inverse conditional propensities are non-trivial, consider a detailed examination of the particular conditional propensities, and the corresponding inverse conditional propensity, for the photon arrangement. The system described above has a propensity to produce photons that impinge upon the mirror at t_2. This propensity and its value are represented by assignment ii) above: $1 > \mathbf{Pr}_{t1}(I_{t2}|B_{t1}) = q > 0$. For all intents and purposes, this propensity is an "absolute" propensity; it is only conditioned on the fact that the system satisfies certain background conditions. The notation of chapter 1 makes this point clear: $\mathbf{Pr}_{t1}(I_{t2}|B_{t1})$ is represented as $\mathbf{Pr}_{t1}(I_{t2}:B_{t1})$. If this (absolute) propensity is realized, or if the system produces a photon that impinges upon the mirror at t_2, then the system may also produce a photon that is transmitted at t_3. This fact is expressed in terms of the conditional propensity (and its value) represented by assignment i) above: $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1}) = p > 0$. Similarly, depending on the value of p, if the system produces a photon that impinges upon the mirror at t_2, then the system may or may not also produce a photon that is not transmitted at t_3. This result is represented by the fact that, according to assignment i) and the additivity axiom for conditional probabilities (1985, 560), the conditional propensity $\mathbf{Pr}_{t1}(\sim T_{t3}|I_{t2}B_{t1}) = 1 - p$.

The system also possesses an absolute propensity to produce a photon that is transmitted at t_3, and this absolute propensity is formally represented as $\mathbf{Pr}_{t1}(T_{t3}|B_{t1})$. According to Humphreys's additivity axiom and total probability theorem for binary events, $\mathbf{Pr}_{t1}(T_{t3}|B_{t1}) = pq$.[2] If this propensity is realized, or

---

[2]See "A.1 A demonstration that $\mathbf{Pr}_{t1}(T_{t3}|B_{t1}) = pq$" in Appendix 1 for the derivation of this result.

if the system produces a photon that is transmitted at t_3, then the system must also produce a photon that impinges upon the mirror at t_2 since assignment iii) states that $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$. This state of affairs can also be represented by the fact that the inverse conditional propensity $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = 1$. Note that the reason for assigning a value of one to the propensity $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ can be based on mathematical, or abstract, considerations involving the relations between conditional probabilities (or propensities) and their inverses, but it is not necessary to consider these relations.

Instead of utilizing the inversion theorems to determine the value of $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$, the value can be arrived at as follows: the value of $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ must be one since the description of the system indicates that the system is arranged in such a manner that if the system produces a photon that is transmitted at t_3, then the system must also produce a photon that impinges upon the mirror at t_2. Indeed, it is assignment iii) that provides the information that the system is arranged in this manner, but it is the arrangement of the photon system itself--and not the value of $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1})$--that demands that $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = 1$. Thus, the reason for assigning a value of one to the propensity $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ is not the fact that the propensity $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1})$ is assigned a value of zero, *per se*. Further, the value of $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ is not arrived at through abstract, or mathematical, considerations of the relations between conditional propensities and their inverses.

To return to the task of describing the photon transmission example in terms of conditional propensities and their inverses, if the system produces a photon that is transmitted at t_3, then the system cannot also produce a photon that fails to impinge upon the mirror at t_2. This is a fact concerning the physical arrangement and this fact is represented by assignment iii), namely that $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$. Consequently, for reasons similar to those above, and without reference to abstract reasoning about the relations between conditional propensities, $\mathbf{Pr}_{t1}({\sim}I_{t2}|T_{t3}B_{t1}) = 0$, since $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$.

Finally, the system possesses an absolute propensity to produce a photon that is not transmitted at t_3: $\mathbf{Pr}_{t1}({\sim}T_{t3}|B_{t1}) = 1 - pq$, by the additivity axiom and the value of $\mathbf{Pr}_{t1}(T_{t3}|B_{t1})$. If this propensity is realized, or if the system produces a photon that is not transmitted at t_3, then the system may or may not produce a photon that impinges upon the mirror at t_2. This fact is represented by the value of the inverse conditional propensity $\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1})$. By the additivity axiom, the total probability theorem for binary events, and the values assigned above, $\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = (q - pq)/(1 - pq)$.[3] Furthermore, if $p < 1$ then $1 > \mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) > 0$, and if $p = 1$ then $\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = 0$ (recall that $q < 1$). Similarly, if the system produces a photon that is not transmitted at t_3, then the system may or may not produce a photon that impinges upon the mirror at t_2. Thus, by the additivity axiom, if $p = 1$ then $\mathbf{Pr}_{t1}({\sim}I_{t2}|{\sim}T_{t3}B_{t1}) = 1$, and if $p < 1$ then $1 > \mathbf{Pr}_{t1}({\sim}I_{t2}|{\sim}T_{t3}B_{t1}) = 1 - [(q - pq)/(1 - pq)] > 0$.

Before considering the representation of this system over time, it is important to note that the propensity assignments described above made no appeal to the inversion theorems. In some cases, Humphreys's propensity

---

[3]See "A.2 A demonstration that $\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = (q - pq)/(1 - pq)$" in Appendix 1 for the derivation of this result.

assignments i) - iii) were used in order to understand the physical system itself, and from this understanding of the physical states of affairs, the appropriate propensity assignments were made. In other cases, an appeal was made to either the additivity axiom--"If A and B are disjoint, then P(AvB|C) = P(A|C) + P(B|C)" (560)--or the theorem on total probability for binary events (TP)--"P(A|C) = P(A|BC)P(B|C) + P(A|~BC)P(~B|C)" (560)--but neither of these theorems makes an explicit or implicit appeal to inverse conditional propensities.[4]

Most important to the task of answering Milne's objection is the fact that it has been established that if $p < 1$ then $1 > Pr_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) > 0$, and if $p < 1$ then $1 > Pr_{t1}({\sim}I_{t2}|{\sim}T_{t3}B_{t1}) = 1 - [(q - pq)/(1 - pq)] > 0$. Consequently, for a photon transmission system described by Humphreys's assignments i) - iii), where $p < 1$, there are non-trivial conditional propensities defined by a propensity function at time t_1. Thus, in establishing the values of the conditional and inverse conditional propensities above, it has been demonstrated that--contrary to Milne--non-trivial conditional propensities are possible when the conditioned event occurs later than the conditioning event.

Of course, systems change over time as the events that they produce are realized. Consequently, the propensities that systems possess, and the values of

---

[4]Although some proofs of TP make an appeal to inversion theorems, TP can be proven without such appeals. See "A.3 A derivation of the theorem on total probability for binary events, without an appeal to the inversion theorems" in Appendix 1.

those propensities, also change over time. In the photon system, if the event I_{t2} occurs at t_2, then it is possible to update the dispositional nature of that system to reflect the dispositional nature of the "new" system as it exists at t_2. Updating the dispositional nature of that system requires the definition of a new propensity function for the system at t_2; call this new function $\mathbf{Pr}_{t2}$. This function is conditioned on a set of background conditions B_{t2} which consists of the conditions expressed in B_{t1} as well as the additional condition that the event I_{t2} occurred at t_2. Furthermore, this function is defined over a new σ-algebra (defined on a new event space), since the event I_{t2} no longer lies in the future of the system. Only the events T_{t3} and $\sim T_{t3}$, of the original events mentioned, remain in the event space. The assignments made by the new propensity function are defined as follows: $\mathbf{Pr}_{t2}(T_{t3}|B_{t2}) = \mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1}) = p$, and $\mathbf{Pr}_{t2}(\sim T_{t3}|B_{t2}) = \mathbf{Pr}_{t1}(\sim T_{t3}|I_{t2}B_{t1}) = 1 - p$.

If the event $\sim I_{t2}$ (rather than the event I_{t2}) occurs at t_2, then a different propensity function is defined for the system that exists at t_2; call this propensity function $\mathbf{Pr}_{t2'}$. This function is conditioned on a set of background conditions $B_{t2'}$ which consists of the conditions expressed in B_{t1} and the condition that the event $\sim I_{t2}$ occurred at t_2. Then $\mathbf{Pr}_{t2'}(T_{t3}|B_{t2'}) = \mathbf{Pr}_{t1}(T_{t3}|\sim I_{t2}B_{t1}) = 0$ and $\mathbf{Pr}_{t2'}(\sim T_{t3}|B_{t2'}) = \mathbf{Pr}_{t1}(\sim T_{t3}|\sim I_{t2}B_{t1}) = 1$. Properly speaking, the events I_{t2} and $\sim I_{t2}$ are not members of the event space for the propensity functions $\mathbf{Pr}_{t2}$ and $\mathbf{Pr}_{t2'}$ since these events cannot be "produced" by the corresponding systems. Nor do these events lie in the future of the systems described by the background conditions B_{t2} and $B_{t2'}$. Thus, for example, the propensity function $\mathbf{Pr}_{t2}$ is not defined over propensities such as

$\mathbf{Pr}_{t2}(I_{t2}|B_{t2})$, $\mathbf{Pr}_{t2}(I_{t2}|T_{t3}B_{t2})$, and $\mathbf{Pr}_{t2}(T_{t3}|I_{t2}B_{t2})$. Consequently, inverse conditional propensities, such as $\mathbf{Pr}_{t2}(I_{t2}|T_{t3}B_{t2})$, are not well-defined. But this is not because the conditioned event occurs before the conditioning event. $\mathbf{Pr}_{t2}(I_{t2}|T_{t3}B_{t2})$ is not well-defined because the event I_{t2} occurs at or before the time that the system and the probability function were defined.

Thus, as a system evolves over time, events are essentially removed from the event space and incorporated into the background conditions. Depending on which events actually occur, this creates a new propensity function that is defined at a different time, for a different system, and over a different σ-algebra of events. Consequently, inverse conditional propensities are susceptible to a propensity interpretation just in case both the conditioned event and the conditioning event are members of the Boolean σ-algebra defined on the event space. That is, *inverse conditional propensities, defined at some time t_i and defined for some system, are well-defined just in case: (1) both the conditioned event and the conditioning event occur after t_i and (2) the system is capable of producing those events.*

The update semantics is more clearly stated in the notation of chapter 1, without temporal indices. In general, for any system S, corresponding event space Ω, and events A, B $\in \Omega$, where A either occurs or does not occur at t_2 and B either occurs or does not occur at some later time t_4: if $\mathbf{Pr}(B|A{:}S) = p$ at t_1 and A occurs at t_2, then there is a new propensity function $\mathbf{Pr}'$ defined at t_2 on a new space $\Omega' = \Omega\text{-}A$ such that $\mathbf{Pr}'(B{:}S') = \mathbf{Pr}(B|A{:}S)$, and S′ = "S&A" is an event description of the "new" system existing at time t_2.

2.4 Humphreys's argument against the inversion theorems

Humphreys 1985 presents an argument that is intended to demonstrate that the "inversion theorems of the classical probability calculus are inapplicable in a straightforward way to propensities" (563). Humphreys's argument is not simply concerned with temporal asymmetries (as discussed in the previous section); his argument is concerned with causal asymmetries. Consequently, the appeal of the previous section to the temporal evolution of propensities alone is not enough to refute Humphreys's argument. Humphreys's argument against the inversion theorems is presented in the form of a *reductio ad absurdum*, and is based on an inconsistency between assignments i) - iii) cited above, a principle concerning causal independence for the photon example (CI), and the propensity interpretation of the inversion theorems. The aim of this section is to refute Humphreys's argument by questioning both his justification for principle CI and his reasons for using principle CI to represent causal independence as he does. First, principle CI, and Humphreys's justification for it, are examined. Next, it is demonstrated--without an appeal to the inversion theorems--that principle CI does not hold for the system in question. Then it is revealed that there is a flaw in both Humphreys's justification for CI and the claim that CI gives a proper statement of causal independence. Finally, the implications of this analysis of Humphreys's argument against the propensity interpretation is discussed.

Humphreys defines principle CI as follows (1985, 561):

$$\text{(CI)}\quad \mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}|B_{t1})$$

Note that there are two questions that arise. First, does principle CI, in fact, hold for the photon arrangement? Second, how is one to interpret the fact that CI holds (or does not hold) for the arrangement? Humphreys's answers to these questions, and his justification for them, are considered first. Humphreys claims that CI does hold for the arrangement, and he reveals that the intended interpretation of CI is that "the propensity for a particle to impinge upon the mirror is unaffected by whether the particle is transmitted or not" (1985, 561). Humphreys considers two alterations to the arrangement described above in order to justify this interpretation of principle CI (1985, 563). According to the first alteration, the mirror is rendered opaque in order to prevent impinging photons from being transmitted. If this is the only alteration, the propensity for a photon to impinge would remain the same. Call the background conditions for this arrangement B'_{t1} and the corresponding propensity function $\mathbf{Pr}'_{t1}$. Then the result of this alteration is that $\mathbf{Pr}'_{t1}(\sim T_{t3}|I_{t2}B'_{t1}) = 1$ and $\mathbf{Pr}'_{t1}(I_{t2}|B'_{t1}) = q$. According to the second alteration, the mirror is rendered transparent in order that (ideally) all impinging photons are also transmitted. Again, the propensity for a photon to impinge would remain the same. Call the background conditions for this set of alterations B''_{t1} and the corresponding propensity function $\mathbf{Pr}''_{t1}$. The results of these alterations are represented by the propensity statements $\mathbf{Pr}''_{t1}(T_{t3}|I_{t2}B''_{t1}) = 1$ and $\mathbf{Pr}''_{t1}(I_{t2}|B''_{t1}) = q$.

The discussion of these alterations demonstrates that the degree to which the mirror is silvered is the only causal factor responsible for the events T_{t3} and $\sim T_{t3}$ *once the photon has impinged upon the mirror at t_2.* Furthermore, the degree of silvering can be isolated as having its effect between the times t_2 and t_3. Thus, it can be assumed that the degree of silvering is directly responsible for the values of $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1})$ and $\mathbf{Pr}_{t1}(\sim T_{t3}|I_{t2}B_{t1})$, and is in no way responsible for the value of $\mathbf{Pr}_{t1}(I_{t2}|B_{t1})$. The value of $\mathbf{Pr}_{t1}(I_{t2}|B_{t1})$ depends on those factors that are responsible for the impinging of a photon on the mirror at t_2. These factors have their effects between t_1 and t_2; for example, these factors determine the specific momentum of the photons as they are emitted from the source.

Humphreys regards these results as crucial to the interpretation of CI. Of these results, Humphreys states on page 563: "given these facts, the events T_{t3} and $\sim T_{t3}$ are irrelevant to the propensity for I_{t2}, and they can be omitted from the factors upon which the propensity is conditioned without altering its value." In other words, the equalities expressed in principle CI-- $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}|\sim T_{t3}B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}|B_{t1})$--are obtained by omitting the "irrelevant" events T_{t3} and $\sim T_{t3}$ from the propensities represented by $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ and $\mathbf{Pr}_{t1}(I_{t2}|\sim T_{t3}B_{t1})$. But does the consideration of the altered arrangements together with this method of "omission" provide an adequate justification for the claim that principle CI is true of the photon arrangement? Furthermore, does principle CI provide an appropriate formulation of the manner in which event I_{t2} is "unaffected" by events T_{t3} and $\sim T_{t3}$?

As it is demonstrated below, despite the fact that the *singular* events T_{t3} and $\sim T_{t3}$ are causally irrelevant to the event I_{t2}, principle CI does not, in fact, hold for the original arrangement. The examination of the altered arrangements reveals that, in all three cases, the values of $\mathbf{Pr}_{t1}(I_{t2}|B_{t1})$ do not depend on the values of $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1})$ and $\mathbf{Pr}_{t1}(\sim T_{t3}|I_{t2}B_{t1})$. But nothing has been demonstrated concerning the values of $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1})$ and

$\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1})$--the very propensities from which Humphreys is "omitting" events. In section 2.3, however, it was determined through an examination of the physical characteristics of the system alone--*without any appeal to the probability calculus*--that $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = 1$. Yet, CI states that $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}|B_{t1})$, and assignment ii) states that $1 > \mathbf{Pr}_{t1}(I_{t2}|B_{t1}) = q > 0$. Thus, principle CI fails for the photon transmission arrangement, and the failure of CI does not depend on the inversion theorems--or any theorems--of the probability calculus.

In fact, both equalities in CI fail. Using the total probability theorem for binary events and the additivity axiom, one can demonstrate that $\mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = (p - pq)/(1 - pq) < q$.[5] Consequently, the following strict inequality holds for the arrangement described by Humphreys (1985): $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) > \mathbf{Pr}_{t1}(I_{t2}|B_{t1}) > \mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}B_{t1})$. It is important to note that the demonstration of these values in section 2.3 made no appeal to the standard inversion theorems, such as Bayes's theorem for binary events--"$P(B|AC) = P(A|BC)P(B|C)/[P(A|BC)P(B|C) + P(A|{\sim}BC)P({\sim}B|C)]$" (1985, 560)--or the multiplication principle--"$P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C) = P(BA|C)$" (559). The demonstration of these values only appealed to Humphreys's own probability assignments i), ii), and iii) (561), the additivity axiom--"If A and B are disjoint, then $P(A \vee B|C) = P(A|C) + P(B|C)$" (560)--and

[5]Recall that, according to Humphreys's assignments i) and ii), $0 < p \leq 1$ and $0 < q < 1$. Suppose that $(q - pq)/(1 - pq) > q$, then $(q - pq) > (q - pq^2)$. Thus, $pq < pq^2$, and this implies that $q > 1$. This result contradicts assignment ii). Therefore, $(q - pq)/(1 - pq) < q$ by *reductio ad absurdum*.

the theorem on total probability for binary events--"$P(A|C) = P(A|BC)P(B|C) + P(A|{\sim}BC)P({\sim}B|C)$" (560).

Considering the failure of CI for the photon arrangement, what is the source of the error in Humphreys's appeal to the alterations of the original arrangement and subsequent "omission" of the events T_{t3} and $\sim T_{t3}$ in formulating and justifying CI? The examination of the altered arrangements does *not* demonstrate that the degree of silvering is the *only* causal factor responsible for the events T_{t3} and $\sim T_{t3}$ *between the times of t_1 and t_3*. In the original arrangement, a photon must impinge upon the mirror at t_2 in order for it to be transmitted at t_3. Consequently, the causal factors responsible for the propensity for the arrangement to produce a photon that is transmitted at t_3 consists of *both* the degree of silvering of the mirror *and* those factors responsible for producing photons that impinge upon the mirror at t_2.

The propensities represented in principle CI are conditioned on the events T_{t3} and $\sim T_{t3}$. The propensity values for these events depend on causal factors that have their effect after t_0 and before t_3. These causal factors include those factors responsible for the momentum of the photons (after t_0 but before t_2) which are also responsible for the event I_{t2}. CI fails for this arrangement because of the manner in which the events I_{t2}, T_{t3}, and $\sim T_{t3}$ have common causal factors that are effective *after t_0 and before t_2*. Specifically, the causal histories of these events share certain causal factors that determine whether or not a photon impinges upon the mirror at t_2.

The alterations described by Humphreys indicate that the value of $\mathbf{Pr}_{t1}(I_{t2}|B_{t1})$ does not depend on the values of $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1})$ and

$Pr_{t1}(\sim T_{t3}|I_{t2}B_{t1})$. This fact reveals that the event I_{t2} does not depend on those causal factors that are responsible for the events T_{t3} and $\sim T_{t3}$ and that are effective between t_2 and t_3. The fact remains that, although the events I_{t2}, T_{t3}, and $\sim T_{t3}$ lack common causal factors between the times t_2 and t_3, the events I_{t2}, T_{t3}, and $\sim T_{t3}$ share common causal factors that are effective between t_0 and t_2. Specifically, the photon transmission arrangement itself (described by B_{t1}) provides a host of common causal factors. This fact is responsible for the failure of principle CI: if the system produces event T_{t3}, then it must have exhibited certain causal factors, some of which have an influence on the event I_{t2}. Given that the propensity function is defined *for* a system, these influences are taken into account by the assignment of propensity values. Humphreys's method of justifying CI on the basis of the relationships between singular events I_{t2}, T_{t3}, and $\sim T_{t3}$ does not take these influences into account. For the same reasons, principle CI fails to provide an adequate formalization of the fact that the singular event I_{t2} is unaffected by the singular events T_{t3} and $\sim T_{t3}$. Alternatively, the reason that principle CI fails for the photon arrangement is that it does not provide a statement concerning the causal independence of singular events. Consequently, Humphreys's justification for CI fails because his justification for CI is based on the assumption that CI provides a statement of the causal independence of the singular events involved.

To conclude this section, consider Humphreys's argument against the propensity interpretation of the probability calculus. Humphreys argues that there is a fundamental inconsistency between the propensity interpretation of the probability calculus (especially the inversion theorems) and the fact that principle CI holds for the photon arrangement, and CI is intended to make a claim concerning causal dependence (of singular events) in terms of propensities. From this inconsistency, Humphreys argues that "[the] inversion theorems of the classical probability calculus are inapplicable in a straightforward way to propensities" (563). Given the demonstration that these inconsistencies arise without the inversion theorems, one is left with an inconsistency between the propensity interpretation of probability (without the inversion theorems), and the claim that principle CI holds for the photon arrangement. That is, the source of the inconsistency is not the propensity interpretation of the inversion theorems.

At this point, then, one is forced to either reject the restricted propensity interpretation of probability (without the inversion theorems) or reject the claim that principle CI holds for the photon arrangement. There are two arguments for rejecting the latter. First, based on the physical description of the arrangement--specifically assignment iii) which states that $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$--it follows, without the use of any theorems of probability, that $\mathbf{Pr}_{t1}(I_{t2}|T_{t3}B_{t1}) = 1$. Consequently, principle CI in fact fails for the arrangement. Second, Humphreys's argument for principle CI, based on the fact that the singular events T_{t3} and $\sim T_{t3}$ are causally irrelevant to the singular event I_{t2}, is not valid. As discussed above, the method of omission that is employed in this argument is not justified. The interpretation of principle CI as a statement of the causal independence of singular events is based on a misunderstanding of the photon arrangement and the propensity interpretation of probability.

Thus, the causal interpretation of principle CI is as much at issue as whether principle CI, in fact, holds or not. Furthermore, the controversy that has generally been referred to as ''Humphreys's paradox'' rests on a disagreement over the causal interpretation of propensity statements and not necessarily on a disagreement over the propensity interpretation of probability. Under the propensity interpretation of probability, outlined above, the propensity statements comprising principle CI do not provide a formulation of causal dependence between *singular* events. Conditional propensities do not provide a measure of causal dependence between the (singular) conditioned event and the (singular) conditioning event.

If one maintains that conditional propensities do provide a measure of causal dependence between the (singular) conditioned event and the (singular) conditioning event, then one is no longer dealing with a propensity *interpretation* of probability, one is dealing with a *theory* of propensities. Furthermore, as both Humphreys's discussion and the present discussion point out, no obvious motivation has been provided for a *theory* of propensities to adopt the inversion theorems of the probability calculus. The question then arises: is there a motivation for an interpretation of the probability calculus to adopt talk of propensities? The next section examines both the relationship between conditional propensities and causal dependence, as well as the possible motivations for an interpretation of probability to adopt talk of propensities.

2.5 Principle CI and causal dependence

This section examines the significance of the failure of principle CI towards the propensity interpretation of probability. First, it is argued that, under the propensity interpretation of probability, principle CI must be taken to be a statement of statistical independence rather than a statement of causal independence. Consequently, the failure of principle CI cannot be taken to be proof that the singular events T_{t3} and $\sim T_{t3}$ are "causally relevant" to the singular event I_{t2}. Finally, this section addresses certain ambiguities surrounding "common" propensity statements of the form "the propensity of X to produce Y is p," and examines the relationship between physical propensities and epistemic probabilities.

Given the failure of CI, rather than concluding that the singular events T_{t3} and $\sim T_{t3}$ are causally relevant to the singular event I_{t2}, first consider whether principle CI should be taken as a proper statement of the causal independence of the events I_{t2}, T_{t3}, and $\sim T_{t3}$. Current theories of probabilistic causation provide evidence that (simple) direct conditional propensities do not provide an adequate characterization of causal independence. For example, conditional propensities cannot be equated with a function like I. J. Good's '$Q(E{:}F)$' or '$\chi(E{:}F)$,' "the tendency of F to cause E" and "the degree to which F caused E" respectively (see Good 1983). As Good demonstrates, the explicatum for these functions is more complex than the explicatum for a direct conditional propensity (according to the classical definition of conditional probabilities). Conditional propensities do not represent the degree to which the (singular) conditioning event causes the (singular) conditioned event. Consequently, given the interpretation of principle CI that is outlined above, and assuming that indications from current theories of probabilistic causation like Good's are correct, principle CI cannot be interpreted as a complete statement of *causal* independence of the singular events I_{t2}, T_{t3}, and $\sim T_{t3}$.

As long as the definition of conditional propensities is taken to correspond to the definition of conditional probability, CI must correspond to a statement of statistical independence rather than to a statement of causal independence. Of course it is possible to reject the classical definition of probability as the proper definition of conditional propensities. In fact, this is what is done by advocates of propensity *theories*. But, the motivation of propensity theories is only to construct a calculus that is capable of representing the causal influences between events, and not to preserve the structure of the probability calculus. On the other hand, some advocates of propensity *interpretations* of probability also reject the classical definition, but claim to retain some "probabilistic" status--Fetzer (1981) is an example. The success of these alternative interpretations depends on the proposed definition. Furthermore, for this type of propensity *interpretation*, "success" involves both the degree to which the new calculus is "probabilistic" and the degree to which it can capture the notion of causal independence.

I see no reason to reject the classical definition, as long as one is willing to accept the stochastic, rather than causal, notion of independence that results. The motivation for retaining the classical definition is to have a stronger, richer, and more useful interpretation of the probability calculus, and not to create a causal calculus. Ultimately, the motivation for this "enriched" interpretation of the probability calculus is to provide the basis for the representation of experiments, the interpretation of chance (as demonstrated in chapter 1), and the construction of a causal calculus. It has already been demonstrated that the dispositional nature of the propensity interpretation is consistent with the axioms and theorems of the probability calculus, and it is demonstrated below that the propensity interpretation is particularly well suited to expressing various criteria for identifying causal relationships. By rejecting CI, the propensity interpretation is not rejecting the notion that the singular events T_{t3} and $\sim T_{t3}$ are "causally irrelevant" to the singular event I_{t2}. What is being rejected is the notion that CI provides an adequate formulation of the causal independence of singular events in terms of the propensity interpretation of the probability calc lus.

One issue arising in the previous discussion deserves further consideration: if direct conditional propensities do not provide an adequate representation of causal dependence, then what is the benefit of using the propensity interpretation in the study of causation? The remainder of this section provides a partial answer to this question by addressing certain ambiguities surrounding common appeals to propensity statements, particularly, statements of the form "the propensity of X to produce Y is p." It must be recognized that (generally) propensity statements of this form are not represented by conditional propensities. Instead, they are represented by "absolute" conditional propensity statements of the form "the propensity of the *system* X to produce the *event* Y is p."

The example that is considered here is that of a "can opener factory" as presented by Salmon (1984, 205).[6]

> Consider, for example, a factory that produces can openers. There are only two machines, which we may designate A and B, in this factory. Machine A is ancient; it produces one thousand can openers per day, and 2.5% of these are defective. Machine B is more modern; it produces ten thousand can openers per day, and only 1% of its products are defective. Suppose, at the end of the day, that all of the defective can openers (which have been sorted out by the inspectors) are placed in a box. Someone randomly picks a can opener out of the box, and asks for the probability that it was produced by the modern machine B. We can easily calculate the answer; it is 4/5.

Let A_{t2} be the event that a can opener is produced by machine A at t_2, let B_{t2} be the event that a can opener is produced by machine B at t_2, and let D_{t3} be the event that a can opener is defective at t_3. Also, define F_{t1} as the background conditions concerning the running of the can opener factory for a day once work starts at that factory at some time t_0 earlier than t_1. Given these definitions, Salmon has indicated that $\mathbf{Pr}_{t1}(A_{t2}{:}F_{t1}) = 1/11$, $\mathbf{Pr}_{t1}(D_{t3}|A_{t2}{:}F_{t1}) = 1/40$, $\mathbf{Pr}_{t1}(B_{t2}{:}F_{t1}) = 10/11$, and $\mathbf{Pr}_{t1}(D_{t3}|B_{t2}{:}F_{t1}) = 1/100$. From these propensities and Bayes's theorem, it follows that $\mathbf{Pr}_{t1}(B_{t2}|D_{t3}{:}F_{t1}) = 4/5$.

As Salmon continues, he expresses the following misgivings:

> Nevertheless, I find it quite unacceptable to say that this defective can opener has a propensity of 0.8 to have been produced by machine B. It makes good sense to say that machine B has a propensity of 0.01 to produce defective can openers, but not to say of the can opener that it has a certain propensity to have been produced by that machine. (1984, 205)

First, consider the statement that makes good sense: "machine B has a

---

[6]Salmon has presented a number of examples, most of which have a structure that is similar to this example. See note 1 for citations.

propensity of 0.01 to produce defective can openers." Considering the fact that "propensities" are attributes of systems to produce certain events, this statement is represented as follows. Let M_{ti} be a set of background conditions describing the operation of machine B at some arbitrary time t_i; then there is a propensity function defined for the system "machine B" at any time t_i such that $\mathbf{Pr}_{ti}(D_{tj}:M_{ti}) = 1/100$ where the time t_j is later than t_i. Thus, the propensity at t_i for (the system) machine B *simpliciter* to produce a defective can opener at t_j is 1/100.

The reason that the statement "the defective can opener has a certain propensity to have been produced by machine B" is unacceptable is now clear. A defective can opener *simpliciter* (at t_3) has no causal power (in and of itself) to "produce" the property of being made by machine B (at t_2). That is, there is no system described by some set of background conditions D_{ti} and some corresponding propensity function $\mathbf{Pr}_{ti}$ for which $\mathbf{Pr}_{ti}(B_{tj}:D_{ti}) = 3/4$, where the time t_i occurs later than the time t_j.

But this result is paralleled in the case of the propensity function defined for the factory F. Suppose that the system produces a defective can opener at t_3, and the propensity function $\mathbf{Pr}_{t1}$ for the factory F is "updated" accordingly. The updated propensity function is $\mathbf{Pr}_{t3}$ and is defined for a system that satisfies the background conditions F_{t3} which include the fact that the factory produced a defective can opener at t_3. Then, as described in section 2.3, the propensity function $\mathbf{Pr}_{t3}$ is not defined over the event B_{t2} since this event does not occur after t_3. Thus, the propensity statement $\mathbf{Pr}_{t3}(B_{t2}:F_{t3})$ is not well-defined.

Finally, recall the exact nature of Salmon's characterization of the situation that raises the question of how one should interpret inverse conditional propensities in the first place.[7] The situation is this: ***at the end of the day*** (at t_3), "someone randomly picks a can opener out of the box, and asks for the probability that it was produced by the modern machine B" and the answer "the probability is 4/5" is given. Does this answer imply that the defective can opener, itself, has a propensity of 4/5 to have been produced by machine B? Of course not. At the end of the day, it is known that at a previous time t_1 the propensity for the factory to produce a can opener that is made by machine B at t_2, given that the factory produced a defective can opener at t_3, is 4/5. This is a physical probability that (still) characterizes the running of the factory as it stood at t_1. Also, at the end of the day it is learned that a particular can opener is defective, and was produced by the factory between the times t_1 and t_3. Thus, ***the value*** of the physical propensity (at t_1) for the factory to produce a can opener made by machine B, given that it produced a defective can opener, is simply ***adopted as the value*** of an epistemic probability that this can opener was made by machine B.

"Inverse" propensities do not represent "inverse" dispositions. An "inverse" propensity represents a "forward" disposition: a disposition that produces one future event given that it also produces another future event. Furthermore, the propensities corresponding to inverse conditional

---

[7]I am grateful to William L. Harper for pointing out the necessity of returning to this issue.

probabilities are not "reduced" from physical probabilities to epistemic probabilities. The values of physical propensities are, however, often adopted as the values of epistemic probabilities. Finally, one must recognize two important points concerning the nature of this answer to Salmon's question. The first point is that the solution to the problem of interpreting the conclusion that "the probability that this defective can opener was produced by machine B is 4/5" is an epistemic matter.[8] The second point is that this solution does not rest on interpreting inverse conditional propensities as being "merely" epistemic probabilities and not physical probabilities.

What is to be said of the corpse mentioned in the example from Salmon (1979)? Does the corpse have a propensity to have had its skull perforated by a bullet? The "corpse" *simpliciter* has no such propensity. But at some time earlier than the time at which the fatal shot killed the victim, the circumstances surrounding the shooting death of the victim had a propensity to produce a bullet that would perforate the victim's skull, given that those circumstances would also cause the death of the victim. Without more information concerning the circumstances that produced the fatal shot, not much more can be said.

2.6 Conditional propensities and probabilistic causation

This section uses the interpretation of chance to relate the propensity interpretation of probability to the study of probabilistic causation and causal

---

[8]This point is recognized by Niiniluoto (1988, 103 note 16).

independence. First, two distinct approaches to probabilistic causation--the sufficiency view and the necessity view--are characterized using the propensity interpretation of chance. Next, examples of each view are examined: Hans Reichenbach ([1956] 1991) as a proponent of the sufficiency view, and I. J. Good (1983) as a proponent of the necessity view. Then, the propensity interpretation is used to demonstrate that both of these approaches to causation incorporate chance-like thinking, consequently both approaches encounter the problem of the homogeneous reference class. Finally, it is argued that any account of causation based on chance-like criteria will encounter the problem of homogeneity, and that it is only through an analysis in terms of "chance in a situation" that the problem of homogeneity can be avoided. Of course, as it was demonstrated in chapter 1, the propensity interpretation is particularly well suited to providing an analysis of chance in a situation. Consequently, it is argued that the propensity interpretation is capable of avoiding the problem of homogeneity.

There is a great deal of debate concerning the nature of probabilistic causation. Consequently, the presentation here is quite general. Following Brian Skyrms (1988), discussion focuses on the distinction between two fundamental approaches to causation. The first approach analyzes an event (a potential cause) in terms of the degree to which it provides a sufficient condition for the occurrence of some other event (the potential effect). The second approach focuses on the "necessity" of an event (a potential cause) in producing the occurrence of some other event (the potential effect).

The basis of the first approach--the sufficiency view--usually involves some statement or formulation of the following criterion for causation:

SPCF C is a (sufficient) positive causal factor for E if and only if **Ch**(E|C) > **Ch**(E|~C)

The propensity interpretation of this criterion for causal relevance is as follows. **Ch**(E|C) > **Ch**(E|~C) if and only if **Pr**(E|C:S) > **Pr**(E|~C:S) for *any* relevant system S capable of producing both event E and event C. **Pr**(E|C:S) > **Pr**(E|~C:S) is true if and only if it is found that, among those systems that are capable of producing both events E and C, the systems that produce event C have a stronger propensity to produce event E than do the systems that produce the event ~C. It is on this basis that C is called a positive causal factor for E. The definition of a negative causal factor can proceed in a similar fashion in terms of **Ch**(E|C) < **Ch**(E|~C). The result is that the identification of causal factors is a task of identifying events that always contribute to the occurrence (or nonoccurrence) of an event: if C is a positive causal factor for E then there are no systems for which **Pr**(E|C:S) ≤ **Pr**(E|~C:S).

One manner of approaching the problem of verifying the fact that there are no such systems is to test for systems containing a common cause for C and E. This is the approach of Hans Reichenbach, one of the most notable proponents of the sufficiency approach to probabilistic causation. Reichenbach ([1956] 1991) defines causal relevance as follows:

> An event A_1 is *causally relevant* to a later event A_3 if **P**(A_1, A_3) > **P**(A_3) and there exists no set of events $A_2^{(1)} \cdots A_2^{(n)}$ which are earlier than or simultaneous with A_1 such that this set screens off A_1 from A_3. (204)

The fact that Reichenbach must supplement his "frequentist" definition of statistical independence with the "screening off" condition in order to define causal relevance can be easily understood in terms of the propensity account of **SPCF**. In terms of propensities, Reichenbach's search for a set of events that screens off event A_1 from A_3 in the analysis of causal relevance is the same as the search for systems that "screen off," or do not exhibit, the contribution of C to the production of E. That is, the screening off condition motivates the search for systems for which **Pr**(E|C:S) = **Pr**(E|~C:S). Reichenbach's notion of screening off is built into the propensity interpretation of **SPCF**. Given that chance can be understood as a selection function on propensities, the satisfaction of **SPCF**, above, demands that there exists no system that screens off the effects of C on E. Thus, the use of chance and propensities demonstrates the strength of Reichenbach's screening off condition.

The "necessity" approach usually involves some statement or formulation of the following criterion for causation:

> **NPCF** C is a (necessary) positive causal factor for E if and only if **Ch**(~C|~E) > **Ch**(~C|E) where the event C either occurs or does not occur at some earlier time than the time that event E either occurs or does not occur.

The propensity interpretation of this criterion for causal relevance is as follows. **Ch**(~C|~E) > **Ch**(~C|E) if and only if **Pr**(~C|~E:S) > **Pr**(~C|E:S) for *any* relevant system S capable of producing both event E and event C. **Pr**(~C|~E:S) > **Pr**(~C|E:S) is true if and only if it is found that, among those systems that are capable of producing both events E and C, the systems that *fail* to produce event E have a stronger propensity for failing to produce event C than do the systems that produce the event E. That is, proportionately fewer systems are capable of producing the event E without also producing the event C. The difference between the values of **Pr**(~C|~E:S) and **Pr**(~C|E:S) can be considered as a measure of the degree to which the production of C is necessary for producing the event E. It is in this manner that C is considered a causal factor for E.

The most notable example of this approach to the analysis of causation is I. J. Good's "A Causal Calculus" (1983). Consider, for example, his Q-function. In abbreviated notation, Q(E:F) is read as the "causai support for E provided by F, or the tendency of F to cause E" where F and E are, respectively, the initial and final events in a causal network (198). The full notation reveals that the subjectivist theory of Good relies on many of the same assumptions as does the propensity account.

Good takes Q(E:F) to be an abbreviation for Q(E:F|U.H) where U and H are defined as follows:

> Let H denote all true laws of nature, whether known or unknown, and let U denote the "essential physical circumstances" just before F started. When we talk about "essential physical circumstances" we imply that the exact state has a probability distribution. (200)

The conditioning of Q on "the 'essential physical circumstances' just before F started" is simply taking account of what "system" is being used. Also, the "implication that the exact state has a probability distribution" amounts to the assumption that there is a well-defined propensity function for that system on some appropriate event space. Consequently, by considering the abbreviated definition of Q to be Q(E:F) = log[P(~E|~F)/P(~E|F)], the Q-function most certainly provides a measure of the difference between the values of Pr(~C|~E:S) and Pr(~C|E:S) as described above.

NPCF presupposes that event C precedes event E. This assumption is tantamount to the assumption that there is no "backwards causation." If backward causal systems are being studied, then an appropriate definition of "necessary positive causal factor" will have to be formulated. The reformulation could be as simple as changing the temporal conditions in **NPCF**. Ultimately, the determination of the necessary changes is an empirical matter. In his statement of causal tendency, Good takes events F and E to be the "initial" and "final" events, respectively, in a causal network. Thus, Good's version of **NPCF** consists of a subjectivist statement of statistical relevance, the physical presuppositions involving "the 'essential physical circumstances' just before F started," and the temporal conditions between F and E. Furthermore, the propensity interpretation of chance reveals that these presuppositions are necessary for capturing the force of **NPCF**.

While the propensity interpretation of chance provides an opportunity to compare and contrast the necessity and sufficiency approaches to probabilistic causation, it also provides an opportunity to evaluate them. Both the necessity approach and the sufficiency approach encounter a fundamental problem: both approaches to causation make reference to probabilities that are defined over negated events, and this reference introduces inhomogeneous causal factors into the analysis. For example, according to the propensity interpretation of

Reichenbach's theory, to search for systems that screen off the event C from E, one must consider systems that produce the event E but fail to produce the event C. Without restrictions on similarity of systems, this search for systems that screen off the event C from E can be interpreted as a search for any system that produces the event E and ~C, with no consideration of its ability to produce the event C. The result is that the systems to be considered can include systems composed of extremely diverse--even inhomogeneous--causal factors. This problem is particularly acute for frequentists like Reichenbach. Given that the frequentist's search for a reference class proceeds in terms of sets of events (rather than systems), the threat of inhomogeneity is more evident since it is possible to devise an infinite number of alternative reference classes in terms of sets of events.[9] Of course, in addition to offering a method for identifying this problem, the propensity account offers a method for avoiding or reducing the extent of the problem.

In stating the criteria in terms of propensities, one can demand that the propensity functions are defined over a σ-algebra on the *same* set of outcomes. This restriction results in the consideration of propensity functions that are defined for the same (or similar) systems, since propensity functions are defined ***both*** for a system ***and*** on an event space. As a result, the propensity functions used in the analysis of causation should be defined for systems that

---

[9]See Salmon 1971 (especially 40-42) for a similar criticism of Reichenbach's frequentist approach to identifying an appropriate reference class.

can produce the events E, ~E, C, and ~C. The assumption is that, in most cases, there will be greater similarity between the causal factors involved in these systems than among the "unrestricted" pool of systems allowed by Reichenbach's frequentist account. Traditionally, this problem has been characterized as the problem of the homogeneous reference class. Of course, even with the propensity interpretation of the sufficiency approach, as long as the criteria for causal relevance are expressed in terms of conditional chance, the problem of homogeneity will remain.

The problem of homogeneity can be avoided by basing the analysis of causation on *chance in a situation*, where that situation is *characterized by a homogeneous set of causal factors*. This can be expressed as an analysis of **Ch**(A‖Σ), where Σ is a set of homogeneous systems; systems that are composed of homogeneous sets of causal factors. In practical terms, the same basic experimental set up (characterized by some set of variables) is used to perform a number of experiments with different settings for the variables. The desired results are produced by controlling variables such that the results of the same general experimental setup run with different settings for the variables can be compared. In chapter 3, examples of such controls and comparisons are discussed in terms of the results of a number of versions of the two slit experiment.

Finally, the necessity approach to causal relevance is confronted with similar problems. Advocates of the necessity approach confront the problem of homogeneity with both the potential cause and the potential effect. In response to the problem of homogeneity, various authors[10] have suggested narrowing the reference class for the potential cause or effect by defining "contrastive" causal relevance as follows (for example): "C in contrast to {C′, C″, C‴, . . .} is a positive causal factor for E" or "C has a positive causal factor for E in contrast to {E′, E″, E‴, . . .}." Chapter 3 demonstrates that this strategy is easily incorporated into the strategy of analyzing causal relevance in terms of controlled experiments.

2.7 Concluding remarks

The dissolution of Humphreys's paradox, and most of the issues surrounding the propensity interpretation of inverse conditional propensities, rests on recognizing that conditional propensities do not represent a measure of the causal dependence of the (singular) conditioned event on the (singular) conditioning event. Inverse conditional propensities do not correspond to, or represent, inverse dispositions. Furthermore, inverse conditional propensities can take non-trivial values and, as the update semantics demonstrated, inverse conditional propensities are well-defined as long as both the conditioned event and the conditioning event occur later than the time at which the propensity function is defined. It was also demonstrated that Humphreys's claim that principle CI--$\mathbf{Pr}_{t1}(I_{t2}|T_{t3}:B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}|{\sim}T_{t3}:B_{t1}) = \mathbf{Pr}_{t1}(I_{t2}:B_{t1})$ in the notation of chapter 1--holds for the photon arrangement is false. On the one

[10]For example, Glymour (1986), Good (1983), Holland (1986), and Skyrms (1988).

hand, by examining the physical characteristics of the photon arrangement (without an appeal to the theorems of the probability calculus) it was demonstrated that principle CI does not hold for the arrangement. On the other hand, it was demonstrated that Humphreys's justification for principle CI was based on a misconception of the causal relations expressed by principle CI. Under the propensity interpretation outlined above, principle CI does not provide a direct statement of causal independence: principle CI provides a statement of stochastic independence.

It was demonstrated that the propensity interpretation of chance is particularly well suited to providing a distinction between, and analysis of, two approaches to probabilistic causation. This analysis showed that any account of causation based on criteria expressed in terms of chance, such as **SPCF** and **NPCF**, is subject to the problem of homogeneity. There is a tension in the study of causation. On the one hand, causation must involve the study of the relations between events in more than one particular system. However, the study of more than one system introduces the problem of homogeneity. Any account of causation based on chance will encounter the problem of homogeneity in full force, since chance is a selection function on *all* systems exhibiting a relation between the events in question. Consequently, in order to avoid, or reduce, the impact of the problem of homogeneity, the class of systems that are examined must be restricted in some manner. The propensity account is particularly well suited to imposing such restrictions.

Finally, it has been demonstrated that the dispositional nature of the propensity interpretation that has been outlined above is consistent with the theorems of the probability calculus. Additionally, the enriched nature of the propensity interpretation, resulting from its dispositional character, is the basis for unique insights into the analysis of chance and causation. Moreover, the propensity interpretation provides a link between the "abstract" nature of chance (discussed in chapter 1) and the problem of homogeneity as it is encountered by both the sufficiency and necessity approaches to causation. Furthermore, this analysis provided a strategy for avoiding the problem of homogeneity.

3

Propensities and Quantum Mechanics

3.1 Introduction

Chapter 1 emphasized the fact that propensity functions are defined for particular systems and represent the relations between a specific experimental setup and its possible outcomes. Chapter 1 also introduced four different "versions" of the two slit experiment: one version, namely E_1, in which slit one is open, slit two is closed, and the detector is off; one version, namely E_2, in which slit one is closed, slit two is open, and the detector is off; one version, namely E_3, in which both slits are open and the detector is off; and one version, namely E'_3, in which both slits are open and the detector is on. Clearly, there are relationships between the experimental setups in the different versions. Basically, the values of two variables are changed: the status of the slits and the status of the detector at the slits. Considering the demand that the propensity interpretation puts on the representation of systems and their possible outcomes, an obvious question arises: "Can the propensity interpretation provide a means of representing these relations between different experiments?"

In fact, the update semantics of chapter 2 provided a means of representing the relations between alternative descriptions of the "same" system as it evolves over time. But in the case of the two slit experiment, shouldn't there also be a rather straightforward relationship between the values of $Pr(R{:}E_1)$--representing trials that produce electrons that pass through slit one and hit region R, with the detector off--and the value of $Pr(R\&P_1{:}E'_3)$--representing trials that produce electrons that pass through slit one and hit region R, with the detector on? Given that chance was represented as a selection function on a partition of propensity functions, one must anticipate the possibility of providing a mathematical representation of the relations between different experimental setups and their results in terms of propensity functions. One of the primary aims of this chapter is to demonstrate that the representation of relations between experiments is an empirical enterprise.

In addition to examining the results of the various two slit experiments performed with electrons (as discussed in chapter 1), this chapter examines the results of two sets of macroscopic two slit experiments: one set is carried out with bullets and the other is carried out with water waves.[1] Each set of experiments is considered in turn, beginning with the bullet experiments. First, the bullet experiments and their results are described in terms of the propensity interpretation. Then the relations between these experiments are expressed in terms of the principle of strict summation and the principle of strict composition. Finally, it is demonstrated, by appealing to the theorems of probability and to various empirical results, that these principles are equivalent.

Section 3.3 examines the exact nature of, and the significance of, the

---

[1]These experiments are based on those described in Richard Feynman 1965 (ch. 6) and in Richard Feynman, Robert B. Leighton, and Matthew Sands 1965 (ch. 1)

equivalence of the principles of strict summation and strict composition. It is demonstrated that the equivalence of these principles is tantamount to making an appeal to controlled experiments. Also, an analysis of controlled experiments establishes that controlled experiments provide the basis for defining the relations between propensity functions that are defined for different systems. Furthermore, the study of controlled experiments provides a characterization of the manner in which the bullet experiments themselves share common causal factors.

Section 3.4 provides an examination of the two slit experiments performed with water waves and of the results of these experiments. The results of the experiments with water waves are presented in terms of intensities rather than propensities. First it is demonstrated that the principle of strict summation is violated. Also, it is established that the violation of this principle is due to the causal influence of the status of the slits. Furthermore, these results yield an interference effect that can be expressed in such a manner that they obviously correspond to a violation of the principle of strict summation, but there is no expression corresponding to the violation of the principle of strict composition. It is argued that, due to the nature of intensities and their measurement, there are limitations on the manner in which the results of wave experiments can be expressed. Thus, questions concerning the violation, or satisfaction, of the principle of strict composition cannot be answered.

The two slit experiments performed with electrons, and their results, are examined in section 3.5. It is demonstrated that the principle of strict summation is violated while the principle of strict composition is not violated. Furthermore, it is demonstrated that these results are consistent with the supposition that both the status of the slits and the status of the detector are causally relevant. It is concluded that the results of the two slit experiment with electrons are paradoxical only if one supposes that there is a causal explanation of the manner in which the status of the detector is causally relevant to the results. From this conclusion it is argued that strictly logical and probabilistic attempts to solve the paradoxes, such as those of Hilary Putnam 1979 and Luigi Accardi 1984, fail. Finally, this chapter utilizes the propensity interpretation of the two slit experiment to formulate two fundamental questions that any interpretation of quantum mechanics must answer. Then, the various interpretations of quantum mechanics are characterized, and distinguished from one another, in terms of the possible answers to these two questions.

3.2 The two slit experiment performed with bullets

The aim of this section is to examine the results of the two slit experiment when it is performed with bullets. First, this section describes the apparatus for six closely related two slit experiments performed with bullets. Next, the propensity interpretation of probability is used to represent the results of those experiments and to represent the relations between those results. The relations between the results are presented in terms of the principles of strict summation and strict composition. Finally, it is demonstrated that the principles of strict summation and strict composition provide equivalent representations of the same relation.

The experimental apparatus for the two slit experiment, performed with bullets, is described as follows. The source is a machine gun, randomly firing bullets at the diaphragm. The diaphragm is a bullet proof barrier containing two slits just big enough to allow a bullet to pass through, and is arranged so that the slits can be opened or closed. The screen is a wood wall that absorbs, and retains, all bullets reaching it. The region R is simply a specific (possibly painted or outlined) region of the screen. At the end of an experiment both the number of bullets that hit the region R and the number of bullets that hit the screen can be counted, and only those emissions that reach, or hit, the screen are considered statistically relevant.

The apparatus is also equipped with a device that is capable of detecting whether a bullet passes through slit one and a similar detector for slit two. This device could take the form of a set of infra-red motion detectors appropriately placed behind the diaphragm, or it could consist of the experimenter reviewing high speed film of each trial. The result is that if the detector is "on" during the experiment, then, for each trial, the experimenter would be able to determine whether the bullet passed through slit one or slit two. In this experiment all factors are held fixed except for the status of the slits and the status of the detector. The results of an experiment that is run with the detector on can be recorded using two columns: on trials where a bullet is detected at slit one, "hits" and "misses" of region R are recorded in column 1; and on trials where a bullet is detected at slit two, "hits" and "misses" of R are recorded in column 2. Recall that it was assumed that the detection process is 100 percent accurate; consequently, the number of bullets detected at a given slit is equal to the number of bullets passing through that slit. Thus, the number of "hits of R given that the bullet passed through slit one" can be calculated using the data from column 1, the number of "hits of R given that the bullet passed through slit two" can be calculated using data from column 2, and "number of hits of R" can be calculated using data from both columns.

Define B_1 to be the two slit experiment in which the source emits bullets and only slit one is open, define B_2 as the system where only slit two is open, and define B_3 as the system with both slits open. For each of these systems the detector is left "off." Consequently, for an arbitrary region R, the set of outcomes for each system is $\{R, \sim R\}$. Three new systems can be created by turning the detectors "on" in each of the previous systems. A prime sign (′) indicates that the detector is "on"; for example, the system B'_3 is defined as having both slits open with the detector "on." The set of outcomes for each "primed" system is $\{R, \sim R\}\times\{P_1, \sim P_1\}\times\{P_2, \sim P_2\}$, where P_1 and P_2 are defined as the events "the emission passed through slit one" and "the emission passed through slit two," respectively.

Now, assume that there is a single apparatus that is capable of providing the basis for all six systems: the desired system can be produced by opening and closing the slits and turning the detectors on or off. That is, for some event X, specifying the common causal factors, O_1 = "slit one is open at time t_i," O_2 = "slit two is open at t_i," and event D = "the detector is on," all six systems can be defined as follows: $B_1 = X\&O_1\&\sim O_2\&\sim D$, $B_2 = X\&\sim O_1\&O_2\&\sim D$, $B_3 = X\&O_1\&O_2\&\sim D$, $B'_1 = X\&O_1\&\sim O_2\&D$, $B'_2 = X\&\sim O_1\&O_2\&D$, and $B'_3 = X\&O_1\&O_2\&D$. There are a number of manners in which the empirical results for these experiments can be reported. For example, the distributions of hits of regions in different positions (X) on the screen are given in figure 3.1. And these distributions correspond to the values of $Pr(R:B_1)$, $Pr(R:B_2)$, and $Pr(R:B_3)$, respectively.

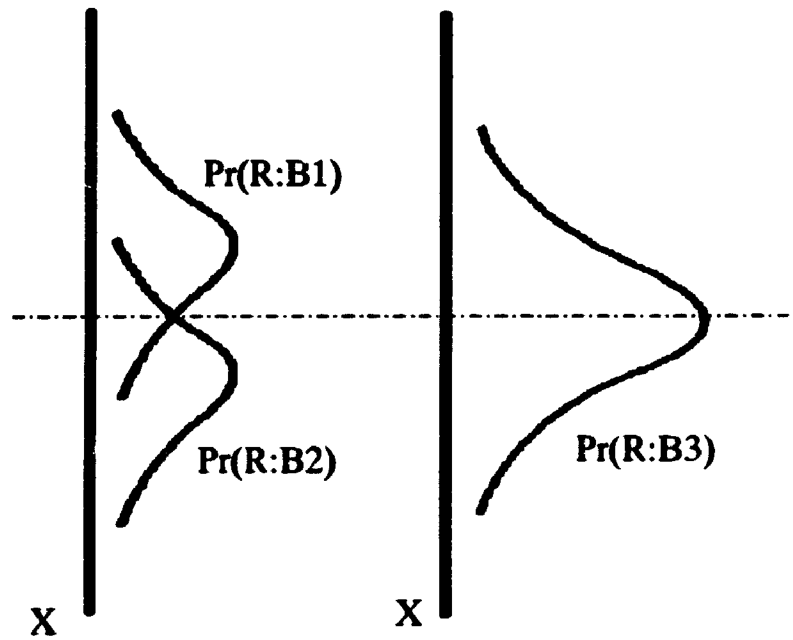

Figure 3.1. Distributions of hits for the bullet experiments.

The results of and relations between these systems have also been represented in terms of the following equalities. The principle of *strict summation* holds between systems B_1, B_2, and B_3:

$$\mathbf{Pr}(R:B_3) = \mathbf{Pr}(R:B_1) + \mathbf{Pr}(R:B_2) \quad (3.1)$$

That is, the results from B_3 can be obtained by simply adding the results of the experiments involving B_1 and B_2. For all systems, the bullets are localized at the screen:

$$\mathbf{Pr}(R\&{\sim}R:B_i) = 0 \text{ and } \mathbf{Pr}(R\&{\sim}R:B'_i) = 0 \quad (3.2)$$

for $i \in \{1, 2, 3\}$. In other words, for all six experiments, there is no trial in which an emission both "hits" and "misses" region R. The results of system B'_3 obey the mathematical principle of *strict composition* :

$$Pr(R:B'_3) = Pr(R\&P_1:B'_3) + Pr(R\&P_2:B'_3)^2 \qquad (3.3)$$

The results of system B'_3 also demonstrates that the bullets are localized in the region of the slits:

$$Pr(P_1\&P_2:B'_3) = 0 \qquad (3.4)$$

That is, system B'_3 produces no bullets that are detected at both slits on the same trial.

Given the similarity of the relationship between the results of systems B_1, B_2, and B_3 and the results of system B'_3--expressed by the principles of strict summation and strict composition, respectively--it is natural to make the claim that the principles of strict summation and strict composition provide alternative statements of the same property, or alternative statements of the relations between the bullet systems. Ultimately, the fact that both strict summation and strict composition hold provides evidence for the claim that all six systems are governed by, or "composed of," the same causal factors.

To end this section, it is demonstrated that the principles of strict summation and strict composition can, in fact, be derived from one another with the consideration of two more sets of empirical results--namely those equalities stated in (3.5) and (3.9) below. The interpretation of the results of the

---

[2]Strict composition is often expressed in terms of conditional probabilities. For simplicity and clarity we use conjunctions. See "A.4 A derivation of alternative expressions of the principle of strict composition" of Appendix 1 for the derivation and statement of equivalent expressions of strict composition.

bullet experiments, the significance of the derivation given below, and discussions concerning the causal similarity of the systems are taken up in the next section.

From empirical results, the following equalities hold:

$$\mathbf{Pr}(R{:}B'_1) = \mathbf{Pr}(R{:}B_1), \tag{3.5}$$

$$\mathbf{Pr}(R{:}B'_2) = \mathbf{Pr}(R{:}B_2), \text{ and}$$

$$\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R{:}B_3)$$

Step 1: by substituting into strict summation--$\mathbf{Pr}(R{:}B_3) = \mathbf{Pr}(R{:}B_1) + \mathbf{Pr}(R{:}B_2)$, equality (3.1) above--according to the equalities in (3.5) it follows that

$$\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R{:}B'_1) + \mathbf{Pr}(R{:}B'_2) \tag{3.6}$$

Step 2: since $\mathbf{Pr}(P_1{:}B'_1) = \mathbf{Pr}(P_2{:}B'_2) = 1$, it must be the case that

$$\mathbf{Pr}(R{:}B'_1) = \mathbf{Pr}(R\&P_1{:}B'_1) \text{ and} \tag{3.7}$$

$$\mathbf{Pr}(R{:}B'_2) = \mathbf{Pr}(R\&P_2{:}B'_2)$$

Step 3: substituting into (3.6) according to the equalities in (3.7) yields

$$\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R\&P_1{:}B'_1) + \mathbf{Pr}(R\&P_2{:}B'_2) \tag{3.8}$$

The following equalities are also empirically verifiable,

$$\mathbf{Pr}(R\&P_1{:}B'_3) = \mathbf{Pr}(R\&P_1{:}B'_1) \text{ and} \tag{3.9}$$

$$\mathbf{Pr}(R\&P_2{:}B'_3) = \mathbf{Pr}(R\&P_2{:}B'_2)$$

Step 4: substituting into (3.8) according to the equalities in (3.9) produces the principle of strict composition: $\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R\&P_1{:}B'_3) + \mathbf{Pr}(R\&P_2{:}B'_3)$, equation (3.3) above.

Note that each step in this derivation can be reversed to produce a derivation of the principle of strict summation from the principle of strict composition. Thus, the results of these experiments are very closely related;

they are so closely related that the principles of strict summation and strict composition must be considered to be equivalent. The next section examines the manner in which these principles are equivalent, by providing an analysis of each step of the derivation given above. Furthermore, the examination of the next section supports the claim that all six systems are governed by the same causal factors.

3.3 Controlled experiments and causal relevance

The aim of this section is to demonstrate the significance of the derivation given in section 3.2. Specifically, it is demonstrated that the derivation cannot proceed in terms of mathematical probability alone, and that the intended use of the argument presupposes a deeper analysis of experimentation than is given by most ''nonpropensity'' interpretations of probability. In fact, it is argued that the derivation essentially presupposes the results of controlled experiments, and that the propensity interpretation can both reveal the fact that this presupposition is made, and reveal the force and reason behind making this presupposition. Also, it is demonstrated that controlled experiments provide the basis for defining the relationship between propensity functions for different systems. Ultimately, it is demonstrated that this presupposition, and the propensity interpretation of the results of the experiments with bullets, are necessary to support the claim that the results of the bullet experiments provide evidence that each experiment is governed by the same ''type'' of causal factors.

Before examining the derivation in section 3.2, it is important to recognize the restrictions imposed by, or limitations of, the propensity calculus. Recall that a propensity function is defined *for* a system and *over* a Boolean σ-algebra on a set of possible outcomes. There is no formal relationship between propensity functions defined for different systems or between propensity functions defined on different event spaces. Although systems B_1, B_2, and B_3 each have {R, ~R} as its outcome space, the status of the slits is different in each system. Consequently, there is no formal relation between the results of the experiments carried out on systems B_1, B_2, and B_3. The examination of the derivation in section 3.2 focuses on the manner in which this derivation establishes, or defines, relationships among propensity functions for different systems and on different event spaces.

The validity of the first step in the derivation--establishing the equivalence of strict summation and equation (3.6)--depends on the equalities expressed in (3.5). In order to understand the significance of step one, first consider the manner in which the equalities in (3.5) represent the nature of the relationships between systems $B_1$and B'_1, systems B_2 and B'_2, and systems B_3 and B'_3. To begin, imagine performing the following *controlled experiment* with system B_1. First, measure the ability of B_1 to produce the event R; that is, measure the value of **Pr**(R:B_1). Then alter B_1 in a controlled manner. In other words, turn the detectors on while holding all other factors fixed; that is, convert B_1 to B'_1. Next, measure the ability of B'_1 to produce the same event R; that is, measure the value of **Pr**(R:B'_1). The results of this controlled experiment are represented by the relationship between the values of **Pr**(R:B_1) and **Pr**(R:B'_1).

According to conventional interpretations of controlled experiments, the status of the detectors is a causal factor for R (in system B_1 or B'_1) if and only if $\mathbf{Pr}(R{:}B_1) \neq \mathbf{Pr}(R{:}B'_1)$. Thus, the first equality of (3.5), $\mathbf{Pr}(R{:}B_1) = \mathbf{Pr}(R{:}B'_1)$, demonstrates that the status of the detector is not causally relevant to the system's ability to produce event R. Since the only difference between systems B_1 and B'_1 is the status of the detectors, and since having the detectors on or off has no effect on a system's ability to produce the event R, the causal factors responsible for the event R must be "within" the common element of systems B_1 and B'_1.

In terms of the propensity interpretation, if the difference between two systems is not statistically relevant to the production of an event, then that difference is not a causal factor in those systems's ability to produce that event. Formally, for any pair of systems S_1 and S_2 such that for some conjunction of events X and some event F, $S_1 = X\&F$ and $S_2 = X\&{\sim}F$, event F (and ~F) is not a causal factor for the event A in systems S_1 and S_2 if and only if $\mathbf{Pr}(A{:}S_1) = \mathbf{Pr}(A{:}S_2)$. In this manner, the propensity interpretation provides a method of representing the results of controlled experiments. Furthermore, *controlled experiments provide the basis for defining the relationships between propensity functions defined for different systems.* Under the propensity interpretation, then, an equality or inequality involving propensity statements for different systems, *on the same event*, is a statement about the causal relevance of the difference between those systems *for that event.*

On this interpretation, the second and third equalities of (3.5) provide similar results concerning the causal influence of the detectors among experiments with slit one closed and slit two open (systems B_2 and B'_2) and the causal influence of the detectors among experiments with both slits open (systems B_3 and B'_3). Consequently the equalities given in (3.5) provide a representation of the fact that the status of the detectors is not causally relevant to any of the six systems. Given this understanding of the equations in (3.5), the physical and causal implications of step 1 are made clear.

According to the propensity interpretation of relations between propensities defined for different systems on the same event, the significance of the principle of strict summation can be interpreted as follows. Given that the status of the slits is the only difference between systems B_3 and B_1 and between B_3 and B_2, strict summation reveals that the status of the slits is causally relevant to the production of the event R since strict summation implies that $\mathbf{Pr}(R{:}B_3) \neq \mathbf{Pr}(R{:}B_1)$ and $\mathbf{Pr}(R{:}B_3) \neq \mathbf{Pr}(R{:}B_2)$, when $\mathbf{Pr}(R{:}B_1) \neq 0$ and $\mathbf{Pr}(R{:}B_2) \neq 0$. Furthermore, the mathematical structure of strict summation provides a statement of the "manner in which" the status of the slits is causally relevant to the occurrence of event R.

Now consider the significance of step 1. Strict summation (3.1) provides a statement of the manner in which the status of the slits is causally relevant to the occurrence of event R among the systems with the detectors turned off. The equalities in (3.5) reveal that the status of the detectors is causally irrelevant to the production of the event R in the bullet systems. From these two facts it follows that the manner in which the status of the slits is causally relevant (or not) to the production of the event R must be the same for the systems with the detector turned on. This relationship of causal relevance is represented by equation (3.6). Thus, in establishing (3.6), step 1 reveals that the manner in which the status of the slits is relevant to the event R is unaffected by the status of the detector.

Step 2 involves relations between propensities that are defined for the same system on the same set of events. Consequently, step 2 proceeds in terms of the axioms and theorems of probability. From a measure theoretic view, any event is equivalent to the conjunction of that event and a tautology. For example, the propensity function for B'_1 assigns the event P_1 a value of one; that is, event P_1 is a tautology for this function. Thus, the propensity function defined for B'_1 must assign the same values to the events R and R&P_1. That is, step 2 establishes that the events R and R&P_1 are considered equivalent, and can be fully interchanged within the scope of the propensity function defined for B'_1 without a loss of "meaning" in measure theoretic terms. Together with similar reasoning in terms of the function for B'_2 and events R and R&P_2, it follows that the equalities in (3.7) represent the fact that the propensity function defined for B'_1 considers the events R and R&P_1 to be equivalent, and that the propensity function defined for B'_2 considers the events R and R&P_2 to be equivalent.

Step 3 can be understood as follows. Given the interpretation of (3.7) above, equality (3.6) can be represented in terms of events R&P_1 and R&P_2 without any significant "measure theoretic" loss of meaning. That is, (3.8) represents the same relationship that (3.6) represents except that (3.6) states the relationship in terms of the events R&P_1 and R&P_2, whereas (3.6) states the relationship in terms of the event R.

Step 4 involves relations between propensities defined for system B'_3 and systems B'_1 and B'_2. Consequently, step 4 depends on empirical results corresponding to appropriate "controlled experiments." The equalities of (3.9) provide these results. According to the first equality of (3.9), the status of slit two is not causally relevant to the production of the event $R\&P_1$ in systems B'_3 and B'_1. Thus, the manner in which B'_1 is causally relevant to event $R\&P_1$ is the same as the manner in which B'_3 is relevant to the event $R\&P_1$. Similarly, the second equality of (3.9) represents the fact that the status of slit one is not causally relevant, in systems B'_3 and B'_2, to the production of the event $R\&P_2$. Therefore, step 4 establishes that the manner in which the events $R\&P_1$ and $R\&P_2$ are produced is unaffected by the status of the stilt through which the bullet does *not* pass. Does this correspond to the propensity interpretation of the principle of strict composition?

The principle of composition states that the causal factors for the event R (in B'_3) can be represented as the sum of the causal factors for the events $R\&P_1$ (in B'_3) and $R\&P_2$ (in B'_3) just as the principle of strict summation states that the causal factors for R (in B_3) can be represented as the sum of the factors for R (in B_1) and the causal factors for R (in B_2). This interpretation is supported by the analysis of the derivation given above. In fact, the analysis of the derivation supports a stronger claim: the principles of strict summation and strict composition are alternative representations of the relationship between the causal factors for the event R in "bullet systems."

The principle of strict composition represents a relationship between different events, and it is expressed in terms of a single propensity function.

On the other hand, the principle of strict summation represents a relationship between different propensity functions, and is expressed in terms of the same event. Each principle provides a truly distinct manner of analyzing the relationship between causal factors for the event R, and the analysis of the derivation given above allows the unification of the results of these distinct methods of analysis. Furthermore, the unifying role of the derivation is a direct result of both the incorporation of empirical results and the propensity interpretation of probability. Equations (3.5) and (3.9) provide empirical evidence that is essential to the derivation of strict composition from strict summation (and vice versa). Moreover, the manner in which this empirical evidence is used is tantamount to supplying the results of controlled experiments. Furthermore, the propensity interpretation of the probability calculus enables the significance of the empirical data (as results of controlled experiments) to be appreciated, and maintained, through each step of the derivation. This appreciation is made possible by the fact that the propensity interpretation represents probabilities as a three-place relation, providing a unified and formal representation of particular experimental setups and their relations to the events they produce and the propensity values assigned to those events. Generally, in so far as non-propensity interpretations of probability do not provide a unified, formal representation of the experimental setup, no other interpretation of probability is able to track this influence of causal factors in terms of the physical experiments and their empirical results within one unified formal system.

Traditionally, the fact that bullets are localized in the region of the slits has been taken to be the most important characteristic of "classical" particle systems. The reason that the principles of strict summation and strict composition hold is the fact that a bullet cannot pass through both slits. It is the following claim that underlies most attempts to characterize B_3 and B'_3 as "particle" systems:

> If the detector had been turned on (and the events P_1 and P_2 were measured) while running experiments with system B_3 then the results associated with B'_3 would have been recorded; specifically, the bullets would have been found to be localized at the region of the slits (3.4), and so strict composition (3.3) would have been found to be true.

The force and intent of this counterfactual can be captured in terms of the propensity account by asserting that the following equalities hold:

$$\mathbf{Pr}(P_1\&P_2:B_3) = 0 \tag{3.10}$$

$$\mathbf{Pr}(R:B_3) = \mathbf{Pr}(R\&P_1:B_3) + \mathbf{Pr}(R\&P_2:B_3) \tag{3.11}$$

Yet, these equalities cannot be empirically tested using B_3 alone, and they are not well-defined according to the formalism of the propensity interpretation since the propensity function for B_3 is not defined over the events P_1 and P_2.

The proper method to assert that systems B_3 and B'_3 have similar causal structure is to note that the difference between them is not causally or statistically significant. This method incorporates the empirical results of controlled experiments and is empirically testable. Moreover, the fact that the bullets are localized in the region of the two slits (3.4) implies strict composition and, through the derivation discussed above, it also implies strict summation. But it is also the case that (3.4) is not equivalent to the principles of strict composition and strict summation.[3] There is a significant difference between claiming that (A) the bullet systems are the same "because" in each system the bullets are localized at the slits, and claiming that (B) the bullet systems are the same "because" in each system the causal factors responsible for the event R are related to each other in the same manner--in the manner captured by strict summation and strict composition.

Statement (A) involves the hypothetical or counterfactual claim that system B_3 behaves in the same manner when it is unobserved as it does when it is observed. This claim is not empirically testable. Statement (B) claims that the systems are equivalent, relative to some empirical criteria. Specifically, statement (B) is made relative to strict summation and strict composition; consequently, statement (B) is empirically testable. Statement (B) does, however, provide evidence, or support, for statement (A). If the causal factors for every event that a group of systems produces are related to each other in every manner then those systems must be equivalent in every sense.

The distinction between statements (A) and (B) raises an important question: "Is there any physically meaningful difference between having a propensity function defined for B_3 and on $\{R, \sim R\}$ and having a propensity function defined for B_3 and on $\{R, \sim R\} \times \{P_1, \sim P_1\} \times \{P_2, \sim P_2\}$?" For, if this distinction is simply an insignificant artifact of the notation that is adopted here, then the distinction between the two propensity functions can simply be

[3]See "A.5 A derivation of the principle of strict composition from the fact that bullets are localized at the slits" in Appendix 1 for a proof of these last two claims.

dropped and the controversy between statements (A) and (B) can easily be avoided.

Recall, first, that the only physical difference between systems B_3 and B'_3 is that the detector is "on" in the case of B_3 and, second, that detection is 100 percent accurate. Furthermore, given that the status of the detector is not statistically relevant to the production of the event R, the only effect of having the detector turned on in the case of system B_3 is that the occurrence of the events P_1 and P_2 can be determined after the initialization of the experiment. It is this physical (or measure-theoretic) condition prior to the start of the experiment (as part of the system) that provides the motivation to include the events P_1 and P_2 in the event space of the propensity function defined for B'_3. There is no similar physical or measure-theoretic motivation for including the events P_1 and P_2 in the event space of the propensity function defined for B_3.

The motivation for considering a propensity function that is defined for B_3 and over $\{R, \sim R\} \times \{P_1, \sim P_1\} \times \{P_2, \sim P_2\}$ is provided by *hypothetical* reasoning which arises from the equivalence of the causal factors for R in the six systems. Given that systems B_3 and B'_3 are considered to be causally equivalent with respect to R, it is legitimate to suppose that the events P_1 and P_2 do, in fact, occur (measured or not) during the experiments performed with B_3. Also, the frequency of the occurrence of P_1 and P_2 (if they were "measured") would have to be described by a propensity function defined *over* $\{R, \sim R\} \times \{P_1, \sim P_1\} \times \{P_2, \sim P_2\}$ and *for* B_3.

Thus, there is a physically and measure-theoretically meaningful difference between having a propensity function defined for B_3 and on $\{R, \sim R\}$ and having a propensity function defined for B_3 and on $\{R, \sim R\} \times \{P_1, \sim P_1\} \times \{P_2, \sim P_2\}$: one function represents a physical system and the other system represents a hypothetical system. Furthermore, there is a legitimate motivation for defining a propensity function for B_3 and on $\{R, \sim R\} \times \{P_1, \sim P_1\} \times \{P_2, \sim P_2\}$: it is defined in order to express the hypothetical consequences of the statistical equivalence of systems B_3 and B'_3 with respect to R. This distinction between propensity functions defined on different event spaces is not simply an insignificant artifact of the notation that is adopted here, and it would be a mistake to simply drop the distinction between the two propensity functions. Equalities like (3.10) and (3.11) have hypothetical force that is based on causal information. Despite the fact that the propensity functions in equalities like (3.10) and (3.11) are defined for a single system, equalities like (3.10) and (3.11) cannot be confirmed by a single experiment, nor do they represent the results of a single experiment.

3.4 The two slit experiment performed with water waves

The aim of this section is to examine the results of the two slit experiment when it is performed with water waves. First, this section describes the apparatus for six closely related two slit experiments performed with water waves. Next, the propensity interpretation is used to represent both the results of those experiments and the relations between those results. It is demonstrated that the relations between the results cannot be presented in terms of the principles of strict summation and strict composition. Specifically, the results exhibit an interference effect and violate the principles of strict summation and strict composition. One interesting result is that this interference effect can be represented only in a limited number of ways. Finally, it is demonstrated that the reason for these failures is that the status of the slits variable is causally relevant to the results.

Consider now a group of experiments utilizing water waves as the emissions. The apparatus is setup in a shallow pool of water. The source is a vibrating "bob" that produces waves of a uniform intensity or size, and the rate at which it produces these waves can be controlled. The diaphragm can be any suitable wave barrier with two closable slits. The screen must make use of a "beach-like" absorber to eliminate any possibility that the waves arriving at the screen are reflected. Similarly, the detector must be capable of recording the height of the waves in the region R without reflecting the waves. Small pieces of cork could be placed in front of a beach so that the experimenter can record the height of the wave as it passes under the cork at some fixed distance (perpendicular) from a specific region of the beach. Thus, the "screen" is in fact an imaginary plane, parallel to the beach, at which the experimenter can make the necessary measurements.

Furthermore, the apparatus is equipped with detectors that are capable of recording the height of the waves at each slit without disturbing the wave. The back of the diaphragm could have markings that would enable the experimenter to simply read off the height of a wave as it passes through the slit. As above, define W_1 as the two slit experiment where the source emits waves and only slit one is open, define W_2 as the system where only slit two is open, and define W_3 as the system with both slits open. For each of these systems the detector is left "off"; that is, the wave height at the slit is not recorded, but the height at region R is recorded. Systems W'_1, W'_2, and W'_3 can be created by turning the detectors "on" in the previous systems. Thus, for systems W'_1, W'_2, and W'_3 the wave height is recorded at both slits and at region R.

Again, assume that there is an apparatus that is capable of providing the basis for all six systems, and that the systems can be expressed as a conjunction of some event X (specifying the common causal factors) and the status of the variables. The first thing to notice about the results of these experiments is that the results are recorded in terms of wave heights rather than propensities, probabilities, or "hits." In fact, the results of these experiments are not represented in terms of wave heights but in terms of wave intensities, where the intensity of a water wave is proportional to the square of the absolute value of its height. Define I(A:S) as "the intensity of the wave at A produced by system S." Then, as in the case of the propensity values for the bullet experiments, the results of the wave experiments can be represented in terms of the distribution of the values of $I(R{:}W_1)$, $I(R{:}W_2)$, and $I(R{:}W_3)$ for different positions (X) of the region R along the screen. This manner of representation is given in figure 3.2.

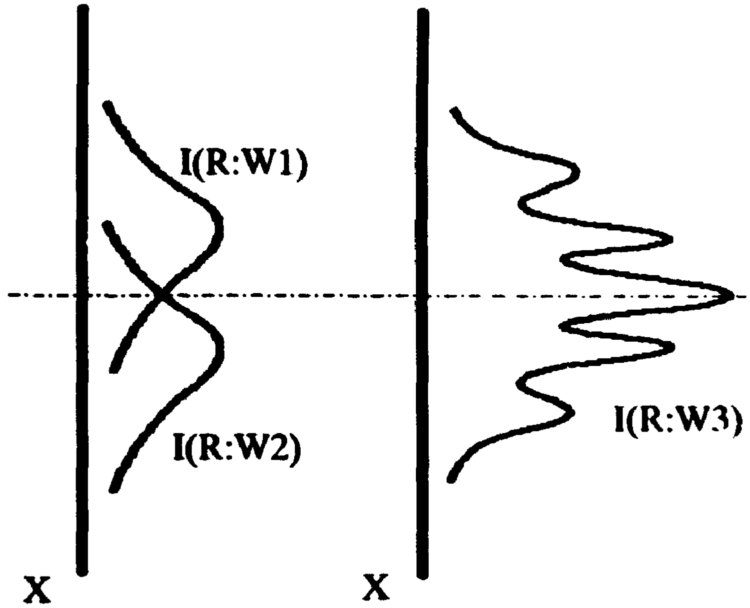

Figure 3.2. Distributions of intensities for the wave experiments.

The other results and relations between these systems are represented in a manner that is similar to the representation of the results of the bullet experiments. First, the principle of strict summation is violated.

$$I(R{:}W_3) \neq I(R{:}W_1) + I(R{:}W_2) \tag{3.12}$$

That is, the results from W_3 cannot be obtained by simply adding the results of the experiments involving W_1 and W_2. In fact, the relation between $I(R{:}W_3)$, $I(R{:}W_1)$, and $I(R{:}W_2)$ can be represented as follows: for $I(R{:}W_1) = |h_1|^2$ and $I(R{:}W_2) = |h_2|^2$, $I(R{:}W_3) = |h_1 + h_2|^2$. Also, water waves are not localized at the screen, for some region R′ distinct from region R.

$$I(R{:}W_i) \neq 0 \text{ and } I(R'{:}W_i) \neq 0, \text{ and} \tag{3.13}$$

$$I(R{:}W'_i) \neq 0 \text{ and } I(R'{:}W'_i) \neq 0, \text{ for } i \in \{1, 2, 3\}$$

That is, there are trials in which the intensity of a wave is spread out over more than one position on the screen. It is also noted that the following equalities hold,

$$I(R{:}W'_i) = I(R{:}W_i), \text{ for } i \in \{1, 2, 3\} \qquad (3.14)$$

That is, the status of the detectors is not relevant to the intensity of the wave at region R in any of the experimental setups. Finally, water waves are not localized at the slits

$$I(S_1{:}W'_3) \neq 0 \text{ and } I(S_2{:}W'_3) \neq 0 \qquad (3.15)$$

That is, non-zero intensities are recorded at both slit one and slit two. System W'_3 produces waves that pass through both slits on the same trial.

Given the similarity of the relations between intensities for the water wave experiments, and the relations between probabilities for the bullet experiments, there is motivation to analyze the results of the wave experiments in a manner that is similar to the analysis of the bullet experiments. Before pursuing this possibility, however, it is important to note the physical explanation for the results in the wave experiments. The physical explanation runs as follows. With one slit open, a wave is emitted by the source and passes through the open slit producing an expanding circular wave centered on the slit. The wave travels across the region between the diaphragm and the screen, and the intensity of the wave decreases as the distance travelled increases. The expanding wave first reaches the screen at a point opposite the slit, then it "rolls" across the screen, away from the point of impact, in both directions. For any point on the screen, the farther it is away from the slit, the longer it takes the wave crest to hit that point. Consequently, the intensity of the wave for region R is proportional to the distance that R is from the slit (as represented in figure 3.2).

With both slits open, a wave is emitted from the source and is split into two expanding circular waves as it passes through both slits. Each wave is centered on the slit of origin and the waves have the same initial intensity (assuming the slits to be of equal size and configuration). Interference takes place as the waves spread out over the region between the diaphragm and the screen. For example, ''double crests'' form where two crests meet, ''double troughs'' form where two troughs meet, and the water is practically undisturbed where crests and troughs meet. The result of this interaction is a new and distinct wave front which moves away from the region of the two slits across the area between the diaphragm and the screen. This new wave front consists of double waves, double troughs, and everything in between. The intensities recorded at the screen vary according to both the particular segment of the wave front that hits the region R (double crest, double trough, or something in between), and the distance between R and the slits. The distribution of the wave intensity over the possible positions of region R on the screen is given in figure 3.2.

Consider now an analysis of the results of the experiments with water waves. Notice that the violation of strict summation (3.14) indicates that the status of the slits is causally relevant to the intensity at region R, since the only difference between the two systems was the status of slits, and the intensities were different in each experiment with the detector turned off. The violation of strict summation represents the ''manner in which'' the status of the slits is relevant to the intensity at R as much as the holding of strict summation represents the ''manner in which'' the status of the slits is relevant to the number of hits at R in the bullet experiments. The manner in which the status of the slits is causally relevant is represented by the equality $I(R{:}W_3) = |h_1 + h_2|^2$, where $I(R{:}W_1) = |h_1|^2$ and $I(R{:}W_2) = |h_2|^2$. The causal factors responsible for the intensity of waves at region R for the three wave systems, W_1, W_2, and W_3 are not "additive" in the same manner as the causal factors for bullets hitting region R.

Notice that, as indicated by the equalities in (3.14), the status of the detector is not causally relevant to the intensity of a wave at region R. Thus, it follows that $I(R{:}W'_3) \neq I(R{:}W'_1) + I(R{:}W'_2)$, in an analogous manner to the derivation of $\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R\&P_1{:}B'_1) + \mathbf{Pr}(R\&P_2{:}B'_2)$ in step 1 of the bullet derivation. The intensity of a wave at R for system W'_3 can be expressed in terms of the intensities of waves produced by W'_1 and W'_2, and the manner of expressing this relationship is the same whether the detector is on or off. Considering the analysis of particle systems, one can infer that the causal factors responsible for the intensity of the waves at region R in W'_1, W'_2, and W'_3 are related in the same manner as those factors in W_1, W_2, and W_3. Both W_3 and W'_3 exhibit the effects of interference.

Also, notice that the bullets are localized at the slits, whereas the water waves are not. Considering the connection between the localized nature of bullets and strict composition, does it follow that an equality analogous to strict composition is also "violated" in the case of the water wave experiments? In other words, can the relationship stated above in terms of W'_1, W'_2, and W'_3 be expressed solely in terms of components of W'_3, in the same manner that strict composition for the bullet experiments is expressed solely in terms of components of B'_3? In the case of water waves it is not obvious that a physical interpretation for intensities of conjunctive or conditional regions can be provided. For example, how can the value of the intensity analog of $Pr(R\&P_1{:}B'_3)$ be measured? On the one hand, "the intensity of the wave at slit 1 and region R" is obviously not the desired quantity. The desired quantity corresponds to something like "the contribution of a wave originating at slit one to the intensity measured at R," yet it is not clear that the value of this quantity can be measured on system W'_3.

It appears that the relationship stated in the violation of strict summation (3.12) simply cannot be expressed solely in terms of system W'_3. There is no analog of the violation of strict composition in the case of water waves. Generally, it is not clear that the intensities are defined over a Boolean σ-algebra on the set of possible outcomes Ω: it may be the case that the "intensity function" can be defined only over the set Ω. Note that the fact that the intensity function has a restricted domain (relatively speaking) does not deny the existence of a relationship between the results of the experiments; it simply denies the ability to measure that relationship and represent that relationship solely in terms of the system W'_3.

Bullet experiments are "additive," and this additivity can be expressed in a number of distinct ways: it can be represented in terms of alternative events, and in terms of alternative systems. The water experiments are not additive; they show the effects of interference. Furthermore, the interference effect can be represented as a relationship between systems and events. But, as it turns out, this relationship cannot be represented in as wide a variety of manners as additivity can for bullets. The limitations in representing the relationships between systems and events may be the result of either the limitations of our systems of measurement or the nature of the phenomenon itself. There are limitations on the manners of expressing the interference effect imposed by measurement.

3.5 The two slit experiment performed with electrons

The aim of this section is to examine the results of the two slit experiment when it is performed with electrons. First, this section describes the apparatus for six closely related two slit experiments performed with bullets. Next, the propensity interpretation is used to represent the results of those experiments and to represent the relations between those results. Specifically, it is demonstrated that (like the water wave experiments) the results violate the principle of strict summation, but (like the bullet experiments) the results satisfy the principle of strict composition. Then it is demonstrated that the explanation of these results is that both the status of the slits and the status of the detector are statistically relevant. From this fact it is argued that the "paradox" of the two slit experiment arises from the assumption that there is a causal explanation of the manner in which the status of the detector is relevant to the results, and the paradox does not arise from the empirical results themselves. Finally, the propensity analysis of the electron systems is used to refute both the quantum probability arguments by Luigi Accardi (1984, sec. 2) to reject Bayes's theorem, and the quantum logic arguments by Hilary Putnam (1979, sec. 4) to reject distribution.

The experimental apparatus for the two slit experiment performed with electrons is described in detail in chapter 1: the basic apparatus (without a detector) is described in section 1.2, and the apparatus with a detector is described in section 1.4. Recall that the detector is arranged so that all electrons passing through the slits are detected when the detector is on. Also, only those electrons reaching the screen are statistically relevant. Thus, all electrons are detected at one slit or the other, and all electrons proceed to the screen to hit or miss region R. Define E_1, E_2, E_3, E'_1, E'_2, and E'_3 in an analogous manner to the previous experiments: the E indicates that the experiment is performed with electrons; the subscripts indicate the status of the slits (1 = only slit one open, 2 = only slit two open, and 3 = both slits open); and the prime indicates the status of the detector (no prime indicates that the detector is off, and a prime indicates that the detector is on).

The first thing to notice is that the results of the experiment can be recorded in terms of the number of electrons arriving at the screen or passing through the slits. In fact, in the case of using an electron sensitive film with a pre-marked region R, the experimenter can count the number of "hits" or absorptions in the pre-marked region R, and the number of "misses" or absorptions on the rest of the film in a manner similar to counting bullet holes in the wood screen. Of course, this method assumes that there are no overlapping holes or points of absorption that may interfere with the accuracy of the counting procedure; it may be best to "count" after each trial and to replace the film or wood screen on each trial. Alternatively, the experimenter could simply count the clicks of the geiger counters. In any event, the results can be represented in terms of propensity values (based on numbers of hits and misses rather than the intensities of waves) and the relations between them, as in the case of the experiments with bullets.

Just as with the waves, however, strict summation is violated.

$$\mathbf{Pr}(R{:}E_3) \neq \mathbf{Pr}(R{:}E_1) + \mathbf{Pr}(R{:}E_2) \quad (3.16)$$

That is, the results from E_3 cannot be obtained by simply adding the results of the experiments involving E_1 and E_2. In fact, the distributions of $\mathbf{Pr}(R{:}E_1)$, $\mathbf{Pr}(R{:}E_2)$, and $\mathbf{Pr}(R{:}E_3)$ for different positions (X) of R are the same as those given in figure 3.2 for water waves. Furthermore, for two complex numbers ϕ_1 and ϕ_2, $\mathbf{Pr}(R{:}E3) = |\phi_1 + \phi_2|^2$, where $\mathbf{Pr}(R{:}E_1) = |\phi_1|^2$ and $\mathbf{Pr}(R{:}E_2) = |\phi_2|^2$. Thus, there is evidence of "wave-like" interference among the experiments with the detectors turned off. However, electrons are localized at the screen for all systems.

$$\mathbf{Pr}(R\&{\sim}R{:}E_i) = 0 \text{ and } \mathbf{Pr}(R\&{\sim}R{:}E'_i) = 0, \text{ for } i \in \{1, 2, 3\} \quad (3.17)$$

There is no trial on which an emission both hits and misses region R; this is a characteristic of particle systems. The principle of strict composition holds for system E'_3 just as it does for system B'_3.

$$\mathbf{Pr}(R{:}E'_3) = \mathbf{Pr}(R\&P_1{:}E'_3) + \mathbf{Pr}(R\&P_2{:}E'_3) \quad (3.18)$$

In fact, the distributions for the values of $\mathbf{Pr}(R{:}E'_3)$, $\mathbf{Pr}(R\&P_1{:}E'_3)$, and $\mathbf{Pr}(R\&P_2{:}E'_3)$, for different positions (X) of R, are the same as the distributions for $\mathbf{Pr}(R{:}B_3)$, $\mathbf{Pr}(R{:}B_1)$, and $\mathbf{Pr}(R{:}B_2)$, respectively, as in figure 3.1. Finally, the electrons are localized at the slits for E'_3.

$$\mathbf{Pr}(P_1\&P_2{:}E'_3) = 0 \quad (3.19)$$

That is, system E'_3 produces no electrons that are detected at both slits on the same trial, and this fact is evidence of particle-like locality.

Equality (3.16) indicates that, with the detector turned off, there is wave-like interference, and it may be stated that the electron experiments admit to a wave description but do not admit to a particle description. But, as indicated by equality (3.16), with the detector turned on, this interference is eliminated and there is strict summation. Consequently, it may be stated that the electron experiments admit of a particle description but do not admit of a wave description. These two results can be restated in terms of the principle of complementarily: neither the wave description nor the particle description provides a complete description of the electron experiments, but together they are capable of providing a complete (but complementary) description of the electron experiments. The paradox of the two slit experiment (and of quantum mechanics) rests on the supposition that there is a single (non-complementary) complete physical or causal description of the electron experiments.

The results of the electron experiments clearly indicate that both the status of the slits, and the status of the detector, are causally relevant to the production of event R. In fact, $\mathbf{Pr}(R{:}E_3) \neq \mathbf{Pr}(R{:}E'_3)$ although $\mathbf{Pr}(R{:}E_1) = \mathbf{Pr}(R{:}E'_1)$ and $\mathbf{Pr}(R{:}E_2) = \mathbf{Pr}(R{:}E'_2)$. But it is not this fact alone that presents us with a "paradox." The paradox arises from a consideration of the *manner in which* the status of the detector is causally relevant to the event R. With the detector turned off, the manner in which the status of the slits is causally relevant to R is represented by (3.16). With the detector turned on, the manner in which the status of the slits is causally relevant to R is represented by (3.18). Thus, turning the detector on or off changes the manner in which the status of the slits is causally relevant to R. Neither the wave description nor the particle description offers an explanation of the mechanism by which turning the detector on or off changes the manner in which the status of the slits is causally relevant to R. The paradox of the two slit experiment rests on the supposition that there is an explanation of the mechanism by which the status of the detector "changes" the manner in which the status of the slits is relevant to the production of event R.

The first point that must be recognized concerning the interpretation of these results is that there is no contradiction between the results of the experiments with electrons and the axioms and theorems of logic or probability theory. For example, both Putnam (1979, sec. 4) and Accardi (1984, sec. 2) argue as follows: according to the axioms and theorems of probability, we predict that the following equality holds for the two slit experiment

$$P(R|P_1 \vee P_2) = P(R|P_1)P(P_1) + P(R|P_2)P(P_2) \qquad (3.20)$$

but, experimental results reveal that in fact

$$P(R|P_1 \vee P_2) \neq P(R|P_1)P(P_1) + P(R|P_2)P(P_2) \qquad (3.21)$$

Thus, some fundamental principle used in the derivation of equality (3.20) must be rejected. Putnam argues for the rejection of distributive laws and Accardi argues for the rejection of Bayes's theorem.

The arguments of both Putnam and Accardi fail to maintain the distinction between systems. Yet if they are to maintain that (3.20) and (3.21) are the basis for a contradiction, they must maintain that (3.20) and (3.21) describe the same state of affairs. For example, Putnam and Accardi would have to maintain that (3.20) and (3.21) both describe the same single system, such as E'_3, or the relationships among the same group of systems, such as E_1, E_2, and E_3. The propensity interpretation, however, demonstrates that it is not possible for (3.20) and (3.21) to describe the same state of affairs. On the one hand, equality (3.20) does hold of E'_3 but empirical data demonstrate that inequality (3.21) does not hold of system E'_3. On the other hand, as discussed in section 3.3, if either equality (3.20) or inequality (3.21) describes some relationship among systems E_1, E_2, and E_3, then it must hold in some hypothetical or counterfactual manner, rather than as some direct empirical result, since the propensity function defined for these systems is not defined on the events P_1 and P_2. Yet, the only empirical basis for stating that equality (3.20) holds hypothetically or counterfactually would be that the status of the detector is not causally relevant to R. But according to the empirical results of the experiments above, the status of the detector *is* causally relevant to R. Thus, there is no formal reason to say that equality (3.20) describes some relationship among systems E_1, E_2, and E_3, and there is no empirical evidence to support a hypothetical version of (3.20) describing a relationship among systems E_1, E_2, and E_3.

According to the examination of the electron experiments given above, there is no other single experiment or group of experiments that (3.20) and (3.21) could reasonably be intended to describe. Thus, equations (3.20) and (3.21) cannot provide a description of the same state of affairs.

Consequently, neither (3.20) and (3.21) can provide the basis for a contradiction, or paradox. Therefore, the arguments of Putnam (for the rejection of distributive laws) and of Accardi (for the rejection of Bayes's theorem) are not well-founded.

3.6 The interpretation of quantum mechanics

The aim of this section is to consider the implications of the propensity interpretation of the two slit experiment for current theories of quantum mechanics. First, a review of the salient results of the propensity interpretation of the two slit experiment is given. Then it is demonstrated that the interpretation of quantum mechanics rests on the answers to two fundamental questions concerned with the status of propensity statements. Finally, a consideration of the possible answers to these questions provides the basis for distinguishing among the approaches of hidden variable theory, quantum logic, quantum probability, and the rest of quantum theory.

In the case of the bullet experiments, it was determined that it is physically meaningful to speak of the values of $\mathbf{Pr}(R\&P_1{:}B'_3)$ and $\mathbf{Pr}(R\&P_2{:}B'_3)$, and that it is possible to measure the values of those propensities. Given the physical meaning of these propensities, it was argued that although the propensities $\mathbf{Pr}(R\&P_1{:}B_3)$ and $\mathbf{Pr}(R\&P_2{:}B_3)$ cannot be empirically measured, they can be considered to have hypothetical, or counterfactual, meaning. Furthermore, it was argued that the equivalence of strict summation and strict composition provide evidence for the claim that $\mathbf{Pr}(R\&P_1{:}B_3) = \mathbf{Pr}(R\&P_1{:}B'_3)$ and $\mathbf{Pr}(R\&P_2{:}B_3) = \mathbf{Pr}(R\&P_2{:}B'_3)$. One result of this analysis was that it was determined that there are many alternative expressions of the common manner in which systems B_3 and B'_3 produce the event R. Some alternatives involve the event R, while others involve the events $R\&P_1$ and $R\&P_2$. Some alternatives are expressed hypothetically, while others are expressed in physical terms. In the case of the water waves, it was found that it is meaningless to speak of the intensities $I(R\&P_1{:}W_3)$ and $I(R\&P_2{:}W_3)$, hypothetically or not. Consequently, it is meaningless to speak of the equalities $I(R\&P_1{:}W_3) = I(R\&P_1{:}W'_3)$ and $I(R\&P_2{:}W_3) = I(R\&P_2{:}W'_3)$. The result is that there is a limited number of alternative expressions of the common manner in which systems W_3 and W'_3 produce the event R. Specifically, the interference effect cannot be expressed in terms of the events $R\&P_1$ and $R\&P_2$, hypothetically or physically.

In the case of the electron experiments, the difference between E_3 and E'_3 is causally relevant. That is, measurement has an effect on event R. If there is some common manner in which E_3 and E'_3 produce the event R, then it must be expressed in some manner other than the direct comparison of the results of systems E_3 and E'_3 and event R. As in the cases of the bullet systems and the wave systems, the search turns to the events $R\&P_1$ and $R\&P_2$ (or conditional events $R|P_1$ and $R|P_2$) as the source of alternatives. This raises two fundamental questions for interpretations of quantum mechanics. Specifically,

(1) Are the following propensities physically or hypothetically meaningful?

$\mathbf{Pr}(R\&P_1{:}E_3)$ and $\mathbf{Pr}(R\&P_2{:}E_3)$

and, conditional on an affirmative answer to (1),

(2) Is there justification or evidence for the following equalities?

$$\mathbf{Pr}(R\&P_1{:}E_3) = \mathbf{Pr}(R\&P_1{:}E'_3) \text{ and } \mathbf{Pr}(R\&P_2{:}E_3) = \mathbf{Pr}(R\&P_2{:}E'_3)$$

Hans van den Berg, Dick Hoekzema, and Hans Radder (1990) point out that most interpretations of quantum mechanics give a negative answer to question (1), and this is tantamount to denying that the electron systems are classical particle systems. Given the manner in which measurement is causally relevant in the electron experiments, although it is empirically meaningful to discuss the values of $\mathbf{Pr}(R\&P_1{:}E'_3)$ and $\mathbf{Pr}(R\&P_2{:}E'_3)$, it is not meaningful to discuss the values of $\mathbf{Pr}(R\&P_1{:}E_3)$ and $\mathbf{Pr}(R\&P_2{:}E_3)$. There is no meaningful discussion, hypothetical or not, concerning the events in $\mathbf{Pr}(R\&P_1{:}E_3)$ and $\mathbf{Pr}(R\&P_2{:}E_3)$. On this account, the difference between the "meaningfulness" of the propensities $\mathbf{Pr}(R\&P_1{:}E'_3)$ and $\mathbf{Pr}(R\&P_2{:}E'_3)$, on the one hand, and $\mathbf{Pr}(R\&P_1{:}E_3)$ and $\mathbf{Pr}(R\&P_2{:}E_3)$, on the other hand, is just another point of distinction between systems E_3 and E'_3. It may be pointed out that this approach reinforces the wave-particle duality picture of "electron systems": sometimes electrons behave like particles and sometimes they behave like waves.

Hidden variable theorists provide a positive answer to question (1) and a negative answer to question (2). Consequently, the hidden variable theorists explain the difference between the systems as follows. There is another factor that provides a manner of distinguishing between E_3 and E'_3. In the case of Bohm's hidden variable theory, the "quantum potential" of the system depends on the status of slits, and the quantum potential determines (in part) the values of the propensities in question (2). The "common manner in which" E_3 and

E'_3 produce the event R can be described in terms of the quantum potential. Hidden variable theorists claim that there is a common manner in which the electron experiments produce the event R, except that this manner cannot be described in terms of traditional events such as $R\&P_1$, $R|P_1$, or R. The common causal relation can be described only in terms of a "new event," namely, the value of some hidden variable.

Quantum logic and quantum probability answer both questions in the affirmative. Their goal is to express some common relation in terms of the traditional set of events, but with an alternative set of rules for logic or probability, respectively. In terms of the propensity examination of the problem, quantum probability theorists are looking for an algebra that defines the relations between the different propensity functions that are defined for different systems. Furthermore, the relations between propensity functions defined for different systems are determined by the relations between the empirical results of different experiments. Thus, quantum probability and logic are both searches for an algebraic representation of the empirical relations among experiments.

Putnam (1975, secs. 1 and 7) draws an analogy between the status of the parallel postulate in physical geometry and the status of distributive laws in quantum logic. Similarly, Accardi (1984, sec. 1) draws an analogy between the status of the parallel postulate in physical geometry and the status of Bayes's theorem in quantum probability. Putnam and Accardi agree with other quantum logic and quantum probability theorists that the choice of a logic or probability is an empirical matter. The problem with Putnam's and Accardi's individual approaches is that they also presuppose that equality (3.20) and (3.21) describe the same system. But, as was demonstrated above, this presupposition is tantamount to assuming that the status of the detector is not causally relevant to R. Yet, the examination of the electron experiments demonstrated that the status of the detector is statistically relevant to the event R.

3.7 Concluding remarks

Determining the relations among propensity functions is, in general, an empirical matter: there is no purely logical or algebraic method for determining the relations among propensity functions. The study of the relations among propensity functions that are defined for different systems, and over the same set of events, corresponds to the study of controlled experiments. Furthermore, the principles that are used to represent the relations among propensity functions for different systems are used to characterize "types" of systems. The principles of strict summation and strict composition characterize the "additive" nature of bullet systems. The interference effect and the *violation* of strict composition were used to characterize the "non-additive" nature of the wave systems. The non-additive nature of the electron systems is different from the non-additive nature of the water wave systems; the non-additive nature of electron systems is characterized by the interference effect and the principle of strict composition whereas the non-additive nature of water systems is characterized by the interference effect and the violation of the principle of strict composition.

Controlled experiments and the relations among propensity functions also provide the basis for defining the causal factors in a system. By comparing the differences among types of systems, and the results of those systems, it was demonstrated that neither the status of the slits nor the status of the detector is a causal factor in the bullet experiments. In the case of the water experiments, the status of the slits is a causal factor, but the status of the detector is not a causal factor. For the electron systems, both the status of the slits and the status of the detector are causal factors. This study of causal relations in terms of controlled experiments avoids the problem of the reference class. But there are two conditions that must be recognized. First, the causal relations that are confirmed are expressed in terms of the ability of some factor to produce a given event *in a specified group of systems*. Second, it is often the case that extending the group of systems to a "type" of system involves counterfactual claims, as in the case of stating that the status of the slits is a causal factor in "unobserved" bullet systems.

It was demonstrated that there is no inconsistency in the empirical results of the experiments, and that the "paradox" of the two slit experiment arises only when it is assumed that there is a causal explanation of the manner in which the status of the slits is causally relevant to the event R. Furthermore, it was demonstrated that, if the propensity interpretation is correct, then strictly logical or probabilistic arguments against the consistency of the methods of representing the results of the experiments, such as those of Accardi (1984) and Putnam(1979) respectively, fail. Finally, two questions concerning the interpretation of propensity statements involving certain events were presented, and it was demonstrated that the possible answers to these questions can provide the basis for distinguishing among the various interpretations of quantum mechanics.

4

Propensities and Contrastive Explanation

4.1 Introduction

Theories of contrastive explanation, such as those put forward by Bas van Fraassen (1977 and 1980) and Alan Garfinkel (1981), claim that explanations are not answers to general why questions such as "Why P," but are answers to contrastive why questions such as "Why P rather than Q." The components of contrastive why questions are referred to as the *fact* (P) and the *foil* (Q). The aim of this chapter is to use the propensity interpretation to, first, demonstrate the role of the foil in contrastive explanation, and, second, to identify the advantage of contrastive explanation over non-contrastive explanation.

Central to the presentation in this chapter is a recent discussion of contrastive explanation provided by Peter Lipton (1991a and 1991b). This chapter both criticizes and expands upon Lipton's discussion of contrastive explanation. First, an outline of Lipton's account of contrastive explanation and the presuppositions that are made by this account are provided. Then it is argued that, although Lipton's concepts of a "difference condition" and of a "corresponding cause" help to improve our intuitive understanding of contrastive explanation, there is much more improvement to be made in terms of explaining both the distinguishing features of and the advantages of contrastive explanation. Specifically, it is argued that Lipton's difference condition does not provide adequate restrictions on what qualifies as a corresponding cause, and his account fails to provide a complete analysis of the role of the foil in contrastive explanation.

Next, this chapter applies the propensity interpretation of the relationship between experiments and their possible outcomes (as outlined in chapter 1), as well as the propensity kinematics (as developed in chapter 2), to contrastive explanation. First, it is demonstrated that the propensity account provides the basis for identifying the causal history of events. Then, it is established that the propensity interpretation is capable of formalizing both the presuppositions of Lipton's account as well as the constraints imposed by Lipton's difference condition. By relating this formalization of the difference condition to the discussion of causation in chapters 2 and 3, it is argued that the propensity interpretation provides justification for the difference condition. Then it is demonstrated that the propensity interpretation is capable of elucidating the mechanism by which the difference condition identifies events in the causal history of the fact P.

Finally, the role of the foil in contrastive explanation is examined. First, it is argued that the role of the foil in contrastive explanation is to narrow the reference class. The foil essentially selects a partial causal history of the fact P. Given this role, the foil facilitates the application of the propensity interpretation to the causal history of the fact P and the application of the difference condition to that causal history. It is argued that the notion of a corresponding cause results from the application of the difference condition to the causal history of the fact P that is selected by the foil Q. In conclusion, it is argued that it is the role of the foil in both identifying corresponding causes, and narrowing the reference class, that confers an advantage on the contrastive theory of explanation. In the case of non-contrastive theories of explanation, it is argued that the problems of homogeneity and explanatory ambiguity are encountered in much the same manner (and for the same reasons) that the problem of homogeneity is encountered by theories of chance and causation. But, through the role of the foil, the contrastive approach to explanation is able to avoid the problem of homogeneity and is able (in many cases) to overcome the problem of explanatory ambiguity.

4.2 The presuppositions of contrastive explanation

The aim of this section is to examine the adequacy of Peter Lipton's (1991a and 1991b) account of contrastive explanation. First, the presuppositions of Lipton's account are presented in terms of an example involving the explanation of "why David Lewis went to Monash rather than to Oxford in 1979." This presentation includes a discussion of Lewis's criteria for contrastive explanation and a demonstration that Lewis's criteria are too weak. Then Lipton's difference condition and his resulting notion of a corresponding cause are discussed and evaluated. It is argued that Lipton's account is susceptible to Philip Kitcher and Wesley Salmon's (1987) criticism of Bas van Fraassen's (1977 and 1980) account of contrastive explanation. Specifically, it is maintained that, according to Lipton's account almost anything can qualify as a corresponding cause. Although it may be acknowledged that

Lipton's difference condition overcomes the problems associated with Lewis's account, Lipton's difference condition does not guarantee that the appropriate corresponding cause is identified. Moreover, although Lipton provides an intuitive notion of corresponding cause, this notion fails to provide a detailed understanding of contrastive explanation: it does not provide an account of the role of the foil in contrastive explanation and of the specific advantages of contrastive explanation.

Lipton discusses and defends four presuppositions of contrastive explanation. In order to introduce these presuppositions, imagine, as Lipton does, that we wish to explain why David Lewis went to Monash rather than to Oxford in 1979. In offering an explanation of this event, it is first presupposed that David Lewis actually went to Monash in 1979 and that he did not go to Oxford in 1979. Also, it is presupposed that Lewis could have gone to Oxford. Accordingly, the first two presuppositions of contrastive explanation are that, first, at some time, both the fact P and the foil Q were possible, and second, at some other (presumably later) time, the fact P actually occurred and foil Q did not. The third presupposition is that explaining an event involves providing information concerning the causal history of that event.[1] Thus, explaining why Lewis went to Monash rather than to Oxford involves citing information concerning the events that lead to his trip to Monash, but not to Oxford, and if there are no such events then there is no explanation of why Lewis went to

---

[1]Note that this is not a presupposition of all contrastive theories of explanation. For example, van Fraassen (1977 and 1980) does not make this presupposition.

Monash rather than to Oxford. The motivations for these first three presuppositions, and further discussion of their role in contrastive explanation, are addressed in the next two sections.

The final presupposition involves the possibility of reducing contrastive explanation to deductive explanation. First, notice that an explanation of "P rather than Q" may be as simple as the citation of one event. In the case of Lewis's trip to Monash, "why Lewis went to Monash rather than to Oxford" can be explained by simply citing the fact that he was invited only to Monash. The explanation of "Why Lewis went to Monash," however, would require a complete description of various events leading to his actual trip to Monash. Similarly, the explanation of "Why Lewis did not go to Oxford" would require a complete description of various events leading to his not going to Oxford. Thus, to explain "P rather than Q" is not the same as to explain "P," to explain "P and not-Q," or to explain "not-Q." In particular, the fourth presupposition is that contrastive explanation offers a *partial* explanation of the fact *in terms of the foil.*

The remainder of this section focuses on the degree to which Lipton's account can explain this presupposition. This presupposition is also discussed in section 4.4. In examining the exact nature of this type of partial explanation, Lipton first considers David Lewis's (1986) account. On Lewis's account, "P rather than Q" is explained by simply citing the event or events that are in the causal history of P but would not have been in the history of Q. Thus, if only Monash invited him to visit, the explanation of why he went to Monash rather than to Oxford is simply that Monash invited him. But as Lipton points out, Lewis's account is too weak. Consider the case in which Lewis is invited by both Monash and Oxford. Then the explanation of why he went to Monash rather than to Oxford could still be provided by the event "Monash invited him" since this event still would not cause him to go to Oxford. Yet, in this case, the invitation to Monash does not explain why he actually went there. Given that the partition of the history of the event "Lewis went to Monash in 1979" consists only of the events "Lewis was invited to Monash" and "Lewis was invited to Oxford," there simply is no explanation of "Why Lewis went to Monash rather than to Oxford" in the case that he is invited to both universities.

Lipton suggests an improvement on Lewis's account that is based on John Stuart Mill's (1904) method of difference. Lipton describes the improvement as follows:

> I propose that, for the causal explanations of events, explanatory contrasts select causes by means of what I will call the "difference condition." To explain why P rather that Q, we must cite a causal difference between P and not-Q consisting of a cause of P and the absence of a corresponding cause in the history of not-Q. Instead of pointing to a counterfactual difference, a particular cause of P that would not have been a cause of Q as Lewis suggests, contrastive questions select as explanatory an actual difference between P and not-Q. (1991a, 693)

Lipton's suggestion certainly works. In the case that Lewis is invited only to Monash, there is no (actual) event "corresponding" to the invitation to Monash in the history of going to Oxford--there is no invitation to Oxford. Similarly, in the case that Lewis is invited to both Monash and Oxford, there is a "corresponding cause": the invitation to Oxford corresponds to the invitation to Monash. Furthermore, attention to the presence or absence of a corresponding cause does appear to serve to identify the underlying "mechanism" of contrastive explanation.

Something is lacking in Lipton's account, however. Lipton does not discuss the requirements on "correspondence" among causes and such discussion is required in order to provide a detailed analysis of the manner in which the difference condition is to be applied. Lipton provides no procedure for applying the difference condition--the very "mechanism" for identifying corresponding causes and for producing explanations. On page 694, for example, Lipton (1991a) states that "the difference condition . . . [requires] only that the cited cause of P finds no corresponding cause of not-Q, where a corresponding cause is something that would bear the same relation to Q as the cause of P bears to P." In so far as Lipton has not specified the constraints on what a corresponding cause is, he has not specified the constraints on explanatory relevance.

This is an important point because the chief criticism of contrastive explanation is that it does not provide definite constraints on explanatory relevance. For example, van Fraassen's (1977 and 1980) theory of contrastive explanation is based explicitly on a relevance relation, R. In our terminology, van Fraassen's relevance relation is defined as follows: a proposition A is called relevant to the explanation of "P rather than Q" exactly if A bears the relation R to the couple <P, X> where X = {P, Q}. X is called the contrast class of the why question and is generally the union of the fact with all of the foils (see van Fraassen 1980, 143). Lipton's relation of being a "corresponding cause" satisfies van Fraassen's definition of a relevance relation. Consequently, Lipton's account is certainly a version of van Fraassen's. But, as Philip Kitcher and Wesley Salmon (1987) have argued, the lack of any definite constraints on the relevance relation R in van Fraassen's theory leads to the fact that, in any situation, almost anything can count as explanatorily relevant. One may argue that this lack of constraints is an advantage since the theory is applicable to a more diverse range of explanatory acts. On the other hand, if one is an advocate of causal explanation, one could argue (with respect to Kitcher and Salmon's criticisms) that Lipton's account improves upon van Fraassen's account in so far as the relation must be causal.

On either account, however, Lipton has correctly emphasized that the distinguishing feature of the contrastive account is that it provides partial explanations of the fact *in terms of the foil.* But the question remains: exactly how does the foil serve to call for one explanation rather than another? Considering Kitcher and Salmon's criticisms of van Fraassen, contrastive theorists must reveal the manner in which consideration of the foil serves to identify explanatorily relevant events. Furthermore, contrastive theorists must demonstrate the manner in which the consideration of the foil provides them with an advantage over non-contrastive approaches to explanation. For Lipton, the question is this: what is the role of the foil in identifying the "corresponding cause" rather than some other cause? For that matter, what exactly is a corresponding cause--doesn't just about anything count as a corresponding cause? Lipton has not provided a mechanism or method by which the foil (or fact-foil pair) serves to identify the corresponding cause (or any other cause) in his account of explanation. The next two sections provide answers to these questions.

4.3 Propensities and the causal history of events

The aim of this section is to demonstrate the manner in which the propensity interpretation can recover and expand upon the presuppositions of Lipton's analysis along with the basis for his difference condition. First, a brief review of the semantics for the propensity interpretation of experiments and their possible outcomes is provided. This review includes a discussion of the manner in which the update semantics can be used to recover causal histories of events. Then the propensity interpretation is used to formalize both the first three presuppositions of Lipton's account as well as his difference condition. Next, the example involving Lewis's trip to Monash is presented in terms of the propensity formalization of the three presuppositions and the difference condition. By appealing to the propensity interpretation of causation in chapter 3, the manner in which the difference condition identifies events in the causal history of the fact is explained and justified. Finally, one new presupposition is made explicit: in order to provide a causal explanation of an event, a detailed causal history of that event must be available.

Propensity statements are taken to be of the following form: **Pr**(A:S) = p which is read as "the propensity for a system S to produce an event A is p." The symbols occurring within the parentheses and before the colon denote propositions concerning the occurrence of events. Symbols occurring within the parentheses but after the colon denote propositions comprising a set of background conditions that describe an experimental setup or system.

Reference is often made directly to the events, background conditions, experimental setups, and systems themselves. A propensity function, '**Pr**', is a probability measure defined over a Boolean σ-algebra $\mathscr{B}$ on a set of events Ω, where each member of Ω is a possible outcome of some experimental setup or system. It is important to note that the propensity function takes both the members of the σ-algebra $\mathscr{B}$ (defined on a set of possible outcomes Ω) and the system S as its domain: a propensity function is defined *over* a Boolean σ-algebra on a set of possible outcomes and *for* a system. The distinction between, and definition of, the set of outcomes in Ω (and consequently the members of the σ-algebra $\mathscr{B}$) and the system S depend primarily on a consideration of the experimental setup.

A system is quite simply the experimental setup as it is arranged prior to the running of the experiment. A system is represented as the statistically or causally relevant events that *actually occur* at some time *before* the experiment begins. The notion of relevance used here is discussed further below. The set of possibie outcomes of an experiment is composed of measurable events that *either occur or do not occur* at some time *after* the experiment begins. Alternatively, the "occurrence" of the system-events is intentionally brought about (or allowed) by the experimenter in order to initiate the experimental process, whereas the occurrence of the outcome-events are spontaneous (once the experiment begins). Given the temporal nature of the distinction between events describing the system and the possible outcomes, and the assumption that systems are described only by relevant events, once the experiment is over, the sequence of "relevant" events leading to the outcomes of the experiment are given in a large part by the event description of the system. The only relevant events not given by the system are those events that occur between the time that the experiment is started and the time that the outcome occurs.

In order to better understand the analysis of experiments and propensities in terms of systems and possible outcomes, consider the application of the propensity interpretation of probability to the example of Lewis's trip to Monash. For simplicity, suppose that Lewis is planning a trip at some time prior to t_1, that he receives invitations to various possible destinations at some time t_2 later than t_1, and that he takes the trip at some time t_3 after t_2. The propensity analysis of this example presupposes that there is some representation of the events and factors (actually taking place prior to t_1) that actually influence Lewis's decision. This assumption is discussed further below. Let 'S_{t1}' be the system representing these events and factors. Then there is a propensity function '$\mathbf{Pr}_{t1}$' defined over the σ-algebra on the set of outcomes 'I_M' = Lewis is invited to Monash, 'I_O' = Lewis is invited to Oxford, 'M' = Lewis travels to Monash, and 'O' = Lewis travels to Oxford. Figure 4.1 provides a diagram representing this system (presented in outlined font) and its possible outcomes (presented in normal font).

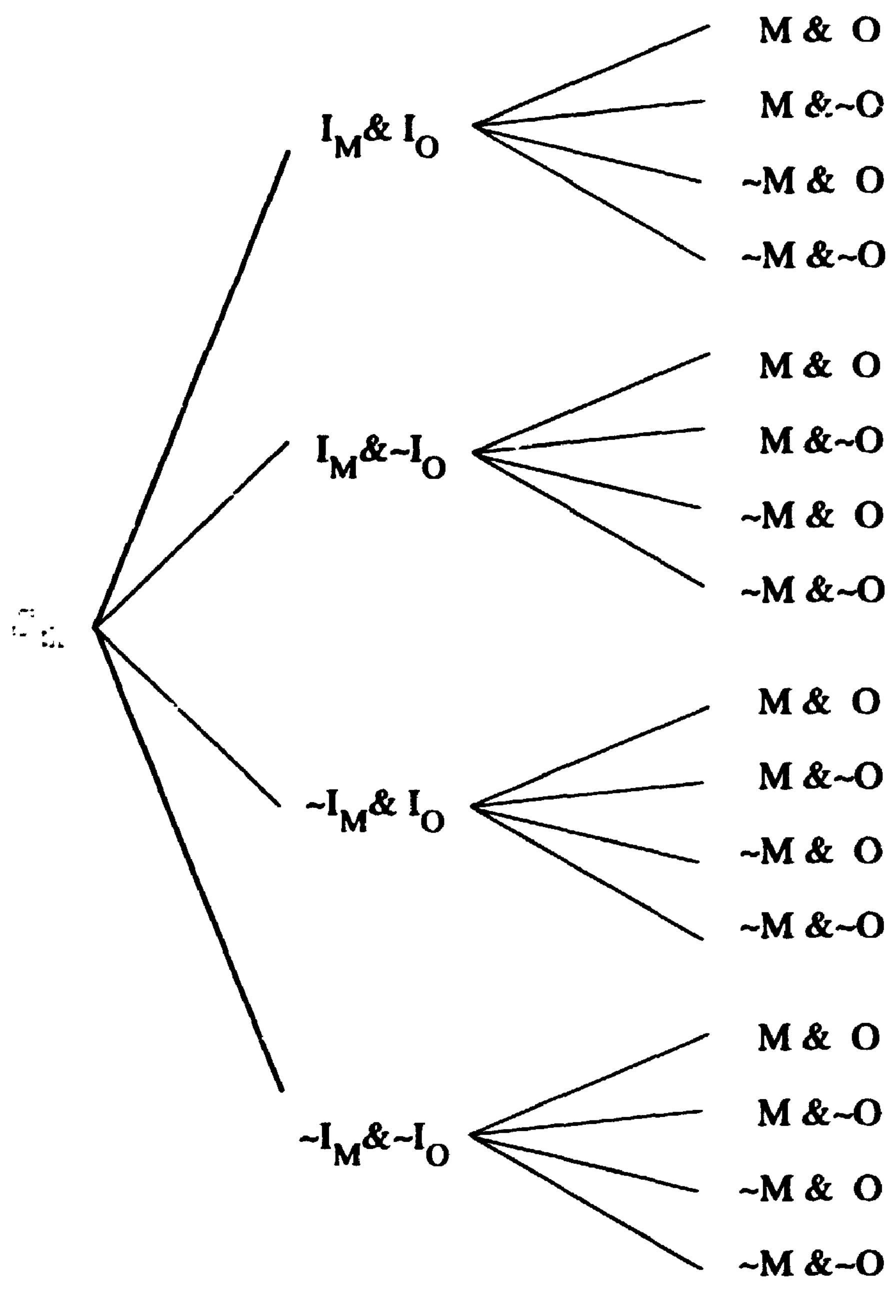

Figure 4.1. Lewis is invited either to Monash or to Oxford.

Given this propensity representation of the state of affairs concerning the system S_{t1} and its ability to produce (or result in) the event that Lewis travels to Monash, the values of various propensities can be calculated as they exist at t_1. Of course, systems change over time as the events that they produce actually occur or do not occur. Consequently, the propensities that systems possess, and the values of those propensities, also change over time. In the system described above, if the event I_M occurs and I_O does not occur at t_2 then it is possible to update the dispositional nature of that system to reflect the dispositional nature of the "new" system as it exists at t_2. This updating procedure creates a new σ-algebra (defined on a new outcome space) since the events I_M and I_O no longer lie in the future of the system. Of the events in the original outcome space, only the events M and O remain in the new outcome space. This updating procedure also requires the definition of a new propensity function. This new function is defined over the new σ-algebra described above and for the new system as it exists at t_2. The system itself may have to be re-defined. In particular, if the event I_M is relevant to the events in the new outcome space, then the fact that the event I_M occurred is incorporated into the description of this new system. But, if the event I_M is not relevant to the events in the new outcome space, then I_M is not incorporated into the description of the system. The reason for the different treatments of events is that the system consists only of those (actual) events that are causally relevant to those events that remain in the outcome space.

For present purposes, it is assumed that all of the events discussed here are relevant to future events. Thus, in the case that I_M occurs at t_2 and I_O does not, the description of the new updated system at t_2 consists of the conditions expressed in S_{t1} and the fact that the event I_M occurred at t_2 and I_O did not occur at t_2. Figure 4.2 provides a representation of this new system

(S_{t1}--I_M&~I_O) and its possible outcomes (M and O) in terms of previously adopted convention of placing the system events (or actual events) in outline font and placing the remaining possible outcome events in normal font.

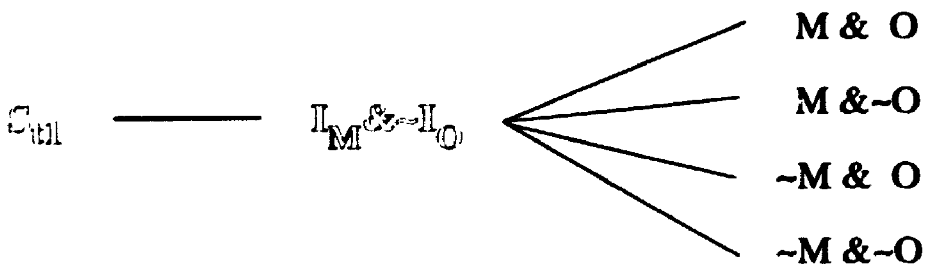

Figure 4.2. Lewis is invited to Monash but not to Oxford.

If the event I_O occurs and the event I_M does not occur at t_2 then a different propensity function is defined for the system that exists at t_2. The propensities defined by this function are conditioned on a set of background conditions consisting of the conditions expressed in S_{t1}, also the fact that the event I_O occurred at t_2, since it is assumed that I_O is causally relevant to the remaining future events. If both I_M and I_O occur at t_2 then still another function and system are defined, and the resulting system consists of the original description plus the fact that both I_M and I_O occurred at t_2. Following the previous convention of outlining system events (the actual events that are causally relevant) and not outlining the remaining possible outcomes, these two new systems and their possible outcomes are represented in figures 4.3 and 4.4, respectively.

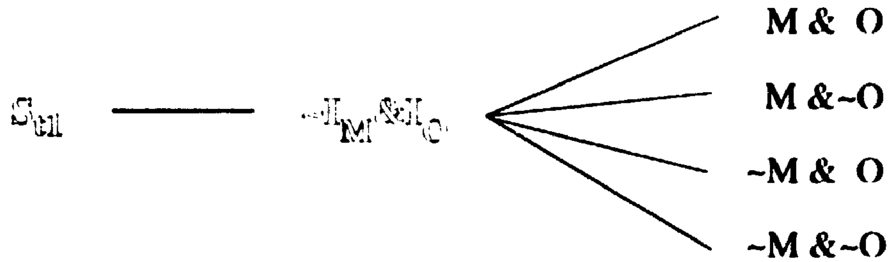

Figure 4.3. Lewis is invited to Oxford but not to Monash.

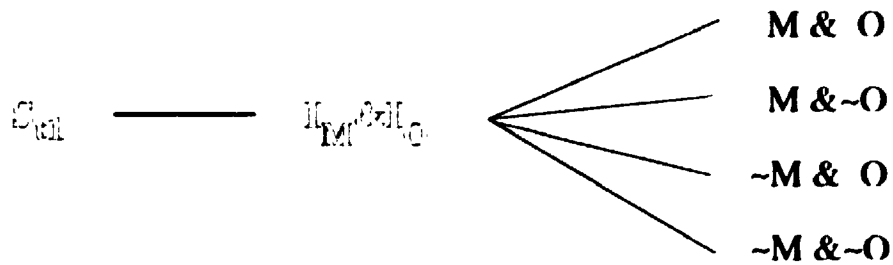

Figure 4.4. Lewis is invited both to Monash and to Oxford.

Each of the "new" systems are defined at time t_2, and at time t_2 the events I_M and I_O have either occurred or did not occur. At time t_2 the events I_M and I_O do not lie in the future of the new systems. As a result, the events I_M and I_O are not members of the outcome space for the new propensity functions (defined for the new systems) since these events cannot be "produced" by the new systems. Consequently, as a system evolves over time, events are essentially removed from the outcome space and those *relevant events that actually occur* are incorporated into the background conditions of the emerging system. Depending on which events actually occur, this creates a new propensity function that is defined at a different time, for a different system, and over a different σ-algebra of events. Thus, a propensity analysis of the *actual evolution* of a particular system provides the causal history of the most recent event to be incorporated into the system.

Given this analysis of the actual evolution of systems, the application of the propensity interpretation of probability to contrastive explanation is quite straightforward. The first two presuppositions of Lipton's account can be expressed as follows: for some appropriate propensity functions defined for some systems S_{t1}, S_{t2}, and S_{t3}, where $t_1 < t_2 \leq t_3$, and defined over the fact P and the foil Q,

$$1 > \mathbf{Pr}_{t1}(P: S_{t1}) > 0 \text{ and } 1 > \mathbf{Pr}_{t1}(Q: S_{t1}) > 0, \qquad (4.1)$$

$$\mathbf{Pr}_{t2}(P: S_{t2}) > 0 \text{ and } \mathbf{Pr}_{t2}(Q: S_{t2}) = 0, \text{ and} \qquad (4.2)$$

$$\mathbf{Pr}_{t3}(P: S_{t3}) = 1. \qquad (4.3)$$

Lipton's first presupposition is provided by condition (4.1): according to the propensity function $\mathbf{Pr}_{t1}$, the fact P and the foil Q are both possible at some time t_1. Lipton's second presupposition is provided by conditions (4.2) and (4.3). In particular, the propensity function $\mathbf{Pr}_{t2}$ of condition (4.2) states that the foil Q is no longer possible at some later time t_2, and condition (4.3) states that, according to propensity function $\mathbf{Pr}_{t3}$ the fact P actually occurred at some time t_3 (possibly simultaneous with t_2).

Lipton's third presupposition, as well as the basis for his difference condition, is inferred from conditions (4.1), (4.2), and (4.3) and the following propensity version of the difference condition (**DC**):

DC if system S_3 evolved from system S_2, and S_2 evolved from system S_1, then the explanation of "Why P rather than Q" consists of the events that provide the *actual* difference between system S_1, and its future events (up to time t_2), and system S_2.

Lipton's third presupposition follows immediately: given that system S_2 evolved from system S_1, the difference between system S_1, and its future events (up to time t_2), and system S_2 provides information concerning the causal history of the event P. In the next section, it is argued that the difference between these two systems is the event or events that caused "P rather than Q," and that this difference is the basis for identifying a "corresponding cause." But, before examining the role of the difference condition in these matters, the manner in which the difference condition identifies the event or events in the actual causal history of the fact P must be considered.

In order to examine the manner in which the difference condition identifies the event or events in the actual causal history of the fact P reconsider the example of explaining Lewis's trip to Monash. In the case that Lewis is invited to Monash but not to Oxford, conditions (4.1), (4.2), and (4.3) are satisfied by systems S_{t1} and S_{t1}--I_M&~I_O; moreover, system S_{t1}--I_M&~I_O evolved from system S_{t1}. Recall that as relevant events actually occur, they are incorporated into the system description. This process of incorporation is represented, diagrammatically, by placing the events that actually occur into the outline font that is used to identify systems and actual events. Consequently, the difference between the outlined portions of figures 4.1 and 4.2 (above) provides the actual difference between the causal histories of the occurrence of the fact (M) and the nonoccurrence of the foil (~O). In comparing figures

4.1 and 4.2 (above), the only *actual* difference between these two systems is the event I_M. In the case that Lewis is invited to both Monash and Oxford, however, there is no actual difference between the causal history of the occurrence of the fact and the nonoccurrence of the foil (compare figures 4.1 and 4.4 above).

Thus, the basis for Lipton's difference condition is evident. The propensity account identifies an actual difference in the causal histories of the events P and ~Q where such a difference exists. Furthermore, the application of the difference condition, and the manner in which it identifies the difference in the causal history, does not make an appeal to hypothetical or counterfactual events. It must be noted, however, that in order to avoid such hypothetical or counterfactual claims, this account presupposes a detailed (propensity) account of the actual causal history of the occurrence of the fact and the nonoccurrence of the foil. Given this requirement, two additional observations must be made. First, that this is a reasonable presupposition because it would be unreasonable to require that one can provide a causal explanation of an event without knowing the causal history of that event. If one does not know the causal history of an event, then he or she cannot possibly give a causal explanation of that event. The second observation is that the establishment of the knowledge of the causal history of the fact may include an appeal to multiple experiments and sequences of events rather than to the single actual sequence of events resulting in the fact P. In particular, given that a detailed causal history of the fact is required, the methods of causal analysis must be used. Furthermore, as discussed in chapter 3, the methods of causal analysis typically make an appeal to controlled experiments. Thus, the requirement of a detailed causal history of the fact P may require an appeal to causal sequences that do not involve the *actual* fact P. Still, however, as discussed in chapter 3, these appeals to other causal sequences do not necessarily require an appeal to hypothetical or counterfactual occurrences of events.

In conclusion, the propensity account adequately captures Lipton's first three presuppositions for contrastive explanation. Furthermore, the use of propensities to represent the actual causal histories of events provides the basis for the application of the difference condition. A complete examination of the difference condition has not been given, however. Also, Lipton's fourth presupposition, that contrastive explanation offers a partial explanation of the fact in terms of the foil, has not been discussed in detail. Consequently, the following question remains: how does the identification of the actual difference between P and ~Q identify the corresponding cause, and in what manner does it provide a partial explanation of the fact in terms of the foil? The next section considers the answer to this question.

4.4 The role of the foil in contrastive explanation

The aim of this section is to demonstrate the role of the foil in providing a partial explanation of the fact in terms of the foil. Ultimately, it is the role of the foil that provides the basis for describing the mechanism by which the difference condition identifies a *corresponding* cause rather than some other cause. First, an outline of the manner in which a change in the foil produces a change in the propensity analysis of the causal history of the fact is given. It is argued that the foil selects partial causal histories of the fact, and that the application of the difference condition to these partial causal histories produces the partial explanation. This selection procedure serves to identify causal histories based on a particular type of causal relevance. Moreover, the notion of corresponding cause is defined in terms of being the type of cause that is selected by the foil. Finally, the implications of this notion and the advantages of contrastive explanation are examined. First it is argued that the foil serves to narrow the reference class and to reduce the problem of encountering inhomogeneity. Consequently, the foil also serves to reduce the problem of explanatory ambiguity.

In order to appreciate the role of the foil in contrastive explanation, consider the task of answering the following contrastive why question: why did Lewis fly to Monash rather than sail to Monash in 1979? According to the first three presuppositions, it is assumed that (1) at some time it was possible for Lewis to fly to Monash and it was possible for him to sail to Monash; (2) at some later time Lewis actually did fly to Monash (and did not sail to Monash); and (3) the explanation of why Lewis flew to Monash rather than sailed to Monash consists of information concerning the causal history of his flight to Monash. Furthermore, according to the additional assumption that a complete causal history of Lewis's actual flight to Monash is available, suppose that the causal history is provided as follows. At some time after accepting the invitation to Monash in 1979, Lewis was confronted with the option of flying to Monash or sailing to Monash. Additionally, the only factor left to influence his decision to fly to Monash or to sail to Monash was the price of each option. That is, all causal factors for his flying or sailing to Monash have been fixed at some time t_1, and can be represented by some system S'_{t1}. For example, factors such as invitations, getting time-off for the trip, the length of a flight versus length of a voyage, and the prospect of seeing friends have all been captured in the description of system S'_{t1}. Finally, assume that, given the manner in which these factors are causally relevant, the reasonableness of the prices of flying to Monash and of sailing to Monash are causally relevant.

Define the events P_F = "The price of flying is reasonable"; P_S = "The price of sailing is reasonable"; F = "Lewis flew to Monash"; and S = "Lewis sailed to Monash." Then, given this causal description of the states of affairs at t_1, and assuming that it is not known which prices are reasonable, the causal process leading to Lewis's flight or voyage to Monash in 1979 is represented in figure 4.5.

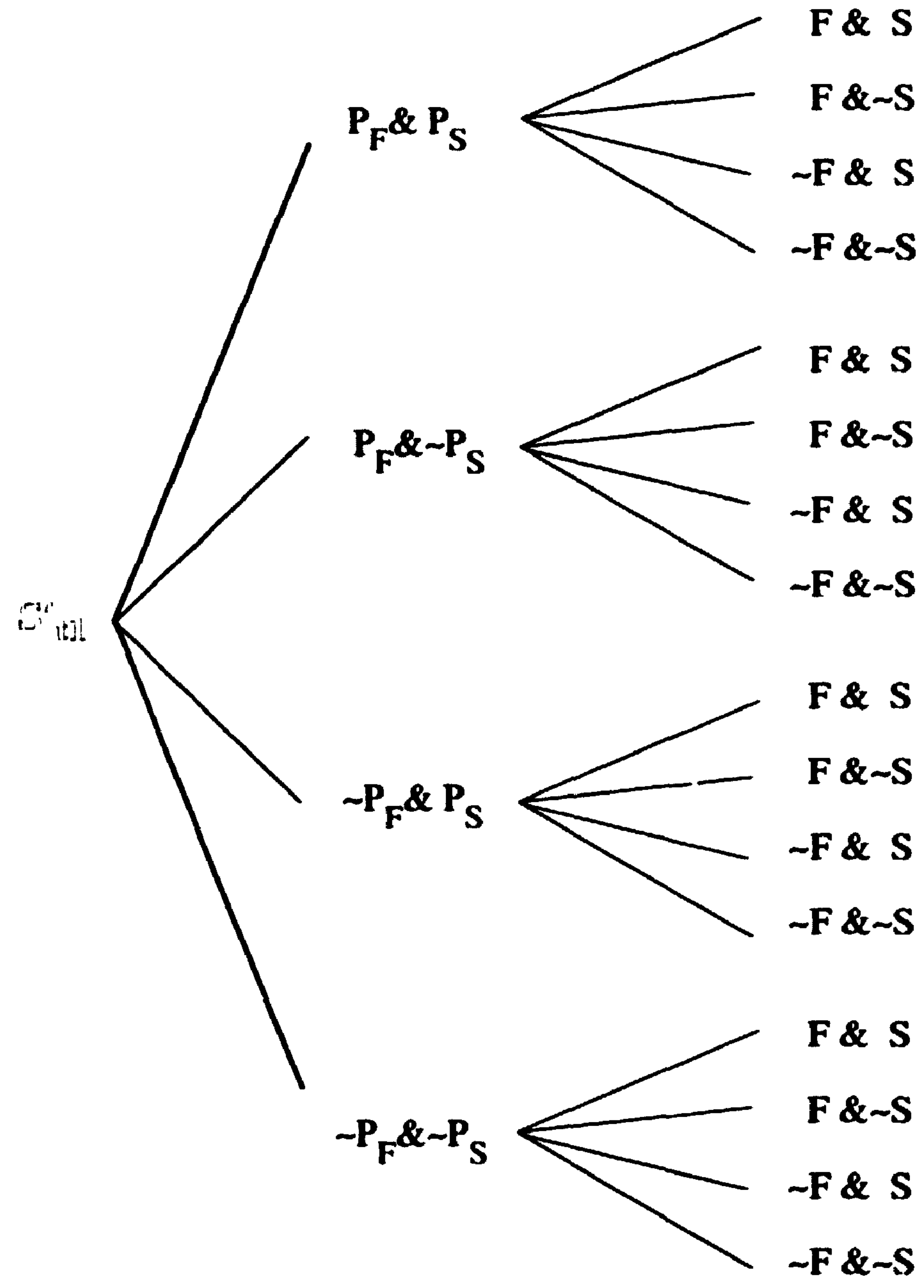

Figure 4.5.

Either the price of flying is reasonable or the price of sailing is reasonable.

If it is actually the case that the price of flying is reasonable and the price of sailing is not reasonable, the causal history of Lewis's flight or voyage to

Monash is represented by figure 4.6.

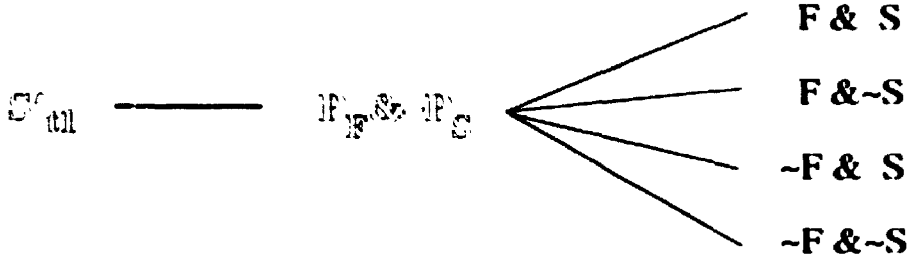

Figure 4.6. The price of flying is reasonable but the price of sailing is not.

On the other hand, if it is actually the case that the prices of both modes of transportation are reasonable, the causal history of Lewis's flight or voyage to Monash is represented by figure 4.7.

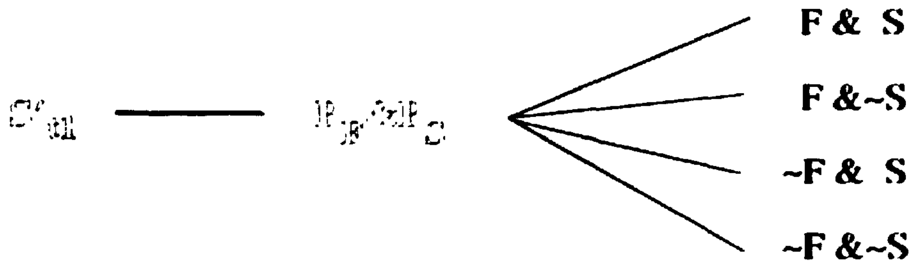

Figure 4.7. Both the price of flying and the price of sailing is reasonable.

Given this causal story describing Lewis's flight or voyage to Monash in 1979, it is possible to apply the difference condition to Lewis's flight (rather than voyage) to Monash in the same manner as it was applied to the previous causal story about Lewis's travel to Monash (rather than Oxford) in 1979. In the case that the price of flying is reasonable and the price of sailing is not reasonable (as represented by figure 4.6), the difference condition identifies the actual difference between figures 4.5 and 4.6. This application of the difference condition reveals that the actual difference is the event "the price of flying is reasonable and the price of sailing is not." Consequently, the explanation of why Lewis flew to Monash rather than sailed to Monash is the fact that the price of flying was reasonable and the price of sailing was not. But in the case that both modes of transportation are reasonably priced (represented by figure 4.7), the difference condition identifies the difference between figures 4.5 and 4.7. In this case there is no actual difference between the two representations of the causal history. Thus, there is no explanation of why Lewis flew, rather than sailed, to Monash.

Clearly, the role of the foil in contrastive explanation is to help identify the causal story to which the difference condition must be applied. Suppose that, in fact, Lewis flew (rather than sailed) to Monash (rather than Oxford) in 1979. Also, assume that Lewis was invited to Monash, but not to Oxford, and that the price of flying was reasonable, whereas the price of sailing was not reasonable. Then, the complete causal history of Lewis's trip to Monash would consider a vast array of events including invitations to Monash and to Oxford, as well as the price of flying and the price of sailing. But the explanation of why Lewis went to Monash rather than to Oxford would consider only a *partial* causal history of Lewis's trip to Monash. As described above, this explanation is the result of applying the difference condition to the causal history concerned with invitations, and the explanation would be that only Monash invited him. In a similar manner, the explanation of why Lewis flew rather than sailed to Monash would consider the partial causal history concerned with the price of travel, and the explanation would be that only the price of flying was reasonable.

The only difference between the two contrastive why questions is the foil, and the foil serves to identify a particular partial history of the fact. In the case of the two contrastive why questions considered above, the foils identify different (partial) causal histories. In answering each contrastive why question, the difference condition is applied in exactly the same manner: the difference condition is applied to the partial causal history that is identified by the foil in the contrastive why question. Consequently, in answering the two contrastive why questions, the difference condition is applied to two different (partial) causal histories. In so far as the explanation consists of some event or events taken from an appropriate partial causal history, the resulting explanation provides a *partial* explanation of the fact. Furthermore, since the explanation is taken from one of the actual partial causal histories of the fact P (and the nonoccurrence of the foil Q) and does not consist of a complete causal history of the fact or the nonoccurrence of the foil, explaining "P rather than Q" does not reduce to explaining "P," explaining "P and ~Q," or explaining "~Q."

The following question remains: how does the selection of the foil determine the partial causal history to use? The answer to this question provides the basis for our understanding of the manner in which a contrastive explanation provides a partial explanation of the fact *in terms of the foil*. When confronted with a contrastive why question, the propensity function that is defined for the system describing the relevant causal history must be defined over an outcome space consisting of the fact and the foil (and their negations). Furthermore, if the outcome space is restricted to the fact and the foil (and their negations), then different contrastive why questions yield different outcome spaces. Given the description of the manner in which causal histories are constructed, different outcome spaces produce different causal histories since the causal history includes only those events that are causally relevant to the fact and the foil.

The result of this restriction is that different contrastive why questions request explanations in terms of different "types" of causal factors. For example, the explanation of why Lewis traveled to Monash rather than to Oxford in 1979 proceeds in terms of those causal factors that are responsible for Lewis travelling to Monash and not travelling to Oxford, whereas the explanation of why Lewis flew to Monash rather than sailed to Monash proceeds in terms of those causal factors responsible for Lewis flying to Monash and not sailing to Monash. This difference provides the basis for claiming that the resulting explanation provides a partial explanation of the fact *in terms of the foil.*

Of course, some events may be causally relevant to more than one foil and may therefore have a role in more than one partial causal history of the fact. For example, Lewis may have been invited to both Monash and Oxford, but may have made his decision based on the reasonableness of the prices of flying to Oxford and flying to Monash. Consequently, "the price of flying to Monash is reasonable" is an event in both the causal history of why Lewis travelled to Monash rather than to Oxford and the causal history of why Lewis flew to Monash rather than sailed to Monash. Despite this fact, the "type" of causal relevance is determined by, and expressed in terms of, the foil. Thus the notion of a corresponding cause can be understood as follows. An event A corresponds to an event B if and only if A and B are causally relevant to the fact P in the same way. In particular, A and B are causally relevant to both the fact P and the foil Q. Consequently, the notion of a corresponding cause is dependent on what is *actually* causally relevant to both the fact and the foil, and to those events that are necessary to produce the fact and the foil.

In conclusion, this section will examine the degree to which the propensity interpretation of contrastive explanation encounters or avoids the traditional problems of the homogeneity of the reference class and of explanatory ambiguity. The problem of homogeneity arises in different manners for different theories of explanation and has been discussed from many perspectives and by many authors.[2] The propensity account of contrastive explanation , for the most part, avoids the problem of homogeneity. One of the major presuppositions that is made by the propensity version of contrastive explanation is that the actual (partial) causal history of the fact is the broadest homogeneous description available. As described above, in the present account there is an explicit assumption that such a reference class is

---

[2]See for example, J. Alberto Coffa 1974; James H. Fetzer 1974a, 1974b, and 1977; Joseph F. Hanna 1981 and 1983; Carl G. Hempel 1962, 1965, and 1968; Peter Railton 1978 and 1981; and Wesley C. Salmon 1971, 1974, 1977, and 1984.

available. In particular, this account of explanation presupposes that a theory of causation is available to produce these "broadest homogeneous causal histories." Once an appropriate causal history is acquired, the contrastive account outlined above simply applies the difference condition **DC** to the appropriate causal histories.

It must be acknowledged that the requirement that the reference class is provided by a theory of causation, as a strategy for avoiding the problem of homogeneity, is available to other (non-contrastive) approaches to explanation. Yet, it must also be realized that the contrastive account has a distinct advantage in using this strategy because the types of causal histories demanded by the contrastive approach are more easily provided by a theory of causation than are the types of causal histories demanded by non-contrastive accounts of explanation. In most non-contrastive accounts of explanation, the reference class must take the form of a complete causal history of the fact. But, as discussed in chapter 2, a causal analysis of the foil *simpliciter* employs chance-like thinking and encounters the problem of homogeneity directly. On the other hand, the contrastive approach to explanation requires an analysis of a partial causal history, and this requirement effectively narrows the reference class. Most importantly, the reference class is narrowed by restricting the causal analysis to only those events that are actually causally relevant to both the fact and the foil. Thus, the reference class is narrowed along causal lines. For example, the events effecting Lewis's decision to fly to Monash rather than Oxford in 1979 are less diverse than the events effecting his decision to fly to Monash in 1979. Consequently, the manner in which the foil serves to narrow the reference class helps to make the reference class that is required more homogeneous. Thus, the contrastive approach to explanation facilitates the theory of causation's task of defining the broadest homogeneous reference class.

This discussion of the manner in which the contrastive approach avoids the problem of homogeneity provides the basis for explaining the manner in which the contrastive approach avoids the problem of explanatory ambiguity. If an account of explanation leaves the task of producing the broadest homogeneous reference class to a theory of causation, then the account of explanation must be able to provide that causal theory with a specific and unambiguous event that is, in fact, susceptible to a unique and homogeneous causal history. This problem has been described by Joseph F. Hanna (1981, 413) as "the ambiguity of event descriptions." As Hanna notes, "the request for an explanation is itself ambiguous, because there is no unequivocal fact . . . has been specified" (1981, 413). It will be argued, below, that non-contrastive approaches often appeal to pragmatic methods of avoiding this problem, whereas the contrastive approach to explanation provides a direct and formal method of overcoming this problem.

Most non-contrastive approaches to explanation presuppose that a complete and homogeneous reference class or causal story is available. Consequently, if a causal theory that can, in fact, provide them with such a causal history (in spite of the difficulties with homogeneity), then these theories of explanation are provided with a large array of contributing and counteracting causes that can be extraordinarily diverse in nature.

Subsequently, non-contrastive theories of explanation encounter the overwhelming task of selecting a relevant event, or events, from that large array of contributing and counteracting causes. Often, this task is achieved by imposing pragmatic criteria for explanatory relevance or salience to this complete causal history.[3] In the case of contrastive explanation, as described above, the answers to contrastive why questions provide *partial* explanations of the fact *in terms of the foil*. This type of explanation is based on, or presupposes, a partial causal history of the fact, and this type of causal history is more easily partitioned into a homogeneous set of causal factors. Furthermore, the foil serves to identify the partial causal history, and by applying the difference condition to this causal history, a specific event or group of events is selected in the case that their is no corresponding cause. Consequently, the foil serves to remove the ambiguity associated with singular (complete) events. The advantage of the contrastive account is that it provides a method for focusing the search on a specific part of the complete causal story, or alternatively, it provides a method for narrowing the reference class. This advantage serves to facilitate that application of causal theory, as well as the application of the difference condition.

4.5 Concluding remarks

The contrastive approach to explanation is based on providing *partial explanations of the fact in terms of the foil*. Moreover, this type of explanation

---

[3]See for example Carl G. Hempel 1965 and Peter Railton 1978 and 1981.

certainly differs from non-contrastive explanations--explanations of the fact *simpliciter*. One of the motivations for this chapter is that advocates of the contrastive approach to explanation are still struggling to explain this difference, as well as the advantage that this difference confers on the contrastive approach to explanation. Lipton's (1991a and 1991b) account has made an effort to explain these differences, but, as indicated above, Lipton's account fails to demonstrate the manner in which the difference condition identifies a corresponding cause rather than some other type of cause. Based on this inadequacy, this chapter examines both the role of the foil in identifying corresponding causes, and the advantages of contrastive explanation.

First, it was demonstrated, by appealing to the distinction between systems and events, as well as to the update semantics, that the propensity interpretation of probability is capable of providing a representation of the actual causal history of an event. Then it was demonstrated that the propensity interpretation is capable of formalizing the four presuppositions of (causal) contrastive explanation: (1) at some time, both the fact and the foil were possible; (2) at some later time, the fact actually occurred and the foil did not occur; (3) the explanation of an event consists of providing information concerning the foil's actual causal history; and (4) "explaining P rather than Q" is not reducible to "explaining P," "explaining P&~Q," or "explaining ~Q." Furthermore, it was argued that causal theories of explanation must also assume that an analysis of the actual causal history of the fact is available prior to the explanation. Finally, the propensity representation of the causal histories of events was used to provide an understanding of the difference condition. Most importantly, the role of the foil in contrastive explanation was examined. It was demonstrated that the foil serves to identify partial causal histories of the fact in terms of the foil, and that it was the application of the difference condition to these partial causal histories that made it possible to identify corresponding causes rather than some other cause. Moreover, it is the application of the difference condition to the causal histories selected by the foil that produced *partial* explanations of the fact *in terms of the foil*.

Finally, the advantages of the contrastive approach to explanation were demonstrated. Theories of explanation based on explaining the fact *simpliciter* encounter the problem of homogeneity in the same manner as the theories of chance and causation encounter the problem. Specifically, by considering events *simpliciter*, these theories must consider a diverse and vast array of causal factors. The contrastive approach avoids the problem of homogeneity by passing it on to a theory of causation, but it passes this problem on in a beneficial and responsible manner. The foils serves to narrow the reference class by requesting a causal history that is defined according to a specific type of causal relevance. Thus, there is a relative lack of diversity within the causal histories that are requested by the contrastive approach to explanation and the causal analysis of these causal histories has a better chance to be homogeneous. Furthermore, given that the foil selects the partial causal histories that it does, and given that the difference condition is applied to these causal histories, the problem of explanatory ambiguity is avoided with the normal application of the formalism of the theory. In the case of the non-contrastive approach, however, an appeal to pragmatic considerations is usually required in order to avoid this problem. Thus, the propensity version of the theory of contrastive explanation provides an account that is capable of avoiding or minimizing the effects of both the problem of homogeneity and the problem ambiguity. Furthermore, the propensity version is unified with, and facilitates the use of, theories of causation and depends only on the difference condition, thereby avoiding an appeal to the pragmatics of explanation.

Appendix 1

Derivations Involving Probabilities and Propensities

A.1 A demonstration that $\mathbf{Pr}_{t1}(T_{t3}|B_{t1}) = pq$

Note the following theorems of the probability calculus. All page references are to Humphreys 1985.

The theorem of additivity for conditional probabilities (p. 560):

Add If A and B are disjoint, then $P(A \vee B|C) = P(A|C) + P(B|C)$

The theorem of total probability for conditional probabilities (p. 560):

TP $P(A|C) = P(A|BC)P(B|C) + P(A|{\sim}BC)P({\sim}B|C)$

The demonstration that $\mathbf{Pr}_{t1}(T_{t3}|B_{t1}) = pq$ proceeds as follows:

(1) $\mathbf{Pr}_{t1}(T_{t3}|I_{t2}B_{t1}) = p$ (p. 561, assignment i)

(2) $\mathbf{Pr}_{t1}(I_{t2}|B_{t1}) = q$ (p. 561, assignment ii)

(3) $\mathbf{Pr}_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$ (p. 561, assignment iii)

(4) $\mathbf{Pr}_{t1}({\sim}I_{t2}|B_{t1}) = 1 - q$ (Add; 2)

(5) $\mathbf{Pr}_{t1}(T_{t3}|B_{t1}) = pq$ (TP; 1, 2, 3, 4)

QED

A.2 A demonstration that $Pr_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = (q - pq)/(1 - pq)$

Recall the following theorems of the probability calculus. All page references are to Humphreys 1985.

The theorem of additivity for conditional probabilities (p. 560):

Add If A and B are disjoint, then $P(A \vee B|C) = P(A|C) + P(B|C)$

The theorem of total probability for conditional probabilities (p. 560):

TP $P(A|C) = P(A|BC)P(B|C) + P(A|{\sim}BC)P({\sim}B|C)$

The demonstration that $Pr_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = (q - pq)/(1 - pq)$ proceeds as follows:

(1) $Pr_{t1}(T_{t3}|I_{t2}B_{t1}) = p$ (p. 561, assignment i)

(2) $Pr_{t1}(I_{t2}|B_{t1}) = q$ (p. 561, assignment ii)

(3) $Pr_{t1}(T_{t3}|{\sim}I_{t2}B_{t1}) = 0$ (p. 561, assignment iii)

(4) $Pr_{t1}({\sim}I_{t2}|B_{t1}) = 1 - q$ (Add; 2)

(5) $Pr_{t1}(T_{t3}|B_{t1}) = pq$ (TP; 1, 2, 3, 4)

(6) $Pr_{t1}(I_{t2}|T_{t3}B_{t1}) = 1$ (demonstrated in ch. 2, sec. 2.3)

(7) $Pr_{t1}({\sim}T_{t3}|B_{t1}) = 1 - pq$ (Add; 5)

(9) $Pr_{t1}(I_{t2}|{\sim}T_{t3}B_{t1}) = (q - pq)/(1 - pq)$ (TP; 2, 6, 5, 7)

QED

A.3 A derivation of the theorem on total probability for binary events, without an appeal to the inversion theorems

Note the following definition, axioms, and theorems of the probability calculus:

The definition of conditional probabilities:

Def $P(A|B) = P(A\&B)/Pr(B)$

The general addition axiom:

Add If A and B are disjoint, then $P(A \vee B|C) = P(A|C) + P(B|C)$

The theorem for logically equivalent events:

Equiv If A and B are logically equivalent, then $P(A) = P(B)$

The theorem for tautologies:

Taut If A is a tautology, then $P(A|B) = 1$

The theorem of total probability for conditional probabilities:

TP $P(A|C) = P(A|BC)P(B|C) + P(A|{\sim}BC)P({\sim}B|C)$

The derivation of theorem of total probability for conditional probabilities proceeds as follows :

(1) $P(B \vee {\sim}B|AC) = P(B|AC) + P({\sim}B|AC)$ (Add)

(2) $1 = P(B|AC) + P({\sim}B|AC)$ (1; Taut)

(3) $P(AC) = P(AC)P(B|AC) + P(AC)P({\sim}B|AC)$ (2; multiply by $P(AC)$)

(4) $P(AC) = P(BAC) + P({\sim}BAC)$ (3; Def)

(5) $P(AC)/P(C) = P(BAC)/P(C) + P({\sim}BAC)/P(C)$ (4; divide by $P(C)$)

(6) $P(A|C) = P(BAC)/P(C) + P({\sim}BAC)/P(C)$ (5; Def)

$$(7)\quad P(AC) = [P(BAC)/P(C)][P(BC)/P(BC)] + [P(\sim BAC)/P(C)][P(\sim BC)/P(\sim BC)] \qquad (6;\ A/A = 1)$$

$$(8)\quad P(AC) = [P(BAC)/P(BC)][P(BC)/P(C)] + [P(\sim BAC)/P(\sim BC)][P(\sim BC)/P(C)] \qquad (7;\ (A/B)(C/D) = (A/D)(C/D))$$

$$(9)\quad P(AC) = [P(ABC)/P(BC)][P(BC)/P(C)] + [P(A\sim BC)/P(\sim BC)][P(\sim BC)/P(C)] \qquad (8;\ Equiv)$$

$$(10)\quad P(A|C) = P(A|BC)P(B|C) + P(A|\sim BC)P(\sim B|C) \qquad (9;\ Def)$$

QED

A.4 A derivation of alternative expressions of the principle of strict composition

Recall that only those emissions that hit the screen are statistically significant. Also, note the following definition and theorems for propensities.

Only those emissions that hit the screen are statistically significant:

SS $\quad \mathbf{Pr}(P_1 \vee P_2{:}B'_3) = 1$

The principle of strict composition, as expressed by (3.3) in ch. 3, sec. 3.2:

SC $\quad \mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R\&P_1{:}B'_3) + \mathbf{Pr}(R\&P_2{:}B'_3)$

Definition of conditional propensities:

DCP $\quad \mathbf{Pr}(A|B{:}S) = \mathbf{Pr}(A\&B{:}S)/\mathbf{Pr}(B{:}S)$

The theorem for tautologies for conditional propensities:

TCP $\quad$ If $\mathbf{Pr}(A \vee B{:}S) = 1$ then $\mathbf{Pr}(C|A \vee B{:}S) = \mathbf{Pr}(C{:}S)$

The derivation of alternative expressions of the principle of strict composition proceeds as follows:

(1) $\mathbf{Pr}(P_1 \vee P_2{:}B'_3) = 1$ (SS)

(2) $\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R\&P_1{:}B'_3) + \mathbf{Pr}(R\&P_2{:}B'_3)$ (SC)

(3) $\mathbf{Pr}(R{:}B'_3) = \mathbf{Pr}(R|P_1{:}B'_3)\mathbf{Pr}(P_1{:}B'_3) + \mathbf{Pr}(R|P_2{:}B'_3)\mathbf{Pr}(P_2{:}B'_3)$ (DCP; 2)

(4) $\mathbf{Pr}(R|P_1 \vee P_2{:}B'_3) = \mathbf{Pr}(R|P_1{:}B'_3)\mathbf{Pr}(P_1{:}B'_3)$
$\quad + \mathbf{Pr}(R|P_2{:}B'_3)\mathbf{Pr}(P_2{:}B'_3)$ (TCP; 1, 3)

Therefore, (2), (3), and (4) provide equivalent statements of the principle of strict composition.

QED

A.5 A derivation of the principle of strict composition from the fact that bullets are localized at the slits

Recall that the fact that bullets are localized at the slits is represented by equality (3.4). Note that references are made to the results of A.4 above. Also note the following theorem of the propensity calculus:

Theorem of composition of conditional propensities:

CCP $\quad Pr(C|A \vee B:S) = Pr(C|A:S)Pr(A:S) + Pr(C|B:S)Pr(B:S) - Pr(C|A\&B:S)Pr(A\&B:S)$

The derivation of the principle of strict composition proceeds as follows:

(1) $Pr(P_1\&P_2:B'_3) = 0$ (3.4)

(2) $Pr(R|P_1 \vee P_2:B'_3) = Pr(R|P_1:B'_3)Pr(P_1:B'_3) + Pr(R|P_2:B'_3)Pr(P_2:B'_3) - Pr(R|P_1\&P2:B'_3)Pr(P_1\&P_2:B'_3)$ (CCP)

(3) $Pr(R|P_1 \vee P_2:B'_3) = Pr(R|P_1:B'_3)Pr(P_1:B'_3) + Pr(R|P_2:B'_3)Pr(P_2:B'_3)$ (1, 2)

(4) $Pr(R:B'_3) = Pr(R\&P_1:B'_3) + Pr(R\&P_2:B'_3)$ (results of A.4; 3)

QED

Note that this derivation is not "reversible." That is, strict composition implies that $Pr(R|P_1\&P2:B'_3)Pr(P_1\&P_2:B'_3) = 0$. This condition can be satisfied if either $Pr(R|P_1\&P2:B'_3) = 0$ or $Pr(P_1\&P_2:B'_3) = 0$.

Bibliography

Accardi, Luigi. 1984. "The Probabilistic Roots of Quantum Mechanical Paradoxes." In *The Wave-Particle Dualism: a Tribute to Louis de Broglie on his 90 th Birthday*, proceedings of an international symposium held in Perugia April 22-30, 1982, edited by S. Diner, D. Fargue, G. Lochak, and F. Selleri, 297-330. Fundamental Theories in Physics edited by Alwyn van der Merwe. Dordretch: D. Reidel Publishing Company.

Coffa, J. Alberto. 1974. "Hempel's Ambiguity." *Synthese* 28: 19-26.

Fetzer, James H. 1974a. "A Single Case Propensity Theory of Explanation." *Synthese* 28: 87-97.

_______. 1974b. "Statistical Explanations." In *PSA 1972: Proceedings of the 1972 Biennial Meeting of the Philosophy of Science Association*, Vol. 1, edited by Kenneth Schaffner and Robert S. Cohen, 337-47. Boston Studies in the Philosophy of Science, vol. 20. Dordretch: D. Reidel Publishing Company.

_______. 1974c. "Statistical Probabilities: Single Case Propensities vs. Long-run Frequencies." In *Developments in the Methodology of Social Science*, edited by Leinfellner and Köhler, 387-97. Dordrecht: D. Reidel Publishing Company.

_______. 1977a. "Reichenbach, Reference Classes, and Single Case 'Probabilities'." *Synthese* 34: 185-217. Errata, *Synthese* 37 (1978): 113-114.

_______. 1981. *Scientific Knowledge: Causation, Explanation, and Corroboration*. Boston Studies in the Philosophy of Science, vol. 69. Dordrecht: D. Reidel Publishing Company.

_______. (ed.). 1988. *Probability and Causality: Essays in Honor of Wesley C. Salmon*. Synthese Library: Studies in Epistemology, Logic, Methodology, and Philosophy of Science, vol. 192. Dordrecht: D. Reidel Publishing Company.

Feynman, Richard P. 1965. *The Character of Physical Law*. Cambridge: The MIT Press.

Feynman, Richard P., Robert B. Leighton, and Matthew Sands. 1965. *Quantum Mechanics*. Vol. 3, *The Feynman Lectures on Physics*. Reading Mass.: Addison-Wesley Publishing Company.

Garfinkel, Alan. 1981. *Forms of Explanation: Rethinking the Questions in Social Theory*. New Haven: Yale University Press.

Giere, Ronald N. 1973. "Objective Single-case Probabilities and the Foundations of Statistics." In *Logic, Methodology and the Philosophy of Science*, 4, edited by Patrick Suppes et al., 476-83. Amsterdam: North-Holland.

_______. 1979. "Propensities and Necessity." *Synthese* 40: 439-51.

Glymour, C. 1986. "Comment: Statistics and Metaphysics." *Journal of the American Statistical Association* 81: 964-86.

Good, I. J. 1983. "A Causal Calculus." In *Good Thinking: The Foundations of Probability and Its Applications*, by I. J. Good, 197-218. Minneapolis: University of Minnesota Press. (First published in *British Journal for the Philosophy of Science* 11 (1961): 305-18; 12 (1962): 43-51; 13 (1962): 88.)

Hacking, Ian. 1965. *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

_______. 1973. "Propensities, Statistics and Inductive Logic." In *Logic, Methodology, and Philosophy of Science*, Vol. 4, edited by Patrick Suppes et al., 485-500. Amsterdam: North-Holland.

Hanna, Joseph. 1981. "Single Case Propensities and the Explanation of Particular Events." *Synthese* 48: 409-36.

_______. 1983. "Probabilistic Explanation and Probabilistic Causality." In *PSA 1982: Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, edited by Peter D. Asquith and Thomas Nickles, 181-93. East Lansing: Philosophy of Science Association.

Hempel, Carl G. 1962. "Deductive-Nomological vs. Statistical Explanation." In *Scientific Explanation, Space and Time*, Minnesota Studies in the Philosophy of Science, vol. 3, edited by Herbert Feigl and Grover Maxwell, 98-169. Minneapolis: University of Minnesota Press.

_______. 1965. "Aspects of Scientific Explanation." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, by Carl G. Hempel, 331-496. New York: The Free Press.

_______. 1968. "Maximal Specificity and Lawlikeness in Probabilistic Explanation." *Philosophy of Science* 35: 116-33.

Holland, P. 1986. "Statistics and Causal Inference" (with comments by Rubin, Cox, Glymour, and Granger). *Journal of the American Statistical Association* 81: 945-70.

Humphreys, Paul. 1985. "Why Propensities Cannot Be Probabilities." *The Philosophical Review* 94: 557-70.

Kitcher, Philip, and Wesley C. Salmon. 1987. "Van Fraassen on Explanation." *Journal of Philosophy* 84: 315-30.

Kitcher, Philip, and Wesley C. Salmon, eds. 1989. *Scientific Explanation*. Minnesota Studies in the Philosophy of Science, vol. 13. Minneapolis: University of Minnesota Press.

Leggett, A. J. 1986. "Quantum Mechanics at the Macroscopic Level." In de Boer, Dal, and Ulfbeck eds. *The Lesson of Quantum Theory: Niels Bohr Centennial Symposium, 1985*. Amsterdam: North Holland.

Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance." In *Ifs: Conditionals, Belief, Decision, Chance, and Time*, edited by William L. Harper, Robert Stalnaker, and Glenn Pearce, 267-97. The University of Western Ontario Series in Philosophy of Science, vol. 15. Dordrecht: D. Reidel Publishing Company.

_______. 1986. "Causal Explanation." In *Philosophical Papers: Volume 2*, 214-40. New York: Oxford University Press.

Lipton, Peter. 1991a. "Discussion: Contrastive Explanation and Causal Triangulation" *Philosophy of Science* 58: 687-97.

_______. 1991b. *Inference to the Best Explanation*. Philosophical Issues in Science Series, edited by W. H. Newton-Smith. London: Routledge.

Mill, John Stuart. 1904. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. 8th edition. London: Longmans, Green, and Co.

Miller, David. 1991. "Single-Case Probabilities." *Foundations of Physics* 21: 1501-16.

Milne, P. 1986. "Can there be a Realist Single-Case Interpretation of Probability?" *Erkenntnis* 25: 129-32.

________. 1987. "Physical Probabilities." *Synthese* 73: 329-59.

Niiniluoto, I. 1988. "Probability, Possibility, and Plentitude." In *Probability and Causality*, 91-108. *See* Fetzer 1988.

Popper, Karl R. 1957. "The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory." In *Observation and Interpretation*, edited by S. Korner, 65-77. London: Butterworths.

________. 1959a. *The Logic of Scientific Discovery*. New York: Basic Books.

________. 1959b. "The Propensity Interpretation of Probability." In *The British Journal for the Philosophy of Science* 10: 25-42.

Putnam, Hilary. 1979. "The Logic of Quantum Mechanics." In *Mathematics, Matter and Method: Philosophical Papers, Volume 1*, by Hilary Putnam, 174-97. Second edition. Cambridge: Cambridge University Press. (First published as "Is Logic Empirical?" In *Proceedings of the Boston Colloquium for the Philosophy of Science 1966-1968*, edited by R. Cohen and M. Wartofsky, 199-215. Boston Studies in the Philosophy of Science, vol. 5. Dordretch: D. Reidel Publishing Company, 1968.)

Railton, Peter. 1978. "A Deductive-Nomological Model of Probabilistic Explanation." *Philosophy of Science* 45: 206-26.

________. 1981. "Probability, Explanation, and Information." *Synthese* 48: 233-56.

Reichenbach, Hans. 1949. *The Theory of Probability*. Berkeley: University of California Press.

________. [1956] 1991. *The Direction of Time*. Edited by Maria Reichenbach. Reprint, with foreword by Hilary Putnam. Berkeley: University of California Press.

Salmon, Wesley C. 1971. "Statistical Explanation." In *Statistical Explanation and Statistical Relevance*, edited by Wesley C. Salmon, Richard C. Jeffrey and James G. Greeno, 29-87. Pittsburgh: University of Pittsburgh Press. (First published in Robert G. Colodny, ed., *The Nature and Function of Scientific Theories*, 173-231. Pittsburgh: University of Pittsburgh Press, 1970.)

———. 1974. "Comments on 'Hempel's Ambiguity' by J. Alberto Coffa." *Synthese* 28: 165-69.

———. 1977a. "Objectively Homogeneous Reference Classes." *Synthese* 36: 399-414.

———. 1977b. "A Third Dogma of Empiricism." In *Basic Problems in Methodology and Linguistics*, edited by Robert Butts and Jaakko Hintikka, 149-66. Dordrecht: D. Reidel Publishing Company.

———. 1979. "Propensities: A Discussion Review." *Erkenntnis* 14: 183-216.

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

———. 1988. "Dynamic Rationality: Propensity, Probability, and Credence." In *Probability and Causality*, 3-40. *See* Fetzer 1988.

———. 1989. "Four Decades of Scientific Explanation." In *Scientific Explanation*, 3-219. *See* Kitcher and Salmon 1989.

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

———. 1984. *Pragmatism and Empiricism*. New Haven: Yale University Press.

———. 1988. "Conditional Chance." In *Probability and Causation*, 161-78. *See* Fetzer 1988.

van den Berg, Hans, Dick Hoekzema, and Hans Radder. 1990. "Discussion: Accardi on Quantum Theory and the 'Fifth Axiom' of Probability." *Philosophy of Science* 57: 149-57.

van Fraassen, Bas C. 1977. "The Pragmatics of Explanation." *American Philosophical Quarterly* 14: 143-50.

———. 1980. *The Scientific Image*. Oxford: Clarendon Press.

von Mises, Richard. 1964. *Mathematical Theory of Probability and Statistics*. New York: Academic Press.