A Generalized Probabilistic Theory of Causal Relevance
Author(s): Christopher Read Hitchcock
Source: *Synthese*, Vol. 97, No. 3 (Dec., 1993), pp. 335-364
Published by: Springer
Stable URL: http://www.jstor.org/stable/20117848
Accessed: 23/09/2013 10:45

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Synthese*.

http://www.jstor.org

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

CHRISTOPHER READ HITCHCOCK

A GENERALIZED PROBABILISTIC THEORY OF CAUSAL RELEVANCE*

ABSTRACT. I advance a new theory of causal relevance, according to which causal claims convey information about conditional probability functions. This theory is motivated by the problem of disjunctive factors, which haunts existing probabilistic theories of causation. After some introductory remarks, I present in Section 3 a sketch of Eells's (1991) probabilistic theory of causation, which provides the framework for much of the discussion. Section 4 explains how the problem of disjunctive factors arises within this framework. After rejecting three proposed solutions, I offer in Section 6 a new approach to causation that avoids the problem. Decision-theoretic considerations also support the new approach. Section 8 develops the consequences of the new theory for causal explanation. The resulting theory of causal explanation incorporates the new insights while respecting important work on scientific explanation by Salmon (1971), Railton (1981), and Humphreys (1989). My conclusions are enumerated in Section 9.

1. INTRODUCTION

Traditional theories of causation have tried to analyze causes as being necessary or sufficient conditions for their effects. Over the past several decades, however, philosophers have become increasingly interested in probabilistic theories of causation, which characterize causation in terms of probability relations. Thus smoking causes lung cancer, not because all or only smokers develop lung cancer, but because smokers are *more likely* to develop lung cancer than non-smokers. This and similar examples lend to the probabilistic approach an air of plausibility. But with increasing philosophical interest has come a barrage of criticisms and counterexamples. In this paper I will discuss one particular problem that plagues most probabilistic theories of causation in one form or another: the problem of disjunctive factors.[1] This problem illustrates the need to move to a more generalized probabilistic theory of causation. Although motivated by the problem of disjunctive factors, I believe the generalized theory that emerges stands on its own merits; in particular, I argue that this account meets the needs of a theory of causal explanation.

I will take as my point of departure Eells's probabilistic theory of causation, together with the hypothetical limiting frequency conception of probability that he grudgingly adopts (Eells, 1991). It is my hope,

Synthese **97**: 335–364, 1993.
© 1993 *Kluwer Academic Publishers. Printed in the Netherlands.*

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

however, that the suggested generalization may be appended *mutatis mutandis* to the reader's favourite probabilistic theory of causation as well.

2. NOTATION

Before proceeding, it is necessary to settle on some conventions for notation and terminology. A probability space is a triple $\langle \Omega, \mathcal{F}, P \rangle$, where Ω is a set, $\mathcal{F}$ a σ-field over Ω, and P a probability function with domain $\mathcal{F}$. (Definitions of all probability concepts used in this paper are given in the Appendix.) A lower case 'ω' will be used to denote members of Ω. Upper case Roman letters from the beginning of the alphabet (A through G), possibly primed or subscripted, will be used to denote members of $\mathcal{F}$. Members of $\mathcal{F}$ will be called 'events'; this word is intended here only in the set-theoretic sense – it is not intended to invoke any metaphysical theory of events. Members of $\mathcal{F}$, as sets, are subject to the operations of union, intersection, and complementation, denoted '$\cup$' and '$\cap$' and '$\sim$' respectively. The symbols '$\bigcup$', '$\bigcap$' will be used to denote generalized union and intersection. Upper case Roman letters from the middle of the alphabet (H through K), possibly primed or subscripted, will be used to refer to sets of real numbers; H will only be used for Borel sets. Lower case Roman letters will play different roles, frequently standing for individual real numbers or functions. The upper case Roman letters from the end of the alphabet (W through Z), possibly primed or subscripted, will stand for random variables. For ease of notation, '$X \in H$' will abbreviate '$\{\omega\colon X(\omega) \in H\}$', and similarly for '$X = x$'. Conditional probabilities of the form $P(E|X = x)$ are generalized conditional probability functions, and may be defined even if $P(X = x)$ is zero.

3. A PROBABILISTIC THEORY OF CAUSATION

Probabilistic theories of causation have centred on the following idea: C is a cause of E if $P(E|C) > P(E|{\sim}C)$. This idea needs to be expanded, of course, and there are many different variations on the basic theme. The best developed probabilistic theory of causation is that of Eells (1991), which is based loosely on the theory advanced in Cartwright (1979). For definiteness, we will work within the framework of Eells's theory. Of necessity, the following exposition achieves brevity at the

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

cost of clarity; interested readers are encouraged to study Eells's more leisurely presentation.

Before delving into the details, I would like to draw attention to three respects in which the theory to be sketched here differs slightly from that of Eells. First, Eells intends his theory to give an account of type-level causation; that is, it is supposed to capture causal generalizations, such as 'smoking causes cancer'. Eells offers an independent theory of token-level causation to capture singular causal claims such as 'Harry's smoking caused him to develop lung cancer'. However, not all authors who have written on the topic of probabilistic causality agree that a theory of the sort sketched below is inappropriate for singular causation. Humphreys (1989), for example, offers a variant of this theory as an account of singular causation, while Suppes (1970) seems to take his theory to be neutral between the two levels of causation. I will maintain a position of neutrality on this issue.

Second, Eells, following Cartwright, argues that it is not possible to provide a reductive analysis of causation in terms of probabilities. (For an attempt at such a reductive analysis, see Papineau, 1989.) Instead, Eells introduces four primitive relations: positive causation, negative causation, mixed causation, and neutrality. The first three relations are different species of causal relevance, the fourth a species of causal irrelevance. The probabilistic theory then imposes constraints upon these relations. I suggest a slight modification of Eells's theory at this point. It suffices to begin with one primitive relation, causal relevance, which is extensionally equivalent to the union of Eells's three relations of causal relevance. The resulting theory not only provides constraints on this primitive relation, but also provides a reduction of the three species of causal relevance to probabilities and the primitive relation of causal relevance. Restructuring the theory in this way makes explicit an important contribution made by probabilistic theories of causation: they provide *taxonomies* of causal relevance. Consider a paradigmatic causal claim: 'smoking causes lung cancer'. There is much to this claim that is not captured by the relevant probability relations, such as the existence of processes in the lungs that lead to the formation of cancer cells. What the probability relations *do* capture is the sense in which smoking (which is causally relevant to lung cancer) *promotes* lung cancer, rather than inhibits it.

Third, Eells conceives of probability as "an objective and a physical relation between event types" (1991, pp. 34–35). In order to preserve

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

mathematical clarity, I prefer to talk of probability in the formal mode: probability is a function over set-theoretic entities. Nonetheless, the assumption underlying Eells' conception can be translated: there are objective and physical relations between "event types" that admit of probabilistic *representation*. Events in a σ-field can represent the sorts of entities that stand in causal relations (Eells usually calls them "factors"), and probability assignments to those events can represent objective features of such entities. The formal primitive 'causal relevance' must then be a two-place relation between events in a σ-field, so it too is abstract. Again, we can assume that this formal relation corresponds to a physical relation (which need not be physically or philosophically primitive).

Eells's causal relations have four argument places. Claims about the causal relevance of C for E are always made relative to a population p and a population-type t. The choice of C, E, and p constrains the choice of t. For example, the population p must be of type t, and t must not be such as to specify explicitly the frequency of the factors represented by C and E in any population of type t. It is the population-type t that determines the structure of the probability space $\langle \Omega, \mathcal{F}, P \rangle$. As an intuitive crutch, we might think of t as an indeterministic set-up that produces populations, where p is the outcome of one trial of type t. The probability of A will be the limiting relative frequency of the factor it represents in a hypothetical sequence of populations generated in conformity with t.[2] By determining the structure of Ω and $\mathcal{F}$, t also determines which events other than C and E are to be considered in evaluating the causal relevance of C for E.

In order to evaluate the causal relevance of C for E relative to p and t it is necessary to construct a partition $\{G_1, G_2, \ldots\}$[3] of the outcome space Ω determined by t. Each set G_i will be called a *cell* of the partition. Each cell represents a uniform causal background context; the partition divides the entire probability space into all of its possible background contexts. The procedure for constructing the partition is described below.

The construction involves two defined relations: interaction, and causal subsequence. Let $\{A_1, A_2, \ldots\}$ be an arbitrary partition of the outcome space Ω. C *interacts* with this partition, with respect to E, iff for each i, $p_i = P(E|C \cap A_i)$, $q_i = P(E|{\sim}C \cap A_i)$, and $i \neq j$ implies that either $p_i \neq p_j$ or $q_i \neq q_j$. Informally, C interacts with a partition, relative to E, if the different cells of the partition make a difference to the

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

probability of E, conditional on either C or $\sim C$. B is a *subsequent causal factor* to C, with respect to E, iff (i) C is causally relevant to B; or (ii) E is causally relevant to B; or (iii) there is some $D \in \mathcal{F}$, such that C is causally relevant to D, D is causally relevant to E, and D is causally relevant to B; note that all clauses in the definition of causal subsequence should be understood as being relativized to a population and a population-type.

Let $\{F_1, F_2, \ldots\}$ consist of all $F_i \in \mathcal{F}$ such that (i) F_i is a member of some partition with which C interacts with respect to E; and (ii) F_i is not a subsequent causal factor to C with respect to E (relative to p and t). Intuitively, each F_i represents a factor that is relevant to E independently of C; such factors should be held fixed when evaluating the causal relevance of C for E. For example, if C represents smoking, and E lung cancer, F_1 might represent exposure to asbestos, F_2 a genetic predisposition to lung cancer, and so forth. There is a worry that smoking might be positively or negatively correlated with one of these factors. If smoking is negatively correlated with exposure to asbestos, for example, it could turn out that smokers are *less* likely to develop lung cancer than non-smokers in the population as a whole, even though smoking raises the probability of lung cancer both in the presence and in the absence of asbestos exposure. This reversal of probabilistic relevance would occur because smokers would be less likely to be exposed to asbestos. In order to avoid erroneous causal conclusions, the causal relevance of smoking for lung cancer should be assessed while other relevant factors are held fixed in the background. One must take care, however, not to hold fixed those factors that are causally subsequent to smoking. Suppose, for example, that smoking causes lung cancer exclusively by depositing pollutants in the lungs. Then it will turn out that smoking has no effect on the probability of lung cancer if one holds fixed the presence or absence of pollutants in the lungs; nonetheless, we should not conclude that smoking does not cause lung cancer: it causes lung cancer *by* polluting the lungs.

Let the partition $\{G_1, G_2, \ldots\}$ contain all the intersections of maximal consistent sets of the F_is and their complements. This partition is analogous to the set of Carnapian state descriptions over the set of predicates corresponding to $F_1, F_2, \ldots$. Each G_i holds fixed each of the members of F_i either negatively or positively. The partition $\{G_1, G_2, \ldots\}$ is the desired partition for evaluating the causal relevance of C for E.

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

C is said to be *causally positive* for E relative to p and t if $P(E|C \cap G_i) > P(E|\sim C \cap G_i)$ for all G_i in the partition; C is *causally negative* for E if $P(E|C \cap G_i) < P(E|\sim C \cap G_i)$ for all G_i; *causally neutral* if $P(E|C \cap G_i) = P(E|\sim C \cap G_i)$ for all G_i; and *causally mixed* if different relations hold in different cells.[4] C is *causally relevant* to E iff it is not causally neutral; this imposes a formal constraint upon the primitive relation. In normal English usage, the phrases 'C causes E' and 'C promotes E' are used in place of 'C is causally positive for E' similarly, 'C prevents E' and 'C inhibits E' are used for 'C is a negative cause of E'.

Some have challenged the requirement that positive causes must raise the probabilities of their effects in *all* background contexts (and likewise for negative causes), a condition which Eells, following Dupré (1984), calls *context-unanimity*. Skyrms (1980) suggests that a condition of Pareto dominance be employed: C is a positive cause of E if $P(E|C \cap G_i) > P(E|\sim C \cap G_i)$ for some G_i and $P(E|C \cap G_i) \geq P(E|\sim C \cap G_i)$ for all G_i (negative causation can be defined analogously). Dupré (1984) suggests a more radical alternative. Eells argues against these proposals (1991, pp. 94–107). Eells's defense rests in part on considerations of expressive power: his taxonomy permits greater precision in conveying information about the underlying probability relations. This provides a clue to the account that will be developed in Section 6. Until then, it suffices that the causal relevance of C for E in a population depends on the values of $P(E|C \cap G_i)$ and $P(E|\sim C \cap G_i)$ for each cell G_i. In a population-type involving only a single cell, the causal relevance of C for E depends only on the probability values within that cell. In the discussion that follows, then, we will assume the population-type to be sufficiently narrow to give rise to a homogeneous causal background. Note that there can be no mixed causal relevance relative to such a population-type, so I will frequently omit reference to this species of causal relevance when discussing the different types of causal relevance below. Similarly, the relativity of causal relevance to a population and population-type, while important for many purposes, will not be of concern in what follows. I will assume that these relata are fixed, and suppress reference to them at many points, talking instead as if the relations of causal relevance were binary.

4. THE PROBLEM OF DISJUNCTIVE FACTORS

Humphreys offers an example that serves to illustrate the problem of disjunctive factors (1989, pp. 41–42).[5] Suppose that research is being

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

conducted on the efficacy of a new drug in the treatment of a certain disease. The research team is interested in whether the drug is effective in moderate doses, since in large doses the drug has toxic side effects. The thirty subjects in the study all have the disease. Let us suppose, moreover, that they are selected so to be uniform with regard to other factors that might aid or impair recovery, as well as with regard to factors that would interact with the drug. Ten subjects are assigned at random to each of three treatment groups. The first group is given a placebo, represented by C_0, the second receives a moderate dose, C_1, and the third a strong dose, C_2; recovery from the disease within the time period of the study will be represented by E. The causal question guiding the research is: Is C_1 causally positive for E?

The probabilities for recovery are as follows:

$$P(E|C_0) = .2,$$
$$P(E|C_1) = .4,$$
$$P(E|C_2) = .9.$$

(We will assume that the recovery rates within each treatment group accurately reflect the probabilities of recovery for each group.) If we assume that $P(C_0) = P(C_2)$, then $P(E|{\sim}C_1) = (P(C_0)P(E|C_0) + P(C_2)P(E|C_2))/(P(C_0) + P(C_2)) = .5P(E|C_0) + .5P(E|C_2) = .55 > .4 = P(E|C_1)$. According to the theory outlined above, then, C_1 is causally negative for E. This result seems wrong, given the various probabilities for recovery. For example, a doctor treating a patient who cannot afford a strong dose of the drug would be foolish to refrain from prescribing a more moderate dose on the grounds that such a dose would actually prevent recovery from the disease.

Perhaps the hypothesis that $P(C_0) = P(C_2)$ is at fault: if $P(C_0)$ is sufficiently large – if $P(C_0)/P(C_2) > 5/2$ – then the computation yields the result that C_1 is a positive cause of E. It is remarkable, however, that the causal relevance of C_1 for E should depend on the values of $P(C_0)$ and $P(C_2)$ at all. Whether C_1 is causally positive for recovery should depend upon the chemical properties of the drug and human physiology, not on the probabilities of receiving a placebo or a strong dose of the drug. It would be curious, for instance, if the company manufacturing the drug had to offer the following advice to physicians: 'The drug, when given in moderate doses, is efficacious in treating the disease so long as you prescribe at least five placebos for every two strong dosages . . .'.

Three interrelated problems are exposed by this example. In increas-

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

ing order of gravity they are: (i) the values of $P(C_0)$ and $P(C_2)$ are needed in order to compute $P(E|\sim C_1)$, yet these values are not obvious from the description of the experimental set-up; (ii) natural assumptions about the values of $P(C_0)$ and $P(C_2)$ lead to counter-intuitive causal judgements; (iii) according to the probabilistic theory of causation outlined above, the causal efficacy of C_1 for E depends on the ratio of the probabilities of C_0 and C_2, but the causal efficacy of C_1 for E should not depend on these probabilities at all. The problems arise because $\sim C_1$ is disjunctive, being equivalent to $C_0 \cup C_2$, where the two disjuncts confer different probabilities upon the effect E. Hence the name: 'the problem of disjunctive causal factors'.

Similar problems arise when the causal factor, and not just its negation, is disjunctive. Suppose, in the example above, that the researchers are interested not only in whether the subjects recover from the disease but also in whether they survive the trial period with no serious medical problems whatsoever, be they due to the disease, or to side effects of the drug. Let F be survival without medical complications. In moderate doses, the drug is unlikely to produce serious side effects, but in higher doses, it is so likely to produce side effects that this risk outweighs the curative benefits of the drug. Here are the probabilities:

$$P(F|C_0) = .2,$$
$$P(F|C_1) = .3,$$
$$P(F|C_2) = .1.$$

What is the causal relevance of taking the drug (in any quantity) to F? That is, what is the relevance of $C_1 \cup C_2$ to F? Whether $P(F|C_1 \cup C_2)$ is greater than, equal to, or less than $P(F|C_0)$ will depend upon the probabilities of C_1 and C_2. It is clear that analogues of problems (i) and (iii) will arise, although problem (ii) will not arise since we do not have any strong pre-theoretic intuitions about what the causal relevance of $C_1 \cup C_2$ *should* be.

It might be objected that it does not make sense to talk of the effects of such disjunctive causes. How might one argue for such an objection? First, it could be argued that gerrymandered events should not be allowed to stand in causal relations. This restriction seems reasonable enough, but the disjunction of C_1 and C_2 is perfectly natural, so the proscription of gerrymandered events does not solve the current problem.[6] Second, it might be argued that only events that are maximally specific can stand in causal relations. It is hard to see what could

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

motivate such a view: certainly no probabilistic theory of causation has ever insisted that a cause C cannot have any subsets in $\mathcal{F}$. Thus, while we are uneasy about assigning a causal role to $C_1 \cup C_2$, we have, as yet, no principled reason for rejecting such a role.

Before canvassing some proposals for tackling these problems, I want to advance some desiderata that these proposals should satisfy. My primary desideratum is that the resulting probabilistic theory of causation vindicate our strong pre-theoretic intuitions, or reasonable refinements of those intuitions, about what is a cause of what. For example, the revised theory should yield the result that C_1 is a cause of E, and this result should not be contingent upon the values of $P(C_0)$ and $P(C_2)$. There are two secondary desiderata, of roughly equal importance. First, a theory of causation should not leave any meaningful causal claims indeterminate in truth value. Thus, a theory should say whether $C_1 \cup C_2$ is causally positive or negative for F, or show that $C_1 \cup C_2$ and F are not the sorts of events that can be meaningfully said to stand in causal relations. Second, it is a desideratum that a theory of causation allows us to explain why we have some of the causal intuitions that we do; in particular, it would be nice to understand why we feel that C_1 is a positive cause of E, but have no intuition about whether $C_1 \cup C_2$ is a cause of F.

5. THREE PROPOSED SOLUTIONS

The first solution, advocated by Humphreys (1989), is suggested almost immediately by the given example. Our intuition is that C_1 is a positive cause of recovery, E, because the probability of recovery is higher for those that take moderate doses of the drug than for those who take only a placebo; that is, $P(E|C_1) > P(E|C_0)$. This is, no doubt, the comparison the research team would make in evaluating the efficacy of the drug in moderate doses. In general, Humphreys's suggestion is that in order to determine the causal relevance of B for A one must compare $P(A|B)$ with $P(A|B_0)$, where B_0 is an objectively determined *neutral state*, a privileged alternative to B.[7] B is a positive cause of A if $P(A|B) > P(A|B_0)$ (and analogously for other species of causal relevance).[8] There is no universal neutral state: different putative causes will have different neutral states. Determining the appropriate neutral state may itself require causal knowledge. For example, in determining the efficacy of the drug in the example above, the neutral state would

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

not be the absence of any treatment whatsoever, but treatment with a placebo. It was an important discovery that being provided with drugs by a health professional can have curative effects that go beyond the chemical properties of the substance prescribed.[9]

There are problems with this solution to the problem of disjunctive factors, however. First, it does not resolve the problem in the case where the cause is disjunctive. If we compare $P(F|C_1 \cup C_2)$ with $P(F|C_0)$, the probability of F in the neutral state, we are still left with the original problem.

A second problem is that comparison with probabilities in the neutral state will sometimes be impossible or lead to the wrong conclusions. Consider the following example. High blood pressure is known to cause a variety of health problems: let us ask, then, what the causal relevance of having a blood pressure of 180/120 is to survival in the following year. Let these factors be represented by B and D, respectively. (We will again assume that all of the appropriate background factors are held fixed.) To what shall we compare the probability $P(D|B)$? The obvious choice for a neutral state would be zero blood pressure: call it B_0. (Humphreys's comments (1989, pp. 38–41) suggest that the zero level of a variable that is measurable on a ratio scale should be taken as the neutral state relative to other values of that variable.) Insofar as it makes sense to attribute a value to $P(D|B_0)$, this value would have to be zero. There is a non-zero probability of surviving the year with blood pressure 180/120, so $P(D|B) > P(D|B_0)$; having a blood pressure of 180/120 would be causally positive for survival, according to this proposed solution. This result is strongly counter-intuitive.

Perhaps some other level of blood pressure could serve as the neutral state? One possibility would be that having optimal blood pressure would constitute the neutral state B_0. Suppose that in the relevant background context, having a blood pressure of 120/80 maximizes the probability of surviving the year. Now $P(D|B) < P(D|B_0)$, yielding the intuitively correct result that having a blood pressure of 180/120 is causally negative for survival. The problem with this suggestion is that it would force us to say that having a blood pressure of 120/80 is causally *neutral* for survival. (It is a trivial consequence of Humphreys's proposal that any event is always causally neutral for any effect for which it is a neutral state.) Most, however, would be inclined to say that having optimal blood pressure is causally positive for survival; at

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

the very least it should be causally relevant in *some* way, and not lumped with other neutral factors such as shoe size and favourite colour.

Perhaps the average blood pressure for those in the relevant group could serve as a neutral value. This, however, would raise the problem that the causal relevance of blood pressure of 180/120 would depend upon demographics as well as physiology. Moreover, this suggestion would come close to rendering certain causal claims – such as 'the average fifty-year-old male has dangerously high blood pressure' – false by definition. Thus, the example of the relevance of blood pressure to survival shows that Humphreys's solution to the drug example is not universally applicable.[10]

The second proposed solution to the problem of disjunctive factors is due to Eells (1988, 1991). Return to the original example involving the drug, where C_0 represents treatment with a placebo, C_1 treatment with a moderate dose, C_2 treatment with a strong dose, and E recovery; we are interested in whether C_1 is a positive cause of E. Suppose that r is a probability function that assigns probability one to $C_0 \cup C_2$. Consider the conditional: 'if the subject s did not receive treatment C_1, s would have received treatment C_0 or C_2, with probabilities $r(C_2)$ and $r(C_2)$, respectively'.[11] Treating s as a variable, this conditional can be construed as a property; it holds of a subject if the resulting conditional is true when her name is substituted for s. Let the event F_r represent this property. Eells argues that $P(E|{\sim}C_1 \cap F_r) = r(C_0)P(E|C_0) + r(C_2)P(E|C_2)$, so it is possible to compare $P(E|C_1)$ directly with $P(E|{\sim}C_1)$ if F_r is held fixed in the background. Eells provides an independent motivation for holding F_r fixed: let q be a probability function different from r such that $q(C_1) = 0$. Then $P(E|{\sim}C_1 \cap F_q) \neq P(E|{\sim}C_1 \cap F_r)$, so we can expect F_q and F_r to be members of a partition with which C_1 interacts with respect to E (in Eells's sense of interaction defined in Section 3 above). Thus, within any cell of the partition of background contexts, just one event of the form F_r will be held fixed. Since F_r determines a unique probability for $P(E|{\sim}C_1)$, the causal relevance of C_1 for E will be unequivocally determined within each cell of the partition. A generalization of this approach can be applied to the case where the cause is also disjunctive.

This approach has its problems, too. First, there is a technical difficulty. Call a probability function r *trivial* if it assigns probability one to either C_0 or C_2. It is clear that Eells wants his account to apply to

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

non-trivial probability distributions. According to the Lewis–Stalnaker approach to subjunctive conditionals, however, the properties $\sim C_1$ and F_r will be incompatible unless r is trivial; thus $P(E|\sim C_1 \cap F_r)$ will be *undefined* for non-trivial r.[12] On the Lewis–Stalnaker approach, the conditional 'If S were true, then T would be true' is true whenever S and T are both true. In the set-up described in the example of Section 4, any subject who does not receive treatment C_1 will receive treatment C_0 or treatment C_2. Thus, of any possible subject who does not receive C_1, one of the following two conditionals must be true: (i) 'if she did not receive C_1, she would have received C_0'; (ii) 'if she did not receive C_1, she would have received C_2'. Both of these conditionals are incompatible with the subject's having the property represented by F_r unless r is a trivial probability measure. Thus, if r is a probability function that does *not* assign probability one to either C_0 or C_2, then *no* subject could possibly have the property represented by F_r together with the property represented by $\sim C_1$. Put in the language of hypothetical limiting frequencies: in a hypothetical infinite sequence of populations, *no* individuals will have the property represented by $\sim C_1 \cap F_r$. Eells's claim that $P(E|\sim C_1 \cap F_r) = r(C_0)P(E|C_0) + r(C_2)P(E|C_2)$ is thus false: $\sim C_1 \cap F_r = \phi$, so $P(E|\sim C_1 \cap F_r)$ is undefined. Of course, Lewis and Stalnaker do not have a monopoly on theories of subjunctive conditionals,[13] but we are still owed an account that will undergird Eells's proposed computation.

Second, the example of Section 4 gave rise to three inter-related problems, and Eells's proposal aims at resolving only one. Let us assume that Eells can provide an account of subjunctive conditionals to motivate his equation: $P(E|\sim C_1 \cap F_r) = r(C_0)P(E|C_0) + r(C_2)P(E|C_2)$.[14] As the example was described, equal numbers of subjects were randomly assigned to the three treatment groups C_0, C_1, and C_2. For the subjects in this study, then, the following conditional would be true (modulo difficulties described above): 'if the subject had not been in treatment group C_1, she would have been equally likely to have been in groups C_0 and C_2'. Thus $P(E|\sim C_1) = .5P(E|C_0) + .5P(E|C_2) = .55 > .4 = P(E|C_1)$. According to Eells's proposal, then, C_1 will be causally negative for recovery: the counter-intuitive result with which we began.[15] Moreover, this solution to the problem makes the causal relevance of C_1 for E sensitive to the mechanism by which the subjects were assigned to the three treatment groups – another undesirable consequence.

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

The third solution can be found in the accounts of statisticians Paul Holland and Donald Rubin (see, e.g., Holland, 1986 and Rubin, 1974). Their framework is different from that described in Section 3, and they do not specifically address the problem of disjunctive factors, but a feature of their account might be adapted in an attempt to solve the problem. The suggestion is to interpret causal relevance as involving an additional argument place. Thus, one could not say that C_1 is a positive, neutral, or negative cause of E *simpliciter*. C_1 can only bear one of these relations to E with respect to some third event, an alternative to C_1. In general, B will be a positive cause of A relative to B' if $P(A|B) > P(A|B')$. Thus, in the drug example, C_1 would be causally positive for E with respect to C_0 but causally negative with respect to C_2.[16] In the blood pressure example, having a blood pressure of 180/120 is negatively relevant to survival when compared with a blood pressure of 120/80, but positively relevant with respect to zero blood pressure (and other extremely low values). Having a blood pressure of 120/80 would be positively relevant for survival when compared with any alternative blood pressure level.[17]

In order to be in accord with our normal causal judgements, this account would need to be supplemented with a discussion of the pragmatics of making causal claims. In making causal claims, we typically do not specify the additional relatum, the alternative to the cause in question; rather, the alternative cause is usually determined by contextual factors. The suppression of the additional relatum in normal discourse is one reason why the causal relation is often taken to be binary, and why we seem to have intuitions about binary causal claims. Note that the context dependence of causal claims, as typically expressed, does not rob the causal relation of its objectivity: it may still be an objective fact whether a certain triple of events satisfies the relation of positive, negative, or neutral relevance. Contextual factors play a role in determining the third relatum in many causal claims, but this does not mean that they play a role in determining whether the resulting ternary causal claims are true.

This proposal has many virtues. It captures much of what is attractive about Humphreys's account: the neutral state, whether explicitly mentioned or not, is frequently used as the third relatum in causal claims. In the drug example for instance, the claim that moderate doses of the drug are causally positive for recovery would normally be interpreted with the neutral state – the placebo – as the third relatum. It is for this

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

reason that we are inclined to accept the claim that C_1 is a positive cause of E. The neutral state is not always the third relatum, however, as the blood pressure example shows.[18]

The Holland–Rubin proposal does have its shortcomings, however. It does not resolve the problem of disjunctive causal factors for the case where the cause is disjunctive. Recall that, in this example, $P(F|C_0) = .2$, $P(F|C_1) = .3$, and $P(F|C_2) = .1$; and we are interested in the causal relevance of $C_1 \cup C_2$ for F. In this example, C_0 is the only alternative to $C_1 \cup C_2$, so $P(F|C_0)$ is the same as $P(F|\sim(C_1 \cup C_2))$, and explicit relativization of causal claims to an alternative cause leaves us with the original problem.

None of the proposals provides an adequate solution to the problem of disjunctive causal factors. The first proposal yields the result that C_1 is a cause of E, which agrees with our pre-theoretic intuition; however, this proposal clashes with our intuitions on a different example. Moreover, the first proposal was silent on the question of whether $C_1 \cup C_2$ is a cause of F, and did not provide an explanation of our lack of intuitions about this example. The second proposal, if supplied with the necessary technical refinements, could provide an answer to the question of whether $C_1 \cup C_2$ is a positive cause of F: it depends upon the conditional that is held fixed in the background context. Moreover, this proposal explains why our intuitions are silent in this case: the example is underdescribed unless the appropriate background conditional is specified. However, the second proposal does not yield the intuitively correct result that C_1 is a cause of E in the original example. The third proposal agrees with our intuition that C_1 is a cause of E if we are willing to let this intuition be refined slightly. Moreover, the third proposal can be made to agree with our intuitions in the blood pressure example, on which the first proposal ran aground. Unfortunately, the third proposal fails to give an account of the causal relevance of $C_1 \cup C_2$ for F and it fails to explain the silence of our intuitions.

Of the three proposals, the third is the most successful, as it is the only one that completely satisfies the primary desideratum of vindicating our causal intuitions. This proposal introduced two interesting suggestions into the discussion of causation. The first is that events like C_1 cannot be said to be positive or negative causes of events like E in an absolute sense: there are no two-place relations of positive and negative causal relevance.[19] In order to square this idea with our common causal judgements, it was suggested that English expressions which appear to

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

describe binary relations of causal relevance are actually elliptical for more complex attributions of causal relevance. Both of these suggestions will have a useful role to play in the solution advocated below.

6. A NEW SOLUTION

I want to propose a generalization of the third alternative that is a more radical departure from existing probabilistic theories of causation. First, I suggest that we use random variables to represent the causal factors and their alternatives. In the drug example, let X represent the dosage of the drug taken; X will take values in the non-negative real numbers.[20] Similarly, in the blood pressure example, we could let Y be the (vector-valued) random variable that measures blood pressure. Probabilistic theories of causation have typically been concerned with defining the causal relevance of *events* of the form $X \in H$. It would not be torturing the language, however, to talk instead of the causal relevance of *variables* like X.[21] It would be natural to say that the dosage of the drug received is causally relevant to recovery (X is causally relevant to E) and that blood pressure (Y) is causally relevant to survival (D). In standard English, 'affects' and 'influences' often stand in for 'is causally relevant to' in such phrases.

Let us define conditional probability functions as follows: let $f(x) = P(E|X = x)$, and $g(x, y) = P(D|Y = (x, y))$.[22] In the drug example, the probability of recovery increases with the dosage of the drug taken, so $f(x)$ is monotonically increasing, asymptotically approaching one. The blood pressure example is more complex. The function g has two argument places, so its graph would be a two-dimensional surface over the first quadrant of the Cartesian plane.[23] The function would reach its maximum (less than one) at $(x, y) = (120, 80)$. If one were to draw a line of moderate slope, about 2/3, outward from the origin, and consider the restriction of g to this line, the values of g would display an inverted 'U' shape, increasing monotonically to 120/80, and decreasing monotonically beyond that. My suggestion is to let the function $f(x)$ represent the causal relevance of the variable X for E, and $g(x, y)$ of Y for D. Causal relevance, then, is infinite in variety: it does not break down simply into positive and negative relevance.

Typically, we do not know the details of functions like f. Even when we do know the values of the function with some accuracy, there may be no convenient linguistic means of conveying them, or no reason to

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

convey them in excruciating detail. Frequently, then, we will want to communicate certain important features about the general shape of the function. One important way of doing this is by making comparisons of the probability of E conditional upon different values of the variable X. For example, the inequality $P(E|X = m) > P(E|X = n)$ conveys important information about the function $f(x) = P(E|X = x)$. This inequality would be conveyed linguistically by claims of the form: 'doses of strength m tend to cause recovery (when compared with doses of strength n)'. (The parenthetical clause is frequently left implicit.) Thus the language of positive and negative causal relevance, when interpreted along the lines of the Holland–Rubin proposal discussed above, can be used to convey information about more complex causal relations. However, not all causal claims express inequalities of this sort. 'Sufficiently large doses of the drug make recovery virtually certain' is a causal claim, which expresses the following piece of information about the function f: 'there exists an m such that $x > m$ implies that $f(x) > 1 - \epsilon$', where the value of ϵ is typically left vague. The claim would also carry a strong implicature to the effect that $f(x)$ is lower for smaller values of x. This claim could not be expressed using only inequalities of the sort considered by the third proposal. Conversely, it is not the case that arbitrary triples of events can be said to stand in some sort of causal relation: attributions of causal relevance to such triples are only appropriate if they convey information about a conditional probability function. For example, I will suggest below that there is no useful information conveyed by an attribution to $C_1 \cup C_2$ of positive, negative, or neutral relevance for F with respect to C_0.

There are two reasons why we have so long held on to the belief that causal relevance can be partitioned into a few simple varieties, such as positive and negative. The first is that statements about positive and negative causal relevance can convey important information about the structure of functions like f. Because we typically use the language of positive and negative causation to describe the causal relevance of (say) X for E, we have come to believe that these expressions pick out categories in a simple taxonomy of causal relevance. The second reason is that the simple dichotomy is a vestige of determinism. If determinism were true, there would be a set H such that $P(E|X \in H) = 1$, and $P(E|X \in \mathbb{R} - H) = 0$. In this situation, it would be natural to describe $X \in H$ (and perhaps its subsets) as a positive cause of E, and $X \in \mathbb{R} - H$ (and perhaps its subsets) as a negative cause of E.[24] If determinism is

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

false, however, $P(E|X = x)$ can take on any value between zero and one inclusive. For what values of x, then, should we say that $X = x$ is a positive, rather than a negative cause of E? There is no natural way to divide the possible values of X into positive and negative causes; I urge that we stop looking for an unnatural way.

Here, then, is the solution to our original problem. The causal relevance of the drug for recovery is given by the function f. The inequalities $P(E|C_0) < P(E|C_1)$ and $P(E|C_1) < P(E|C_2)$ convey important information about the function f, although they do not exhaustively describe the causal relevance of X for E. The comparison of $P(E|C_1)$ to $P(E|\sim C_1)$ conveys almost no useful information about the function f; $P(E|C_1) < P(E|\sim C_1)$, for example, tells us only that the probability of E is not a maximum at C_1. It does not tell us whether the probability of E increases from C_0 to C_1 to C_2, decreases through this interval, or reaches a minimum at C_1.[25] This probability comparison is uninformative because the probability value $P(E|\sim C_1)$ depends on the values of the quantities $P(C_0)$ and $P(C_2)$, which have nothing to do with the values of f. In many contexts 'moderate doses of the drug are causally positive for recovery' would be elliptical for 'moderate doses of the drug are causally positive for recovery *when compared with treatment by placebo*'. This causal claim expresses the inequality $P(E|C_1) > P(E|C_0)$, which does provide information about f. Although there is no objective relation of positive relevance holding between C_1 and E, we would still be inclined to make the claim of positive relevance as a way of conveying true and useful information. This explains our initial inclination to say that C_1 is a positive cause of E.

In the example where the researchers are concerned with F, survival without any major medical complications, the causal relevance of X for F is given by the function $h(x) = P(F|X = x)$. The inequalities $P(F|C_2) < P(F|C_0) < P(F|C_1)$ provide useful information about h. The comparison of $P(F|C_1 \cup C_2)$ to $P(F|C_0)$, by contrast, provides virtually no information about the function h, for the reasons discussed in the previous example. This explains why we lack intuitions about which causal claim to make in this case: we would not use causal terminology to make such a comparison because it tells us so little about the function h. Nor is the proposed account under any obligation to determine the actual causal relevance of $C_1 \cup C_2$ for F; causal relevance is, at the deepest level, a relation that takes a variable, not an event, as its first relatum. Claims about the causal relevance of events are appropriate

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

only insofar as they provide information about conditional probability functions.

Thus the proposal advanced here meets all the desiderata discussed in Section 4. It vindicates our inclination to say that C_1 is a cause of E and can be made to agree with our intuitions in the example involving blood pressure as well. Moreover, it explains why we have no inclination to attribute to $C_1 \cup C_2$ any sort of causal relevance for F, and why any such attribution of causal relevance would be inappropriate.

Recall, however, that the function f described above determines the causal relevance of the drug for recovery only within one cell of the partition of causal background contexts (and analogously for g and h). In some other cell, some function different from f will describe the causal relevance of the drug for recovery. In a heterogeneous population, the causal relevance of X for E will be determined by an array of functions of the form $f_i(x) = P(E|X = x \cap G_i)$, where $\{G_1, G_2, \ldots\}$ is the partition of the outcome space into possible causal background contexts. Typically, it is not possible to convey information about all of the f_is; again, most causal claims in English give only partial information. Singular causal claims, such as 'the drug caused Mary's recovery' may refer to single cells – in this case, the cell that applies to Mary.[26] A causal claim such as 'the drug promotes recovery in most cases, although resistant strains have been discovered' might mean: for most G_i, f_i is monotonically increasing; however, for G_i in which W (the resistance of the disease) is greater than m, f_i is almost constant. In general, the information about the f_is that is conveyed by a particular causal claim will be qualitative in nature, and will depend upon context as well as the literal meanings of the words used.

I return now to a brief discussion of Eells's context-unanimity condition: the requirement that a cause C must raise the probability of the effect E in *all* causal background contexts. Recall that Eells defended this condition on the grounds of expressive power. Against Skyrms's Pareto-dominance requirement, for example, Eells argues that everything that can be said using the language of Pareto-dominant causation can be said in the language of context-unanimous causation, while the converse does not hold.[27] Eells concedes, however, that:

> [O]ne may carve up all the possibilities however one wants, and if I do it differently from the way you do it, then we simply arrive at *different concepts*. One set of concepts may be more versatile or descriptive than the other for one purpose, and vice versa for another purpose; and each set of concepts may be just as 'legitimate' and coherent as the other.

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

I do not think there is anything conceptually wrong or incoherent with the Pareto revision . . . . (1991, pp. 97–98)

Eells's defense together with this concession suggest that the *point* of making causal claims is to provide information about the underlying probability relations, and not to ascribe objective relations of positive and negative relevance. This is a suggestion that I strongly endorse. Eells prefers the context-unanimity approach because it allows for more precision in describing probability relations. But Eells's account does not provide maximal expressive power, for it permits only the expression of probabilistic inequalities. Complete expressive power comes only with the ability to specify all of the values of the conditional probability functions described above. Unfortunately, this level of expressive power is beyond our linguistic and cognitive powers, and we must resort to more inexact tools of communication. The language of positive and negative causal relevance is just such a tool.

7. ASIDE ON DECISION

The logical geography in the neighbourhood of causation is complex. Many domains of interest to philosophers border that of causation: knowledge, moral responsibility, probability, explanation, and decision are but a few. While explanation will be the causal neighbour of primary concern in this paper, a brief excursion into decision theory will provide some support for the theory of causal relevance outlined in the preceding section. The argument will not hinge on the details of causal decision theory, but on a platitudinous connection between causation and decision: if one is trying to achieve some end, one should perform those actions that tend to cause that end, and abstain from those that tend to prevent it.[28]

Suppose that a doctor is treating a patient for the disease in the example of Section 4. One of her ends will be to cure the patient of the disease, that is, to bring about E. The doctor will have other ends as well, such as avoiding side effects from medication, minimizing the costs to the patient, and so forth. In order to prescribe a dosage of the drug that is suited to the first goal, she will want to have information about the causal relevance of the drug for recovery. What does the doctor need to know about the causal relevance of the drug for E in order to make an informed decision? I contend that the information

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

contained in the function $f(x) = P(E|X = x)$ – and that is considerable information – suffices. She would prescribe the moderate dose C_1, for example, if she thought that this was the dosage for which the chance of recovery most strongly outweighed the chance of suffering from side effects and the financial cost of filling the prescription. The function f tells her what the chance of recovery is for every level of dosage – no more information about the relevance of the drug for recovery is needed for her computation (although she will need information about the relevance of the drug for side effects, financial cost, and so forth). According to the proposals advanced by Humphreys and Eells, however, the doctor could possess all of the information contained in the function f and still not know whether a particular level of dosage like C_1 *causes* recovery. According to Humphreys, she would need to know which value of X served as the neutral state; according to Eells, she would need to know which subjunctive conditional was true of the patient. It is evident, however, that the doctor need not know either of these in order to reach an informed decision.

8. CAUSAL EXPLANATION

Humphreys (1989) presents an account of causal explanation in which explanations have the following grammar: 'E because Φ, despite Ψ', where Φ is a partial list of positive causes and Ψ is a partial list of negative causes. (According to Humphreys, Φ must be non-empty, although Ψ need not be.) Correct explanations need not exhaustively list all causally relevant factors, nor need they cite probability values for E (although the list of causes will entail certain probabilistic inequalities). This picture of causal explanation is appealing: in particular, it seems right that a causal explanation should not only provide a list of causally relevant factors but also say something about the species of causal relevance exhibited by each factor. Suppose it be asked why Harry suffered a heart attack; it would be misleading to respond that he smoked heavily and exercised often, without specifying that the former was a contributing cause, and the latter an inhibiting cause. If the account of the preceding sections is correct, however, there is no neat division of causes into positive and negative; causal relevance comes in all shapes and sizes. I want to suggest a picture of causal explanation that incorporates the ideas of the previous sections, while preserving the appealing features of Humphreys's account.

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

We may take as a clue a parallel between the approach to causation recommended above, and the approach to explanation suggested in Railton (1981). It was argued above that typical causal claims do not tell the complete story about the causal relevance of a variable for an event, but only provide information about this complex relation of causal relevance. Railton argues that for any explanandum there is an "ideal explanatory text" containing all the information necessary for a complete understanding of why the explanandum occurred. However, one is seldom, if ever, in a position to describe completely the ideal text when an explanation is requested. Instead, actual explanations provide information about the ideal text. An explanation is correct insofar as what it says about the ideal text is true; explanatory to the extent that it provides information about the ideal text: the more informative an explanation is about the ideal text, the more explanatory it is. If the ideal explanatory text were to include complete descriptions of the conditional probability functions described above, then one way to provide information about the text, and thus to provide an explanation, would be to describe these conditional probability functions.

The picture of causal explanation I propose is a hybrid of Railton's (1981) theory of explanation and that of Salmon (1971). The core of Salmon's statistical relevance (*S-R*) theory of explanation is contained in the following passage:

> [A]n explanation of the fact that x, a member of A, is a member of B, would go as follows:
>
> $$P(B|A \cap C_1) = p_1$$
> $$P(B|A \cap C_2) = p_2$$
> $$\ldots$$
> $$P(B|A \cap C_n) = p_n$$
>
> where
>
> $A \cap C_1, A \cap C_2, \ldots, A \cap C_n$ is a homogeneous partition of A with respect to B,
> $p_i = p_j$ only if $i = j$,
> $x \in A \cap C_k$.
>
> (Ibid., 1971, pp. 76–77: the notation is changed slightly)

Thus an *S-R* explanation is an assemblage of statistically relevant factors.

My suggestion is that for causal explanations, there is an 'ideal causal array' that looks something like Salmon's collection of statistically relevant factors. Actual explanations, however, can seldom if ever de-

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

scribe this array in its entirety, but can only provide partial information about the array. It is the role of causal claims to provide such partial information. The structure of the ideal array will differ somewhat from that of Salmon's model for *S-R* explanations, in order to accommodate advances in probabilistic theories of causation over the past two decades, and to accommodate the shift in emphasis from events to variables recommended in the previous sections.

I will propose a construction of the ideal array that mirrors Eells's construction as outlined in Section 3. Parallels to other probabilistic theories of causation are no doubt possible. We take as our primitive a relation of causal relevance that holds between a variable and an event.[29] Generalizing this relation, we will say that the random variable W is causally relevant to the variable Z if there is a Borel set H such that W is causally relevant to the event $Z \in H$. Suppose E is the event to be explained, and that χ_E is the characteristic function of E. First, we will need to say what it is for one variable to interact with another, since this notion has only been defined in terms of events. We will say that X_j *interacts with* X_i with respect to E if there exist Borel sets H, $H_1, H_2, \ldots$ such that $\{H_1, H_2, \ldots\}$ is a partition of the range of X_j and $X_i \in H$ interacts with $\{X_j \in H_1, X_j \in H_2, \ldots\}$, a partition of Ω, in the sense of Section 3. Let $\{X_1, X_2, \ldots\}$ be the maximal set of random variables such that each X_i either (i) is causally relevant to E or (ii) interacts with some X_j that is causally relevant to E. The set of all sets of the form $\bigcap_i X_i = x_i$ is a partition of the outcome space, and these sets provide the basis for the ideal array. We may also use the set $\{X_1, X_2, \ldots\}$ to impose a constraint upon the primitive relation of causal relevance. For this, we need an analogue of Eells's notion of causal subsequence. X_j will be causally subsequent to X_i relative to E if: (i) X_i is causally relevant to X_j; (ii) χ_E is causally relevant to X_j; or (iii) there is some X_k such that X_i is causally relevant to X_k, X_k, is causally relevant to E, and X_k is causally relevant to X_j. Let $K_i = \{j: j \neq i$ and X_j is not causally subsequent to $X_i\}$. Causally relevant variables must satisfy the following constraint: X_i is causally relevant to E iff there exist x, y, and x_j for each $j \in K_i$ such that $P(E|X_i = x \cap \bigcap_{j \in K_i} X_j = x_j) \neq P(E|X_i = y \cap \bigcap_{j \in K_i} X_j = x_j)$. In other words, there must be some background context in which some difference in the value of X_i gives rise to some difference in the probability of E.

The ideal causal array will be a probability space over a σ-field that contains E, and on which all of the random variables $X_1, X_2, \ldots$

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

are measurable. This ideal array will contain all information about conditional probabilities of the form $P(E|X_i = x_i \cap \bigcap_{j\in K_i} X_j = x_j)$, where the set K_i is defined as above. In particular, for each of the variables, the ideal causal array will contain all of the information contained in the functions f_i defined in the last section. Let us change the notation slightly, in order to make it explicit that we have such functions for *each* causally relevant variable. Let k_i be a function from K_i to $\mathbb{R}$ such that $k_i(j)$ is in the range of X_j. Then define $f_{i,k_i}(x) = P(E|X_i = x \cap \bigcap_{j\in K_i} X_j = k_i(j))$. Here the constellation of factors represented by $\cap_{j\in K_i} X_j = k_i(j)$ plays the same role as a cell of the partition $\{G_1, G_2, \ldots\}$ in Eells' theory. The causal relevance of X_i for E is given by the set of conditional probability functions of the form f_{i,k_i}. Generalizing, we can say that the causal relevance of X_i for the *variable* Z is given by the conditional *distribution* functions of the form $\mu_{i,k_i}(H, x) = P(Z \in H|X_i = x \cap \bigcap_{j\in K_i} X_j = k_i(j))$. Talking of the causal relevance of one variable for another, rather than of a variable for an event, would be a natural extension of the change in emphasis from events to variables recommended above. In contexts where explanations are requested, however, there is usually some specific event specified as explanandum.[30]

Actual explanations will not be able to describe the ideal array in every detail, just as causal claims in English are typically unable to describe the functions f_{i,k_i} in every detail. Instead, explanations will provide information about this ideal causal array. One way to provide information about the ideal array is to cite some of the causally relevant variables, and to provide some information about the functions that characterize the causal relevance of each – this sort of information is typically conveyed by causal locutions in English. The way in which different causal claims in English can be used to convey this sort of information about the ideal array has already been discussed in Section 6.

Suppose, for example, that Martha was a subject in the study described in Section 4, and that she recovered from the disease. Why did she recover? The ideal causal array will include the event E, representing recovery, and a set of variables. Among these will be X, representing the dosage of the drug taken; a family of variables $\{Y_i\}_{i\in I}$ representing various physiological factors relevant to the probability of recovery, like T-cell count; a family $\{Z_j\}_{j\in J}$ representing features of different strains of the disease, such as their level of resistance to the

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

drug; and so forth. An explanation might cite any of these factors, but let's assume that the person requesting the explanation is interested in finding out about any forms of medical treatment relevant to Martha's recovery. In this context, it would be appropriate to point out that Martha received a moderate dose of the drug, and to describe the functional dependence of the probability of recovery upon doses of the drug. Typically, this functional dependence is described using causal language: 'Martha's taking a moderate dose of the drug (rather than a placebo) was a positive cause of her recovery'. The language of positive, negative, and mixed causal relevance is useful in providing explanations, not because it corresponds to an objective taxonomy of causal relevance, but because it corresponds to the sort of imprecise information that we typically have at our disposal when called upon to provide explanations.

The probability space constituting the ideal array will contain information beyond that contained in the conditional probability functions of the form $P(E|X_i = x_i \cap \bigcap_{j \in K_i} X_j = x_j)$. This raises two questions. First, why is information about these conditional probability functions of particular interest? Second, is information about the underlying probability space which is *not* of this sort ever of interest in explanatory contexts? In answer to the first question, I would speculate that the information contained in the conditional probability functions described above is that which is needed for decision. The brief digression in Section 7 supports this, but clearly more work needs to be done in order to make the intimated connection between causation and decision explicit.[31] In answer to the second question, I have little doubt that other types of information about the underlying probability space will be of interest in some explanatory contexts. It is even likely that causal language could be used to convey some of this information. After all, the discussion has focussed only on that fragment of causal language that deals with different species of causal relevance. Other pieces of causal language might be used to convey different sorts of information about the underlying probability space. Non-causal explanations might also be construed as providing information about a probability space. For example, R. I. G. Hughes (1989) has argued that quantum mechanical probabilities are defined not over a σ-field of events, but over a sub-Boolean lattice. Moreover, he claims that this piece of information about quantum mechanical probabilities is explanatory. Both claims

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

are controversial, but it seems clear that to provide this sort of information about a probability space is not to provide a *causal* explanation.

9. CONCLUSION

I have offered a new approach to probabilistic theories of causation. The main features of this approach are these. One, probabilistic theories of causation do not provide analyses, or reductions, of causation in terms of probabilities; but, given a relation of causal relevance as a primitive, such theories can be seen as attempts to provide non-circular *taxonomies* of causal relevance. Two, causal relevance is best seen not as a relation between two events (in the probabilistic sense), but as a relation between a variable and an event, or even between two variables. Three, there is no natural division of causal relevance into a few simple species, such as positive and negative; rather, causal relevance is infinite in variety. Four, causal claims, such as those made when providing explanations, are used to provide information – usually of a sketchy and qualitative sort – about a complex probability space. The language of positive and negative causal relevance is useful for this end, but it does not correspond to any objective division of causal relevance into positive and negative.

Several arguments support this account. First, traditional probabilistic theories of causation have run aground on the problem of disjunctive factors, whereas the approach outlined above deftly deals with the various aspects of this problem. Second, on this approach, the causal information that is needed for decision is precisely information about causal relevance. Finally, the approach to causation outlined emerges naturally as a hybrid of extant theories of explanation. Since explanation is an important context in which one makes causal claims, this consonance between the suggested approach to causation and existing theories of explanation is highly desirable.

APPENDIX: PROBABILITY

(1) A *probability space* is a triple $\langle \Omega, \mathcal{F}, P \rangle$ where Ω is a set, $\mathcal{F}$ is a *σ-field* of subsets over Ω and P a *probability measure* on $\mathcal{F}$.
(2) $\mathcal{F}$ is a *σ-field* over Ω if $\mathcal{F}$ is a set of subsets of Ω such that:

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

(i) $\phi \in \mathcal{F}$;
(ii) $A \in \mathcal{F}$ implies $\sim A \in \mathcal{F}$;
(iii) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

(3) *P is a probability measure* on $\mathcal{F}$ if it is a function with domain $\mathcal{F}$, such that:

(i) For all $A \in \mathcal{F}$, $P(\phi) = 0 \leq P(A) \leq 1 = P(\Omega)$;
(ii) If $A_1, A_2, \ldots \in \mathcal{F}$, and $A_i \cap A_j = \phi$ for all $i \neq j$, then $P(\bigcup_{i \in \mathbb{N}} A_i) = \Sigma_{i \in \mathbb{N}} P(A_i)$.

(4) $H \subseteq \mathbb{R}$ is a *Borel set* if it belongs to $\mathcal{B}$, the smallest σ-field containing all open intervals (or, equivalently, all closed intervals). This definition can be generalized for subsets of $\mathbb{R}^n$.
(5) X is a *random variable* on the probability space $\langle \Omega, \mathcal{F}, P \rangle$ if it is a real valued function with domain Ω, such that for any Borel set H, $\{\omega: X(\omega) \in H\} \in \mathcal{F}$. *$X$ is said to be measurable* with respect to $\mathcal{F}$ if it satisfies this condition. This definition can be generalized for *random vectors*, which take values in $\mathbb{R}^n$. If X is a random variable on Ω, then $\sigma(X)$ denotes the smallest σ-field over Ω with respect to which X is measurable.
(6) The *conditional probability* $P(A|B)$ (read: the probability of A given B) is standardly defined as the ratio $P(A \cap B)/P(B)$. If $P(B) = 0$, this ratio is undefined. A generalized notion of conditional probability is available, however. Let $\langle \Omega, \mathcal{F}, P \rangle$ be a probability space, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-field over Ω. For fixed A, there will be a random variable Z such that:

(i) Z is measurable with respect to $\mathcal{G}$;
(ii) $\int_B Z(\omega) d(P) = P(A \cap B)$ for all $B \in \mathcal{G}$.

The existence of such a variable is guaranteed by the *Radon–Nikodym* theorem. Any random variable satisfying conditions (i) and (ii) is said to be a *version* of the conditional probability $P[A \| \mathcal{G}]$. Any two versions of $P[A \| \mathcal{G}]$ will differ at most on a set of probability zero; that is, if Y and Z are versions of $P[A \| \mathcal{G}]$, and $D = \{\omega: Y(\omega) \neq Z(\omega)\}$, then $P(D) = 0$. The standard definition of conditional probability follows as a special case. Let $\{B_1, B_2, \ldots\}$ be a countable partition of Ω such that $P(B_i) > 0$ for each $i \in \mathbb{N}$. Let $\mathcal{G}$ be the smallest σ-field that is a superset of this partition. Then the only version of $P[A \| \mathcal{G}]$ is the function Z, defined as follows:

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

$$Z(\omega) = P(A \cap B_i)/(P(B_i), \text{ where } \omega \in B_i.$$

Let X be a random variable, and let Z be a version of $P[A||\sigma(X)]$, then the function $P(A|X = x)$ can be defined as follows:

$$P(A|X = x) = Z(\omega), \text{ for any } \omega \text{ such that } X(\omega) = x.$$

Since Z is measurable with respect to $\sigma(X)$, it follows that Z is constant on the set $\{\omega\colon X(\omega) = x\}$. Note that the values of $P(A|X = x)$ may depend upon the selection of a particular version of $P[A||\sigma(X)]$.

NOTES

* I would like to thank Nuel Belnap, John Earman, Richard Gale, Paul Humphreys, Satish Iyengar, Wes Salmon, and two anonymous referees for comments and discussion. I am also indebted to the members of an audience at the Center for Philosophy of Science at the University of Pittsburgh, where some of the ideas contained in this paper were presented.

[1] The problem is not unique to probabilistic theories of causation; it is particularly pressing for counterfactual theories as well, but the details will not be discussed here.

[2] There remains an important ambiguity in this informal description; see Eells (1991, Chap. 1) for more discussion.

[3] The notation is not intended to indicate that this set is countably infinite. In general, I will use this notation for sets when I do not want to commit myself to the set's being of a particular cardinality.

[4] Eells also imposes a temporal priority condition upon the relations of positive, negative, and mixed causation, in order to ensure that these relations are anti-symmetric. Thus any factors that occur after E are also to be considered causally neutral for E. We will not worry about the details of this addition here.

[5] The moral that Humphreys draws from this example is somewhat different from that developed here.

[6] Sober (1984, pp. 93–95) offers a criterion for distinguishing natural from unnatural disjunctions of causes; $C_1 \cup C_2$ passes.

[7] Humphreys's approach is similar to that taken by David Lewis's counterfactual theory of causation (Lewis, 1973; 1986b). On this account, one compares the probability of the effect in the presence of the cause with the value that the probability of the effect would have had *if the cause had been absent*. Some of Lewis's comments (e.g., 1986b, pp. 210–11) suggest that the situation in which the cause would have been absent should be taken to be something like Humphreys's neutral state.

[8] Note that in the discussion that follows, we will consider the theory outlined in Section 3, emended by Humphreys's suggestion that comparisons be made with probabilities conditional on the neutral state. This hybrid theory is not to be attributed to Humphreys, whose own theory differs in important ways.

[9] It should be noted that Humphreys's treatment of the background context is slightly different from that outlined in Section 3. On the account sketched there, we would not need to take so much care in specifying the neutral state. Since all subjects in the test

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

are put through the same process of being prescribed pills by a health professional, regardless of the actual quantities of the drug consumed, this aspect of the treatment can be taken to be an independent causal factor that is held fixed in the background context. With this factor held fixed, receiving a dosage of zero would automatically entail receiving a placebo.

[10] This discussion is not entirely fair to Humphreys, since he holds that causes may be either states or *changes* in states. The neutral state for a change is the absence of change. For example, in determining the causal relevance of an increase in blood pressure from 160/100 to 180/120, we would compare the probability in the presence of the change with the probability in the absence of change – where blood pressure remains constant at 160/100. This would yield the correct result that the increase is a negative cause of survival. In general, for blood pressure levels higher than 120/80, increases in blood pressure are negative for survival. It is this fact, Humphreys might argue, that we are elliptically referring to when we say that high levels of blood pressure are negative for survival. For more discussion, see note 18 below.

[11] Strictly speaking, the conditional should read: 'if the subject did not receive the treatment *represented by* $C_1, \ldots$', but it is already a mouthful. Where it aids readability, the words 'represented by' will be omitted below.

[12] Generalized conditional probability is of no help here, since the problem is not merely that the set $\sim C_1 \cap F_r$ has measure zero, but that this set is identical with the empty set. Even if there were an appropriate σ-field $\mathcal{G}$, and a version of Z of $P[E\|\mathcal{G}]$, there is *no* $\omega \in \sim C_1 \cap F_r$ such that we might take $Z(\omega)$ to be $P(E|\sim C_1 \cap F_r)$.

[13] For a survey of the logic of conditionals, see Nute (1984). The principle that causes problems for Eells's account is the one that Nute labels **CS**; Nute's discussion makes it clear which theories are committed to **CS** and which are not.

[14] In a terse paragraph, Good (1961–1962, p. 309) suggests that claims about the causal strength of C_1 for E are made relative to a probability distribution over alternatives to C_1. This suggestion could license the equation proposed by Eells: since Good does not require that such distributions are rooted in the truth of subjunctive conditionals, his account does not run into the sorts of problems sketched above. Good's suggestion might also be taken as a generalization of the Holland–Rubin proposal described below.

[15] I have been informed that Eells does not find this consequence counter-intuitive, so we have a clash of intuitions here.

[16] One of the referees has suggested that Eells's theory has the resources to make the same three-place causal claims: C_1 is causally positive for E relative to a population whose members receive either the treatment represented by C_0 or the treatment represented by C_1; but C_1 is causally negative relative to a population receiving C_1 and C_2.

[17] The suggestion that the causal relation has an additional argument place is also easily adapted to counterfactual theories of causation, as Glymour (1986) notes.

[18] Similarly, the Holland–Rubin approach allows for the sorts of probability comparisons made in Humphreys's theory of changes, and for some of those made in Eells's theory of token causation. At the same time, this proposal is not committed to the claim that the appropriate *change* in probability values must occur in order for the comparison of the two values to be relevant.

[19] Of course several causal relations that have been proposed in the literature, such as

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

those of Eells, involve more than two argument places. It is not the move from a binary to a ternary relation per se that is of central interest in the Holland–Rubin account, but the relativization of causal relevance to an alternative causal factor.

[20] We need not explicitly state that $X(\omega) = 0$ represents treatment with a placebo, rather than no treatment at all: see note 9 above.

[21] Indeed this is standard usage in much of the literature on causal modelling and causation in the social sciences. Suppes (1970, Chap. 5) extends some of his ideas to cover causal relations between variables. In particular, his brief description of *functional causes* is suggestive of the ideas contained in this section.

[22] The values of f may depend upon the selection of a version of $P[E||\sigma(X)]$ (and similarly for g). Since none of what is said below requires that f be determined by other features of the probability space, we may assume that there is a privileged version of $P[E||\sigma(X)]$ in which we are interested.

[23] Actually, the graph would be more tightly confined. Since diastolic pressure does not exceed systolic pressure, the graph would be confined between the x-axis and the line of slope one passing through the origin.

[24] This account of positive relevance is roughly equivalent to Mackie's (1974) account of causation in terms of *inus* conditions. According to this account, $X \in H$ would be a cause of E if it is an insufficient but non-redundant part of an unnecessary but sufficient condition. The sufficient condition here is $X \in H$ together with the other factors that are held fixed to determine the background context in the cell of the partition under consideration. This conjunction of factors is not necessary, since the probability of E might be one in other cells of the partition (for certain values of X). Within this cell of the partition, the probability of E will be zero if $X \notin H$, so the factor $X \in H$ is non-redundant. Finally, $X \in H$ will, by itself, be insufficient for E, depending as it does on the other factors making up the background context.

[25] There may, conceivably, be contexts in which the information that the probability of E is not a maximum at C_1 is important, but this is not one of them.

[26] This is very controversial, however. Eells believes that the probabilistic theory of causation sketched in Section 3 does not apply to singular causation. Humphreys (1989) offers a probabilistic theory of singular causation in which the cause must raise the probability of the effect in all physically possible background contexts, not just the actual one.

[27] Actually, the converse only fails if claims that quantify over sub-populations are not allowed, as Eells himself notes (1991, p. 97, n. 25).

[28] For a nice discussion relating probabilistic theories of causation with the means-end relationship relevant to decision, see Mellor (1988).

[29] Recall that this relation is formal, taking set-theoretical entities as relata. It is assumed, however, that there exists a physical relation to which the mathematical primitive corresponds.

[30] Keep in mind that 'event' is here being used in its set-theoretic sense. In particular, this claim should not be understood as asserting that causal explanation involves causal relations at the singular, rather than the general level. Cartwright (1979), for example, argues that causal explanations invoke causal laws rather than singular causal claims.

[31] The suggestion is also in the spirit of Cartwright (1979), who argues that causal laws are needed in order to distinguish effective strategies from ineffective ones.

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions

REFERENCES

Cartwright, N.: 1979, 'Causal Laws and Effective Strategies', *Noûs* **13**, 419–37, reprinted with additions in Cartwright (1983).
Cartwright, N.: 1983, *How the Laws of Physics Lie*, Clarendon Press, Oxford.
Dupré, J.: 1984, 'Probabilistic Causality Emancipated', in P. A. French, T. E. Uehling, H. K. Wettstein (eds.), *Midwest Studies in Philosophy IX*: *Causation and Causal Theories*, University of Minnesota Press, Minneapolis, pp. 169–75.
Eells, E.: 1988, 'Probabilistic Causal Interaction and Disjunctive Causal Factors', in Fetzer (1988, pp. 189–209).
Eells, E.: 1991, *Probabilistic Causality*, Cambridge University Press, Cambridge.
Fetzer, J. H. (ed.): 1988, *Probability and Causality*: *Essays In Honor of Wesley C. Salmon*, D. Reidel, Dordrecht.
Glymour, C.: 1986, 'Statistics and Metaphysics', *Journal of the American Statistical Association* **81**, 964–66, comment on Holland (1986).
Good, I. J.: 1961–1962, 'A Causal Calculus I–II', *British Journal for the Philosophy of Science* **11**, 305–18; **12**, 43–51.
Holland, P. W.: 1986, 'Statistics and Causal Inference', *Journal of the American Statistical Association* **81**, 945–60.
Hughes, R. I. G.: 1989, *The Structure and Interpretation of Quantum Mechanics*, Harvard University Press, Cambridge, Massachusetts.
Humphreys, P.: 1989, *The Chances of Explanation*: *Causal Explanations in the Social Medical, and Physical Sciences*, Princeton University Press, Princeton, New Jersey.
Lewis, D. K.: 1973, 'Causation', *Journal of Philosophy* **70**, 556–67, reprinted with postscripts in Lewis (1986a).
Lewis, D. K.: 1986a, *Philosophical Papers*, Volume II, Oxford University Press, Oxford.
Lewis, D. K.: 1986b, 'Postscripts to "Causation"', in Lewis (1986a), pp. 172–213.
Mackie, J. L.: 1974, *The Cement of the Universe*, Clarendon Press, Oxford.
Mellor, D. H.: 1988, 'On Raising the Chances of Effects', in Fetzer (1988, pp. 229–39).
Nute, D.: 1984, 'Conditional Logic', in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, D. Reidel, Dordrecht, pp. 387–439.
Papineau, D.: 1989, 'Pure, Mixed, and Spurious Probabilities and their Significance for a Reductionist Theory of Causation', in P. Kitcher and W. Salmon (eds.), *Scientific Explanation*, University of Minnesota Press, Minneapolis, pp. 307–48.
Railton, P.: 1981, 'Probability, Explanation, and Information', *Synthese* **48**, 233–56.
Rubin, D. B.: 1974, 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies', *Journal of Educational Psychology* **66**, 688–701.
Salmon, W. C.: 1971, 'Statistical Explanation', in W. C. Salmon, J. G. Greeno and R. C. Jeffrey (eds.), *Statistical Explanation and Statistical Relevance*, University of Pittsburgh Press, Pittsburgh, pp. 29–87.
Skyrms, B.: 1980, *Causal Necessity*: *A Pragmatic Investigation of the Necessity of Laws*, Yale University Press, New Haven.
Sober, E.: 1984, *The Nature of Selection*, MIT Press, Cambridge, Massachusetts.
Suppes, P.: 1970, *A Probabilistic Theory of Causality*, North-Holland, Amsterdam.

Department of Philosophy
Rice University
Houston, TX 77251
U.S.A.

This content downloaded from 129.118.139.129 on Mon, 23 Sep 2013 10:45:53 AM
All use subject to JSTOR Terms and Conditions