

CHRISTOPHER READ HITCHCOCK

THE MISHAP AT REICHENBACH FALL:  
SINGULAR VS. GENERAL CAUSATION\*

(Received 10 November 1993)

1. INTRODUCTION

What is the relationship between claims of singular causation such as

1. David's smoking caused him to develop lung cancer,

and claims of general causation, such as

2. Smoking causes lung cancer?

Hume held that the truth of singular causal claims depended upon the existence of universal regularities in nature. In the first *Enquiry*, for example, Hume wrote that

we may define a cause to be *an object, followed by another, and where all the objects similar to the first, are followed by objects similar to the second.* (Hume [1748] 1955, §VII)

Davidson (1980) concurs, while noting that it may not be apparent which generalization is instantiated by any particular episode of singular causation:

[I]f 'a caused b' is true, then there are descriptions of *a* and *b* such that the result of substituting them for 'a' and 'b' in 'a caused b' is entailed by true premises of the form of (L) and (P) [where (L) provides the form of a causal law, and (P) provides the form of premises describing initial conditions] . . . If this is correct, it does not follow that we must be able to dredge up a law if we know a singular casual statement to be true; all that follows is that we know there must be a covering law. (pp. 159–60)

*Philosophical Studies* 78: 257–291, 1995.

© 1995 Kluwer Academic Publishers. Printed in the Netherlands.

Hume and Davidson may be understood as pursuing the following strategy: analyze the truth-conditions of general causal claims in terms of universal regularities in nature, and then treat singular causal claims as describing instantiations of such regularities. Let us call this the Humean strategy.<sup>1</sup>

In her inaugural lecture at Cambridge University, Elizabeth Anscombe challenged the doctrine that true singular causal claims describe instantiations of universal regularities:

It is over and over again assumed that any singular causal proposition implies a universal statement running "Always when this, then that" ... Even a philosopher acute ... as Davidson, will say, without offering any reason at all for saying it, that a singular causal statement implies *that there is* such a true universal proposition – though perhaps we can never have knowledge of it. Such a thesis needs some reason for believing it! (Anscombe, 1981, p. 147)

One of Anscombe's reasons for rejecting the Humean strategy is the belief that singular causation is compatible with indeterminism. That is, sentence 1 might be true although there is no exceptionless law which subsumes David's smoking and his lung cancer. But even granting the compatibility of singular causation and indeterminism, as I believe we should, the issue is not settled. There remains the option of interpreting general causal claims in terms of statistical rather than universal laws, and then proceeding as before. Let us call this the neo-Humean strategy.

Another type of strategy would be to begin by trying to analyze singular causal claims, say in terms of single case probabilities, and then to try to understand general causal claims as generalizations over singular causal claims. This strategy is suggested by Lewis (1986) and Humphreys (1989), although neither gives an explicit account of how the general causal claims relate to singular causal claims. (Carroll (1991) does attempt to provide such an account.) Let us call this the generalization strategy.

I wish to explore this issue from within the framework of a probabilistic theory of causation. Several proponents of this theory have offered arguments that singular and general causal claims describe distinct species of causal relation. Eells (1991), for example, claims that

Causal claims made on one of the two levels of causation turn out to be quite independent of claims made on the other ... I argue that (1) very little (if anything) about what happens

on the [singular] level can be inferred from [general] level probabilistic causal claims, and that (2) very little (if anything) about [general] level probabilistic causal relations can be inferred from [singular] level probabilistic causal claims. (Eells, 1991, p. 6. I have substituted the words 'singular' and 'general' for 'token' and 'type', respectively.)

I will diagnose the flaw in arguments for this conclusion, and then offer a positive account of the relationship between singular and general causation.

## 2. PROBABILISTIC CAUSATION

The idea that underlies probabilistic theories of causation is that  $C$  is a cause of  $E$  if  $P(E|C) > P(E|\sim C)$ . This idea needs to be elaborated, and it has been elaborated in a number of ways. The best-developed approach is that of Eells (1991), who expands upon the theory advanced in Cartwright (1979). I will here give only an outline of the Cartwright-Eells approach to probabilistic causation.

This theory of causation is formulated within the framework of the mathematical theory of probability. The central concept of this theory is that of a 'probability space'. A *probability space* is a triple  $\langle \Omega, \mathcal{F}, P \rangle$ :  $\Omega$  is a set;  $\mathcal{F}$  is a set of subsets of  $\Omega$  having the structure of a ' $\sigma$ -field' or ' $\sigma$ -algebra';<sup>2</sup> and  $P$  is a 'probability function' which assigns to each member of  $\mathcal{F}$  a 'probability value' between zero and one. Members of  $\mathcal{F}$  are called *events*.<sup>3</sup> I put aside for now questions about the empirical interpretation of these mathematical concepts; questions, for example, about whether probability is to be interpreted in terms of single case propensities or long run frequencies cannot be answered independently of questions about whether it is singular or general causation that is being modelled.

The assessment of the causal relevance of  $C$  for  $E$  is made against a partition  $\{G_1, G_2, \dots\}$ <sup>4</sup> of  $\Omega$ . Each cell of the partition represents a homogeneous causal background context. (See Eells (1991) for an explanation of how this partition is to be constructed.)  $C$  is a cause of  $E$  if  $P(E|C \wedge G_i) > P(E|\sim C \wedge G_i)$  across a sufficient range of cells  $G_i$ . Similarly, we can say that  $C$  is a negative or inhibiting cause of  $E$  – or that  $C$  prevents  $E$  – if  $P(E|C \wedge G_i) < P(E|\sim C \wedge G_i)$  across a sufficient

range of cells. For what range of cells must these inequalities hold? Eells (1991) defends a *context-unanimity* requirement: a cause must raise the probability of its effect in every cell of the partition (and analogously for inhibiting causes).<sup>5</sup> Humphreys' (1989) unconditionality requirement is similar. Skyrms (1980) and Sober (1984) endorse a Pareto-dominance requirement: a cause must raise the probability of its effect in some cell of the partition, and lower the probability of its effect in none. Dupré (1984) can be read as claiming that a cause must raise the probability of its effect in a weighted majority of the cells.<sup>6</sup> We will not try to resolve this dispute here.

Eells (1991) intends his theory to capture only causal relations at the general level (which Eells calls type-level causal relations); he offers a separate theory for singular (token-level) causation. It would beg the question to agree with Eells at this point;<sup>7</sup> but we may note for now that his theory provides a superficially plausible theory of general causation (subject to the refinements introduced below), while the argument considered in section 4 below renders it superficially implausible as an account of singular causation. As an expositional tactic then, I will begin by granting that the probabilistic theory of causation introduced in this section is plausible as an account of general causation. In particular, this account seems to make sense of the sort of statistical methods that are used to assess the truth of claims such as 2.

Elsewhere (Hitchcock, 1993) I have shown that this probabilistic theory of causation needs to be refined. Consider, for example, how this theory would treat the claim that smoking causes lung cancer. If this claim is true, then  $P(LC|S \wedge G_i) > P(LC|\sim S \wedge G_i)$  for a range of background contexts  $G_i$ . (Where  $S$  and  $LC$  have the obvious interpretations.) Let us focus on one of these background contexts, call it  $G_n$ . The theory suggests the binary picture of the probabilities within this cell depicted in Figure 1. There are two possibilities, smoking and non-smoking, and each confers a different probability upon the outcome lung cancer. But a more realistic picture of the underlying probabilities would look more like Figure 2. This figure depicts the function  $f(x) = P(LC|ppd = x \wedge G_n)$  (where  $ppd$  is a variable that measures quantity of smoking in units of packs per day). This picture poses a difficulty for the probabilistic theory of causation, for there are

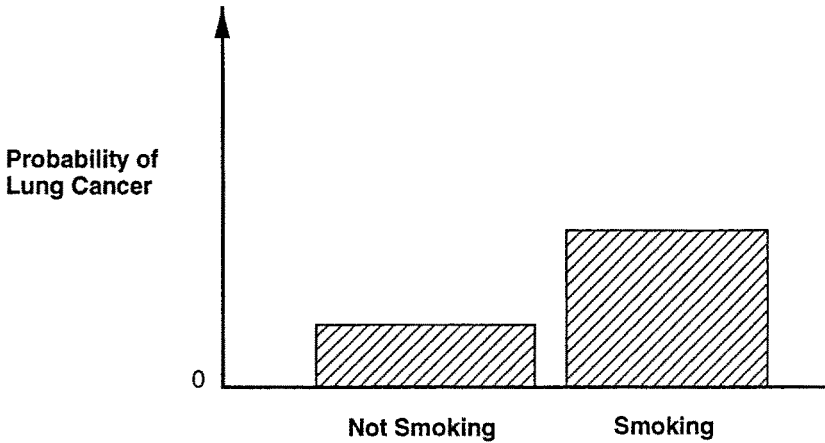


Figure 1. The binary picture.

no longer just two probability values for us to compare. Suppose for example, we are interested in whether smoking one pack of cigarettes per day<sup>8</sup> causes lung cancer. Then, according to the theory sketched above, we must compare  $P(LC|ppd = 1 \wedge G_n)$  to  $P(LC|ppd \neq 1 \wedge G_n)$ . But Figure 2 does not supply a value for this second probability, indeed its value depends upon the probabilities  $P(ppd = x|G_n)$  for  $x \neq 1$ ; so whether smoking one pack of cigarettes a day causes lung cancer depends upon the probability of smoking zero packs a day, half a pack a day, two packs a day, and so on. This, I have argued, is a very counterintuitive consequence of the standard probabilistic theory of causation. The problem is a version of what Eells (1988) has dubbed the problem of ‘disjunctive causal factors’.

The solution advocated in (Hitchcock, 1993) is to view causal claims as describing conditional probability functions, such as the function  $f$  depicted in Figure 2. This is typically done by comparing two different values of the function. It follows that causal claims are typically made relative to an alternative. For example, the claim that smoking one pack per day causes lung cancer would normally be understood as being made relative to the alternative of smoking zero packs per day. This

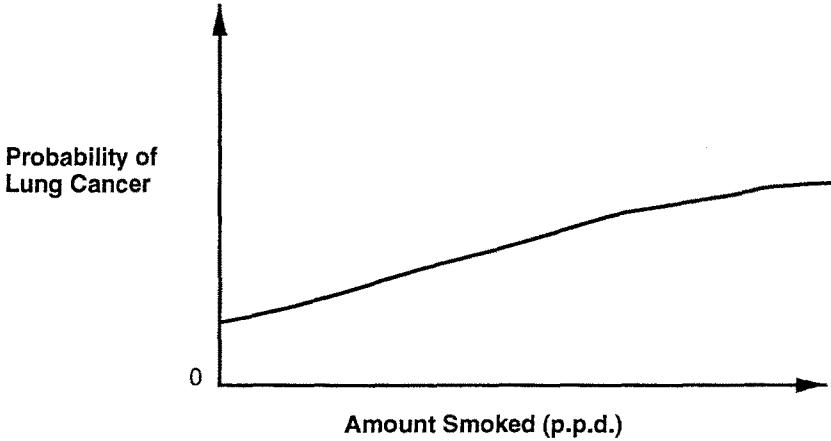


Figure 2. A more realistic picture.

claim would then express the inequality  $P(LC|ppd = 1 \wedge G_n) > P(LC|ppd = 0 \wedge G_n)$ , or more simply  $f(1) > f(0)$ . This gives us qualitative information about the shape of the function  $f$ . In the appropriate context, however, a different alternative might be made salient. We can imagine a country in which almost everyone smokes two packs per day, and in which the surgeon general admonishes citizens to cut back to one pack per day. In such a context, it might be natural to say that smoking (only) one pack per day inhibits lung cancer, describing the inequality  $f(1) < f(2)$ . In general, such causal claims are ambiguous unless an alternative is supplied, either explicitly or implicitly.

Although I have advocated a probabilistic theory of causation, I do not advocate it to the exclusion of other accounts, such as Salmon's (1984) mark-transmission theory or Lewis's (1986) counterfactual theory. With Skyrms (1984), I believe that paradigm examples of causation – baseballs crashing through windows, flipped switches turning on lights, stabbings causing the deaths of dictators – involve an “amiable jumble” of regularities, counterfactual dependence, statistical correlation, spatiotemporal processes, and so on. The probabilistic theory captures an important component of the concept of causation. More

precisely, it enables us to provide a taxonomy of causal relevance, to distinguish 'promoting' from 'inhibiting' causes.<sup>9</sup> This component of our concept of causation is certainly one that has important applications; I will mention two. First, information about promoting and inhibiting causes is relevant to making decisions. If one seeks to avoid lung cancer, one should participate in those activities which inhibit lung cancer, and eschew those which promote it. This idea is developed more fully in Cartwright (1979), who argues that the concept of a 'causal law', given a probabilistic explication, is needed to distinguish between effective and ineffective strategies. Second, the notions of causal promotion and inhibition are of importance within the theory of natural selection. Central to that theory is the distinction between the selection *of* a trait and the selection *for* a trait. In the evolution of hominids there has (presumably) been selection *for* the ability to produce speech; as a byproduct of this selection process, there has also been selection *of* a configuration of the trachea and esophagus which has left us prone to choking on food. In order for a trait to be selected *for*, it must have promoted reproductive success in those organisms that possessed it. Sober (1984) makes use of a probabilistic theory of causation in explicating this notion of 'selection *for*'.

In arguing against the claim that singular and general causal claims describe two distinct species of causal relation, then, I am not denying that different philosophical accounts of causation can complement, rather than compete with, one another. I am denying, however, that the distinction between the probabilistic theory sketched above and, say, Salmon's mark-transmission theory, corresponds directly to the distinction between singular and general causation. For example, I disagree with Sober (1985), who maintains that probabilistic theories provide the best account of general causation, while something like the mark-transmission theory is needed to account for singular causation.

### 3. AGAINST GENERALIZATION

One natural suggestion is that general causal claims such as

2. Smoking causes lung cancer

are generalizations over singular causal claims such as

1. David's smoking caused him to develop lung cancer;

much as the claim

3. All frogs are amphibians

is a generalization over claims of the form

4. If Kermit is a frog, then he is an amphibian.

There is, however, an important disanalogy between 2 and 3. In 3, the terms that flank the copula are pluralized, whereas in 2, the terms that flank 'causes' are not. A better analogue of 3 might be

5. Smokings cause lung cancers,

or slightly more perspicuously 'Episodes of smoking cause episodes of lung cancer'. In 5, the terms that flank the word 'cause' are plural event sortals, and the structure of the sentence suggests that it is describing singular causal relations that hold between events that fall under the sortals. We might use a construction akin to 5, for example, if we are interested in the frequency with which such causal episodes happen. For example, we might claim that

6. Every year, smoking causes thousands of cases of lung cancer.

This causal claim is plausibly analyzed as

7. (For every year  $y$ ) (There exist thousands of  $x$ ) ( $x$ 's smoking causes  $x$  to develop lung cancer in year  $y$ ).



By contrast, 2 appears to be asserting that a causal relation holds, *not* between events that fall under particular sortals, but between the sortals or event-types themselves.<sup>10</sup> Thus the surface grammar of 2 does not readily suggest analysis in terms of generalization over singular causal claims.

Surface grammar aside, let us try to understand 2 as a generalization over claims of singular causation. Here are some candidates for analysans:

8.  $(x)(x\text{'s smoking causes } x \text{ to develop lung cancer})$ ;
9.  $(x)(x \text{ smokes} \rightarrow x\text{'s smoking causes } x \text{ to develop lung cancer})$ ;
10.  $(\exists x)(x\text{'s smoking causes } x \text{ to develop lung cancer})$ .

8 is clearly too strong: 2 is true despite the fact that many people do not smoke. Similarly, 9 is too strong: not all smokers develop lung cancer, so not all smokers develop lung cancer because of their smoking.<sup>11</sup> 10 is true, like 2, but existential claims seem to be too weak to provide an analysis of general causation. To borrow an example from Wayne Davis (1988), even though Jim Fixx's last run caused his death, we would not say that jogging causes death (indeed, we would be inclined to say that it promotes good health).<sup>12</sup> More interestingly, there is also a sense in which the existential analysis is too strong. There can be true general causation claims that have no instances, such as:

11. Eating one kilogram of uranium 235 causes death.

11 is true in virtue of certain features of human physiology and the physics of nuclear chain reactions; however, no one has ever died in this unusual way, and it is unlikely that anyone ever will.

This game of example and counterexample could be continued, and the interested reader should consult the introduction of Eells (1991) for a more thorough discussion of the candidates discussed above (and others). What is wanted, however, is a general line of argument that shows that *no* such strategy can be successful. Such an argument is suggested by example 11. The general causation claim in 11 can be true because it does not entail anything about the instantiation of the

event-types figuring in it. This is precisely what we would expect if the probabilistic theory sketched in the previous section is correct as an account of general causation. According to that theory  $C$  is a cause of  $E$  if  $C$  raises the probability of  $E$  (relative to a certain alternative  $C'$ , and across a sufficient range of background contexts). But from this set of probabilistic inequalities, nothing strictly follows about the actual occurrences  $C$  and  $E$ .<sup>13</sup> For example, it is possible (although very unlikely) that smoking increases the probability of lung cancer even though in a certain population all and only the non-smokers develop lung cancer. In contrast, singular causal claims do entail that the named events occur; 1 would not be true unless David smoked and David developed lung cancer. Thus if general causal claims *could* be analyzed in terms of singular causal claims, general causal claims *should* have implications about the instantiation of the event-types named. By *modus tollens*, claims of general causation cannot be reduced to claims of singular causation.

I do not imagine that this line of argument is watertight, but only that it is a reasonable extrapolation from the responses to particular attempts to implement the generalization strategy. In any event, it does not by itself entail the independence of singular and general causation. It shows that if one starts with an account of singular causal claims, one will not be able to analyze general causal claims as generalizations over singular causal claims and end up with a theory of general causation that looks something like the probabilistic theory sketched in the previous section. In section 1, we saw that there were (at least) two strategies that one might follow in trying to characterize the relationship between singular and general causation. In this section, we have discussed only the generalization strategy, which begins with singular causation and then tries to analyze general causation in terms of it. But one could proceed in the other direction. Here, for example, is a fairly simple suggestion that has not been ruled out by the arguments that have been presented so far:

12. David's smoking caused him to develop lung cancer just in case (i) David smokes (or smoked), (ii) David developed lung cancer, and (iii) smoking causes lung cancer.

This is none other than the neo-Humean strategy. One advantage of this proposal is that it accounts for an important difference between claims of singular causation and claims of general causation that we have already noted; singular causal claims imply that their relata occur, while general causal claims carry no analogous implication. According to the neo-Humean strategy, this implication is tacked on in moving from the general to the singular level.

#### 4. THE MISHAP AT REICHENBACH FALL

The central type of argument for the independence of singular and general causation invokes examples of 'making it the hard way.'<sup>14</sup> These examples present rather direct arguments against the neo-Humean proposal introduced in the previous section, but they also provide a more subtle positive argument for the independence of the two causal levels. In these examples, a relation of negative causal relevance is alleged to hold between two event-types, while the events which instantiate those types stand in a relation of positive causal relevance. I will present an example due to I. J. Good (1961–2) (another example of this sort is found in Eells and Sober (1983); Eells (1991) abounds with them).<sup>15</sup>

In Good's story, Sherlock Holmes is walking through the valley at Reichenbach Fall,<sup>16</sup> while his nemesis, Moriarty, is poised on top of a cliff overlooking Holmes. Moriarty has perched a boulder on the edge of the cliff, so that when he pushes it carefully it will have a ninety percent chance of killing Holmes. Holmes' intrepid companion Watson arrives atop the cliff just as Moriarty is about to push the boulder. Watson cannot see Holmes from his vantage point, so he is not able to push the boulder in such a way that it will be certain to miss Holmes. Nevertheless, Watson reasons that it is better for him to push the boulder in a random direction than to let Moriarty push it while aiming carefully. Acting quickly on this line of reasoning, Watson rushes forward and pushes the boulder off the cliff in such a way that Holmes' chance of dying is reduced to just ten percent. As fate would have it, the boulder crushes Holmes anyway. Although it decreased the chance of Holmes' death, Watson's pushing the rock *caused* Sherlock Holmes to die.

Let us call pushings of similar rocks in similar circumstances with similar force ‘pre-emptive pushings’, and call deaths of healthy men while strolling through similar valleys ‘deaths by crushing’.<sup>17</sup> At the level of general causation, a relation of negative causal relevance seems to hold:

13. Pre-emptive pushing prevents death by crushing.

Watson’s push was of the sort that made it *less likely* that Holmes would die. If the set-up were re-created, and the rock pushed many times – sometimes in Watson’s random manner, sometimes in Moriarty’s calculated manner – fewer deaths would result from the Watson-type pushes. Nonetheless, we are inclined to say:

14. Watson’s pushing the rock caused Holmes to die.

Thus at the level of singular causation, a relation of positive relevance seems to hold.<sup>18</sup> This type of example seems to refute the neo-Humean view suggested in the previous section.<sup>19</sup> Moreover, it presents a general challenge: ‘cause’ and ‘prevent’ are antonyms, so how are we to reconcile claims 13 and 14? The reconciliation that Good and others invite us to accept is that there are two very different types of causation in play: since general causation is independent of singular causation, there is no contradiction between 13 and 14. The standard probabilistic theory of causation seems to yield the right verdict at the level of general causation, but the wrong verdict at the level of singular causation. Thus some other theory of singular causation is needed.

This assessment is strengthened by an example due to Cartwright (1979). Nancy discovers poison oak in her garden, and she decides to spray it with a defoliant. The defoliant claims to be ninety percent effective, meaning that if properly used, the poison oak has a ninety percent chance of dying within a specified period of time – say one month. Let us assume that this claim is accurate. Let us assume, moreover, that if left untreated, the poison oak would have only a ten percent chance of dying within one month. Thus, spraying with this defoliant *decreases* the chances of survival for plants of this kind from ninety percent to ten

percent. As it happens, however, the plant survives. In this example, the probabilities are identical with those in Good's story, with Nancy's spraying being analogous to Watson's pushing, and the poison oak's surviving being analogous to Holmes' *dying*. (Try not to let this confuse you.) In describing the story, one could make the analogue of claim 13:

15. Spraying with ninety percent effective defoliant prevents the survival of healthy poison oak plants.

Yet one would not say:

16. Nancy's spraying the poison oak with ninety percent effective defoliant caused it to survive.

16 is the analogue of claim 14. Although the two stories exhibit identical probability relations, they differ with regard to the causal relations that hold at the singular level. This threatens to undermine the very enterprise of providing a probabilistic account of singular causation,<sup>20</sup> suggesting that singular and general causation demand very different theories.

##### 5. RECONCILIATION

In order to reconcile the apparently contradictory claims 13 and 14, we must invoke the refined probabilistic theory of causation sketched in section 2: causal claims describe functions of probabilistic dependence, typically by contrasting the probability of the effect in the presence of alternative causes. Let us apply this principle to claim 14 above, which asserts that Watson's pushing the rock caused Holmes to die. This tells us that Watson's pushing the boulder *increased* the probability that Holmes would die, relative to some salient alternative. A natural alternative would be that in which no one pushes the boulder, so let us assume that 14 is implicitly being made against this alternative. Let *W* represent Watson's pushing the boulder, *N* represent no one's pushing it, and *H* represent Holmes' death. Then 14 would express the following inequality:

17.  $P(H|W) > P(H|N)$ .

There is, however, another salient alternative to Watson's pushing the boulder: the alternative in which Moriarty pushes the boulder. Let  $M$  stand for this alternative. Then the following inequality will also hold:

$$18. \quad P(H|W) < P(H|M).$$

This inequality could be expressed by the claim that Watson's pushing the rock was a negative cause of Holmes' dying, relative to Moriarty's pushing the rock; or more informally by

$$19. \quad \text{Holmes died } \textit{despite} \text{ Watson's pushing the rock.}^{21}$$

Note that we accept 19 as readily as we accept 14, and that 19 appears to conflict with 14 as directly as does 13. But the apparent conflict between 14 and 19 cannot be resolved by postulating the independence of singular and general causation, for both claims express causal relations at the singular level. Instead, claims 14 and 19 are reconciled by recognizing that they are implicitly being made relative to *different* alternatives.

The relativity of causal claims to an alternative cause holds at the level of event-types as well: pre-emptive pushing prevents death by crushing relative to malicious pushing (the type of event Moriarty was trying to instantiate); but pre-emptive pushing causes death by crushing relative to no pushing at all. In other words, the inequalities in 17 and 18 continue to hold if  $H$ ,  $M$ ,  $N$ , and  $W$  are interpreted as the appropriate event-types, rather than as singular events. But whereas claim 14 was used to describe an inequality such as 17, 13 was used to describe an inequality with the form of 18. The claims 13 and 14 appear to be in conflict, not because they involve different levels of causation, but because they are being made relative to different alternatives.<sup>22</sup>

Let us turn now to Cartwright's example involving the poison oak. In this example, we would accept the analogue of 19:

$$20. \quad \text{The poison oak survived despite its being sprayed with ninety percent effective defoliant;}$$

but not the analogue of 14:

16. Nancy's spraying the poison oak with ninety percent effective defoliant caused it to survive.

Where is the disanalogy with the Sherlock Holmes story: why do we accept 14 in that story, but not 16 in Cartwright's example? The disanalogy is that in Good's story, there is a salient alternative in which Holmes has virtually no chance of dying: the alternative where the rock is not pushed. 14 can be accepted because it can be understood as making a claim of causal relevance that is relative to this alternative. Cartwright's story lacks an analogue: there is no salient alternative in which the poison oak plant has virtually no chance of survival. It is easy enough to introduce one into the story, however. Let us suppose that Nancy chose between two different defoliants before spraying the poison oak. One of the defoliants was ninety-nine percent effective, but much more expensive than the weaker, ninety percent effective defoliant that Nancy eventually decided to use. Now that this alternative has been introduced, we are able to accept 16, for we can understand the contrast to be with the stronger defoliant. (Try saying 16 with the emphasis on 'ninety percent' to suggest this contrast.) Because Cartwright's original example did not provide this salient alternative, there was, after all, a disanalogy between the two examples. Once the disanalogy has been removed, the stubborn poison oak can be treated in the same way as the mishap at Reichenbach Fall.

Using the refined probabilistic theory of causation sketched above, we can resist the conclusion that singular and general causal claims describe independent species of causal relation. Claims of both types are used to furnish information about conditional probability functions. This solution requires that we refine our intuitions about claims 13 through 16 as well as claim 19; we must understand them as being implicitly made relative to alternative causes. But *any* solution to these problems, must involve *some* revision of our intuitions, for claims 14 and 19, each intuitively correct in its own right, seem to flatly contradict one another. Within these constraints, the solution that has been advocated is respectful of the pre-theoretic intuitions evinced by

the stories. Moreover, the solution is not *ad hoc*, but independently motivated by the problem of disjunctive causal factors. (See Hitchcock, 1993).

#### 6. OTHER PROPOSED SOLUTIONS

In this section, I wish to review some alternative attempts to resolve the problems raised by the mishap at Reichenbach Fall and similar cases. The three accounts that I will consider have an important feature in common: each maintains that singular causation and general causation demand different philosophical theories. Each of these proposals, like that of the previous section, requires that we refine some of our intuitions about what causes what in the examples of section 4. This refinement is not a bad thing, for those intuitions are awfully crude. I will, however, impose two desiderata upon such refinements. First, they should be independently motivated. Second, the alternative proposals should not destroy the intuitions that motivate them. Since it was our *unrefined* intuitions about the examples of section 4 that seemed to create the need for independent theories of singular and general causation in the first place, the theories that embrace the independence of the two causal levels should leave our unrefined intuitions relatively intact.

Sober (1985) argues that in light of examples such as those discussed in section 4, no probabilistic theory of singular causation can be made tenable. He suggests that the probabilistic theory provides a reasonable account of general causation, but recommends something like Salmon's (1984) theory of causal processes to account for causation at the singular level. Claim 13 is true in virtue of the probabilistic relations that hold between the event-types, whereas 14 is true because there is a causal process (the falling boulder) connecting Watson's push with Holmes' death.

The presence of such a process is no doubt part of the reason that we are willing to accept causal claims such as 14. We noted in section two that our ordinary notion of causation is an 'amiable jumble'; causal processes and relations of probabilistic relevance may well be distinct but legitimate pieces of that jumble. There is no reason, however, to think that the line between probabilistic accounts of causation and



process accounts of causation will also divide general causation from singular causation. In particular, the causal process theory fails to account for the disanalogy between claims 14 and 16. When Nancy sprayed the plant, there were physical processes that connected the spraying with the live plant one month later (indeed, the plant itself seems to be such a process.) The problem is that the causal process theory is unable to provide a *taxonomy* of causal relevance, and that seems to be precisely what is needed in this case. While Nancy's spraying the poison oak is causally relevant to its surviving, the plant survived *despite* her spraying it (relative to the most salient alternative). Thus there is a need for the taxonomy of causal relevance that the probabilistic theory is able to provide at the level of singular causation as well as at the level of general causation.

Eells (1991, chapter 6) agrees with Sober that singular causation and general causation demand different theories, but he offers a distinct *probabilistic* theory for singular causation. Both singular and general causation involve the increase of probabilities, but this increase takes very different forms in the two cases. In the case of general causation, causes increase the conditional probabilities of their effects in the manner described in section two. (Although Eells may not embrace the refinement recommended there.) In the case of singular causation, however, the probability increase involves the trajectory of probabilities through time. In order to illustrate this idea, let us elaborate on Good's story somewhat. Let us suppose that the boulder is perched unstably on a small promontory, so that when it is pushed, it will fall either to the left or to the right. If it falls to the left, then the probability that Holmes will die is 0.98; if it falls to the right, the probability that Holmes will die is 0.02. The rock is perched so precariously that any push that is strong enough to dislodge it might result in its falling either way. However, by carefully aiming, Moriarty will be able to push the rock in such a way that there is a 0.917 chance that it will fall to the left, resulting in an overall probability of 0.9 that Holmes will die. Watson, by trying to push the rock to the right, can decrease the probability of a leftward fall to 0.083. At the instant at which Watson begins to push the rock, the probability that Holmes will die becomes 0.1, *lower* than it was before Watson managed to beat Moriarty to the rock.<sup>23</sup> Very shortly after Watson pushed

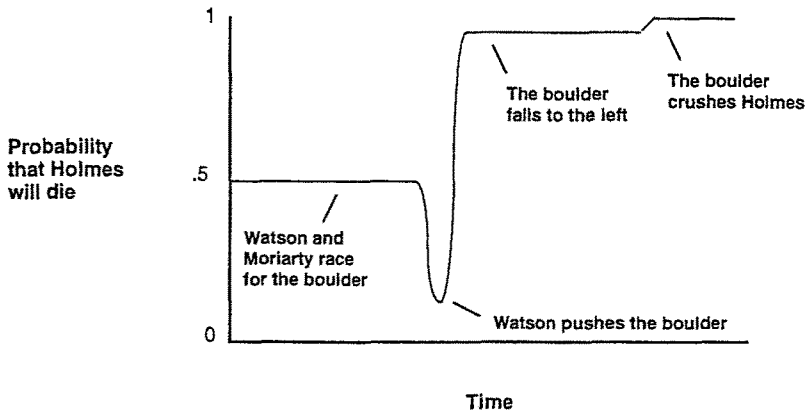


Figure 3. The mishap at Reichenbach Fall.

the rock, however, it teetered to the left (much to Watson's dismay) thus dramatically increasing the probability that Holmes would be crushed. (See Figure 3). It was this increase in probability that occurred shortly after Watson pushed the rock that renders Watson's pushing the rock a *positive cause* of Holmes' dying. By contrast, we reject claim 16 because there was no analogous increase in the probability of the poison oak's survival shortly after it was sprayed by Nancy. (We assume that the plant's probability of survival only gradually climbed back to 1.)

The problem with this account is that it puts exorbitant demands upon our intuitions. We were originally asked to accept claim 14 *without* the extra details laid out in the previous paragraph. If the boulder had fallen to the right (as Watson had intended) but, improbably enough, had killed Holmes anyway, then Eells would have it that Watson's pushing the rock did *not* cause Holmes to die. In that case, Eells would say that Holmes died *despite* Watson's pushing the rock; claim 19 would be true but 14 false. Now it is not illegitimate to suggest refinements of our intuitions in cases like these, but to the extent that our pre-theoretic intuitions become unrecognizable in the process, the philosophical analysis receives little or no support from those naïve intuitions. In this case, our untutored judgments are too imprecise to render the verdict that 14 would be true if the rock teetered to the left before killing Holmes, but that 19 would

be correct if it first teetered to the right; and thus these judgments are too imprecise to be telling in favour of Eells' theory. It is particularly troubling that Eells' theory would have us displace our naive intuitions about claims like 14, for it was those very intuitions that motivated the need for an independent theory of singular causation in the first place. And Eells' theory of singular causation – unlike the revised theory of probabilistic causation that was employed in the previous section – seems to have no independent motivation.

A third approach is suggested by the theories of David Lewis (1986) and I. J. Good (1961–2).<sup>24</sup> This approach can be motivated by observing that singular causation seems to be *transitive*, while general causation is not. To illustrate the failure of transitivity in the case of general causation, we may consider an argument that was presented by a group of conservative Princeton alumni opposed to the role of the university's health services in providing contraceptives to students:

Access to contraception causes sexual activity

Sexual activity causes pregnancy

∴ Access to contraception causes pregnancy

Even if we accept the dubious first premise, the conclusion is clearly false. The failure of transitivity is borne out by the probabilistic theory of causation. Let us suppose that the probability that a Princeton undergraduate will engage in sexual activity if contraceptives are not made available is 0.4, and that the probability is 0.5 if contraceptives are made available. In the absence of sexual activity, there is of course no chance of pregnancy, but let us suppose that the probability of pregnancy conditional upon sexual activity is 0.05 if contraception is used, and 0.7 if it is not. (These probabilities might refer to a couple having sexual intercourse regularly over the course of a year.) According to these probabilities, the two premises of the above argument are true: the probability of sexual activity is greater conditional upon access to birth control, and sexual activity increases the probability of pregnancy both in the presence and in the absence of contraceptive use. But there is another relationship to be considered: access to contraception will increase the probability of contraceptive use (for those who are sexually

active), let's say from 0.4 to 0.8. Now we can calculate the probability of pregnancy conditional on access to contraception, and on its negation: it turns out that access to contraception *decreases* the probability of pregnancy from 0.176 to 0.09. Access to contraception *prevents* pregnancy by promoting a strong inhibitor of pregnancy, namely contraceptive use.<sup>25</sup>

Many believe, however, that singular causation is transitive. Perhaps, then, transitivity should be built into a probabilistic theory of singular causation. *C* will be a cause of *E*, not merely if *C* raises the conditional probability of *E*, but also if there is some sequence  $D_1, D_2, \dots, D_n$ , such that *C* raises the probability of  $D_1$ ,  $D_1$  raises the probability of  $D_2$ ,  $\dots$ , and  $D_n$  raises the probability of *E*.<sup>26</sup> Each of these probabilities must be calculated by holding fixed the previous members of the chain.<sup>27</sup> We may illustrate this approach by again elaborating on the story of our intrepid detective and his faithful companion. Let us suppose now that there are *three* directions in which the rock might fall after being pushed: left, right and straight. If it falls to the left or right, the probability that Holmes will die is 0.98 or 0.02, as before. If, however, the rock falls straight off the promontory, the probability of crushing will be 0.3. In this version of the story, the probability that the boulder will fall to the right, conditional upon Watson's pushing the rock, is 0.714; the probability that it will fall straight is 0.286, and there is zero probability that it will fall to the left. Conditional upon Moriarty's pushing the rock, the probability that it will fall to the left is 0.882, straight, 0.118, and right, negligible. Again these numbers are such that the probability for Holmes to die is 0.1 conditional upon Watson's pushing the rock, and 0.9 conditional upon Moriarty's pushing the rock. Let's suppose that Watson pushes the rock, which falls straight over the promontory and kills Holmes. Watson's pushing the rock increased the probability that it would fall straight, so this forms one link in the causal chain. Given that Watson pushed the rock, it had to fall either straight, or to the right; conditional on Watson's pushing the rock, its falling straight raised the probability that Holmes would die, so this forms the second link. Watson's pushing the rock caused Holmes to die, because it caused the rock to fall straight, which in turn caused Holmes to die. (Note that in this version of the story, Eells'

theory would render the opposite conclusion, since the probability of Holmes' death shortly after the time of the push was only 0.3.)

But this theory also places high demands upon our intuitions. For example, if the rock had fallen to the right after Watson pushed it, and then (improbably enough) killed Holmes, the theory would render the verdict that Watson's pushing the rock did *not* cause Holmes to die. Similarly, in the earlier elaboration of the story, where the rock could only fall to the left or to the right, Watson's push would not have been a cause of Holmes' death no matter which way the boulder fell. This distinction – between cases where Watson's pushing the rock does cause Holmes to die and those where it does not – receives no support from our intuitions. A second difficulty with this theory is that it does not provide an account of the meaning of the 'despite' claim in 19, let alone a resolution of the apparent conflict between 14 and 19.

Our intuitions about claims like 13, 14, and 19 are too unstable to provide knock-down refutations of theories that violate them, but this very instability threatens to undermine the premise upon which these three proposals are built. The intuitions elicited by the examples of section 4 are supposed to press us to admit that we *cannot* apply the same probabilistic theory of causation to both singular and general causation; ironically, the account of those examples that strains our intuitions least – that sketched in section 2 – *does* apply the same probabilistic theory of causation to both singular and general causation.

#### 7. BIG SPACES, LITTLE SPACES

Are we done yet? We have argued that causal claims, at least insofar as they describe promoting and inhibiting causes, can be understood as describing conditional probability functions of a certain sort. This account applies at both the singular and general level. Singular causal claims describe probability relations between singular events (and also state that the named events occurred), while general causal claims describe formally analogous relations between generic event-types. But can we say more about the relationship between the two sets of probability relations?

Practitioners of probability theory are often interested in relations that hold between two (sets of) probability spaces. Many interesting probability spaces are constructed from other probability spaces. Here is a fairly elementary example. Suppose that we have a coin, which is flipped many times, and we are interested in constructing a probabilistic model of the various possible outcomes. Idealizing, we will assume that the coin is flipped infinitely often. The outcome space  $\Omega$  will consist of sequences of  $H$ 's and  $T$ 's, representing sequences of outcomes of heads and tails. Let  $X_1, X_2, \dots$  be a sequence of random variables such that for  $\omega \in \Omega$ ,  $X_i(\omega)$  is the  $i$ th term in the sequence  $\omega$ , representing the result of the  $i$ th flip of the coin in the sequence of flips represented by  $\omega$ . Thus the event<sup>28</sup>  $X_8 = T$  will represent tails on the eighth flip; formally, it is the set of all sequences  $\omega$  that have a  $T$  in the eighth position. In this model, probabilities such as  $P(X_8 = T)$  – the probability of tails on the eighth flip – are well defined. Events such as  $X_8 = T$  are analogous to singular events, since they represent the outcomes of *particular* flips of the coin. It makes no sense in this model to talk of the probability of tails *simpliciter* – the probability function  $P$  is simply not defined on events such as  $T$ , which are analogues of generic event-types. However, we can define *distribution functions* which are defined on the events  $H$  and  $T$ .  $\mu_i$  is the *distribution* of the variable  $X_i$  if  $\mu_i(T) = P(X_i = T)$  (and likewise for  $H$ ). The distribution functions are themselves probability measures, so we now have probability measures that assign values to the events  $H$  and  $T$ . If we were to say that  $\mu_8(T) = 0.5$ , however, we would still be interpreted as saying something about the probability of tails *on the eighth flip*. But suppose that the variables  $X_1, X_2, \dots$  are *identically distributed*, i.e., that the distribution functions  $\mu_1, \mu_2, \dots$  are identical. Let us call the common distribution function ' $\mu$ '. The function  $\mu$  genuinely assigns probability values to the generic events  $H$  and  $T$ . Moreover,  $\mu$  is systematically related to the original probability function  $P$ , such that information about the assignments made by  $\mu$  can also provide information about the assignments made by  $P$ . Thus, although, the original probability function  $P$  is not defined on generic events such as  $T$  and  $H$ , attributions of probabilities to such generic events may still be understood as providing partial information about the function  $P$ .

Since the function  $P$  is defined on a much richer algebra of events than is  $\mu$ , we will call  $P$  together with its corresponding outcome space and  $\sigma$ -field a ‘big’ probability space, and  $\mu$  together with its outcome space and  $\sigma$ -field a ‘little’ probability space. Because  $P$  has a certain symmetry – its variables are identically distributed – it is possible to partially represent  $P$  in terms of the little probability space  $\mu$ . If the variables  $X_1, X_2, \dots$ , are *independent*, as well as identically distributed, then it is possible to express all of the values of  $P$  in terms of the values of  $\mu$ . Such relationships between big and little probability spaces are often of deep philosophical interest. For example, de Finetti ([1937] 1964) showed that if a sequence of variables on a big probability space satisfies the condition of *exchangeability* (which is stronger than identical distribution, but weaker than identical distribution plus independence), then that big probability space could be represented as a weighted average of simpler probability spaces. This result has important consequences both for the foundations of probability, and for confirmation theory. In an important negative result, Arthur Fine (1982) showed that a version of the ‘no hidden variable’ theorem for distant quantum correlations could be couched within the framework of big and little probability spaces: he showed that distribution functions representing certain experimental outcomes fail to meet a consistency condition that is necessary for their joint representation within a big probability space.<sup>29</sup>

The apparatus of interrelated big and little probability spaces allows us to construct a model for the relationship between singular and general causation. Let  $\langle \Omega, \mathcal{F}, P \rangle$  be a probability space, and let  $X_1, X_2, \dots$  be a sequence of random variables on this space. Also, let  $E_1, E_2, \dots$  be a sequence of events in  $\mathcal{F}$ . We are interested in conditional probability functions of the form  $f_i(x) = P(E_i | X_i = x)$ ; as we noted in section two, these are the sorts of conditional probability functions that are described by claims of positive or negative causal relevance. Suppose that the indexing numbers correspond to individuals: David = 1, Mary = 2, and so on.  $E_i$  represents  $i$ 's developing lung cancer, and  $X_i = n$  represents  $i$ 's smoking  $n$  packs of cigarettes per day. Thus claims about the causal relevance of  $i$ 's smoking for  $i$ 's developing lung cancer describe the conditional probability function  $f_i(x) = P(E_i | X_i = x)$ . (For now, we will assume that the other causally relevant factors are

being held fixed in the background.) For example, the claim that David's smoking two packs per day caused him to develop lung cancer (relative to his not smoking) conveys that  $P(E_1|X_1 = 2) > P(E_1|X_1 = 0)$  (and of course it also conveys that David smoked two packs per day and that David developed lung cancer).

Now we want to treat the conditional probability functions  $f_i$  in a fashion analogous to the way that we treat the random variables in the coin-tossing example; i.e., we want to find the analogue of a distribution for a conditional probability function  $f_i$ . It is shown in the appendix that it is possible to construct a probability space  $\langle \Omega_i, \mathcal{F}_i, P_i \rangle$ , such that  $E$  is an event in  $\mathcal{F}_i$  and  $X$  is a random variable on  $\Omega_i$  that is measurable with respect to  $\mathcal{F}_i$ , and such that  $f_i(x) = P_i(E|X = x)$ . (The proof is fairly complex, but the idea is straightforward.) The function  $P_i$  is then the analogue of a distribution function. Moreover, if the conditional probability functions  $f_i(x) = P(E_i|X_i = x)$  are all identical, then the *same* space can be constructed for each one; that is, we can construct a space  $\langle \Omega', \mathcal{F}', P' \rangle$  such that  $P'(E|X = x) = P(E_i|X_i = x)$  for all  $i$ .<sup>30</sup> If the causal relevance of smoking for lung cancer is the same from individual to individual, then it is possible to provide information about the probability space  $\langle \Omega, \mathcal{F}, P \rangle$  by describing the related probability space  $\langle \Omega', \mathcal{F}', P' \rangle$ . And one way to provide information about the latter space is by describing the conditional probability function  $f(x) = P'(E|X = x)$ , which is what we do when we make claims about the causal relevance of smoking for lung cancer. Stricter symmetry conditions on  $P$ , such as analogues of exchangeability or independence plus identical distribution, will lead to tighter connections between  $\langle \Omega, \mathcal{F}, P \rangle$  and  $\langle \Omega', \mathcal{F}', P' \rangle$ .

An advantage of this model is that it predicts the grammatical form of general causal claims. It was noted in section 3 above that the paradigmatic claim of general causation has the form 'smoking causes lung cancer', and not 'smokings cause lung cancers', as one might expect if general causal claims were some kind of generalization. In the model sketched here, general causal claims describe the probability function  $P'$ , which is defined over generic events such as  $E$  and  $X = 1$ . These events lack subscripts, just as the event-types smoking and lung



cancer lack reference to any individual episodes of smoking and lung cancer.

The assumption that  $f_i(x) = P(E_i|X_i = x)$  will be independent of  $i$  may seem implausible; surely some people may smoke a great deal with virtually no chance of lung cancer while others are genetically prone to suffer from lung cancer regardless of their smoking habits. It was assumed, however, that the relevant background conditions were held fixed. Let us now relax that assumption. For each  $i$ , there will be a partition of possible background contexts into cells  $\{G_i^1, G_i^2, \dots\}$ . Then we would not expect  $P(E_i|X_i = x)$  to be independent of  $i$ , for some individuals might be more likely to possess additional risk factors. But it is plausible that the conditional probability functions  $f_{ij}(x) = P(E_i|X_i = x \wedge G_i^j)$  will be independent of  $i$ . To simplify the notation, let  $\mathbf{Y}_i$  be a *random vector* such that  $\mathbf{Y}_i = (x, j)$  is identical to  $X_i = x \wedge G_i^j$ . Then the necessary condition is that  $f_i(\mathbf{y}) = P(E_i|\mathbf{Y}_i = \mathbf{y})$  be independent of  $i$ . This condition says that individuals will have identical probabilities of suffering from lung cancer, conditional upon their smoking the same amount, and being alike with respect to other causally relevant factors. Thus if David and Mary are alike with respect to all of the factors that are causally relevant to lung cancer, David will not have a different probability of lung cancer simply in virtue of being David, rather than Mary. This condition is nothing more than a probabilistic version of the principle of the uniformity of nature.<sup>31</sup> Since the proof of the appendix can be generalized to include random vectors, we can construct a probability space  $\langle \Omega', \mathcal{F}', P' \rangle$  so that  $P'(E|\mathbf{Y} = \mathbf{y}) = P(E_i|\mathbf{Y}_i = \mathbf{y})$  for all  $i$ , whenever the uniformity condition holds.

Note that this probabilistic version of the principle of uniformity of nature is not to be confused with Eells' context-unanimity condition. That condition demanded that  $P(E|C \wedge G_i) > P(E|\sim C \wedge G_i)$  for every cell  $G_i$  of the relevant partition. This formulation does not incorporate the refinement advocated in section 2; a liberal reformulation might be that the conditional probability functions  $f_j(x) = P(E|X = x \wedge G_j)$  are qualitatively similar – say by increasing over a certain interval – for all  $j$ . One can formulate the context-unanimity requirement in the language of big spaces and little spaces. Let  $\langle \Omega, \mathcal{F}, P \rangle$

and  $\langle \Omega', \mathcal{F}', P' \rangle$  be as above. Suppose, moreover, that the little probability space satisfies the further constraint that for every  $G_j$ ,  $f_j(x) = P'(E|X = x \wedge G_j)$  increases over the interval  $0 \leq x \leq 1$ . Then we can define an ensemble of even littler probability spaces. These will consist of probability measures  $P_j$  defined on a  $\sigma$ -field that does *not* include the events  $G_1, G_2, \dots$ , and such that  $P_j(E|X = x) = P'(E|X = x \wedge G_j)$ . Each of these littler probability spaces includes probability relations between smoking and lung cancer, but excludes any reference to background conditions. The set of these little spaces can be treated as representing a vague probability relation, just as vague states of opinion are represented by sets of probability functions in the theory of subjective probability (see, e.g., van Fraassen, 1989, chapter seven). Causal claims, such as that smoking causes lung cancer, might then be interpreted in terms of supervaluations over this set of littler probability spaces. For example, the claim that smoking one pack of cigarettes per day causes lung cancer (relative to not smoking) would be true if  $P_j(E|X = 1) > P_j(E|X = 0)$  for all  $j$ . I do not put forward this suggestion as an endorsement of the context-unanimity view, but to show that the context-unanimity view can be recognized as a concrete proposal about how general causal claims are used to describe little probability spaces.

## 8. CONCLUSION

In Hitchcock (1993) I advocated a refinement to traditional probabilistic theories of causation. Two features of that refinement prove useful in accounting for the relationship between singular and general causation. First, causal claims are contrastive in nature, so that positive and negative causation are relativized to alternative causes. This observation enables us to deflect a class of arguments for the independence of singular and general causation. Second, causal claims are used to provide information about a certain type of probabilistic structure: the conditional probability function. Singular and general causal claims describe different probability spaces, but they describe formally similar structures on those spaces. Claims of singular causation describe con-

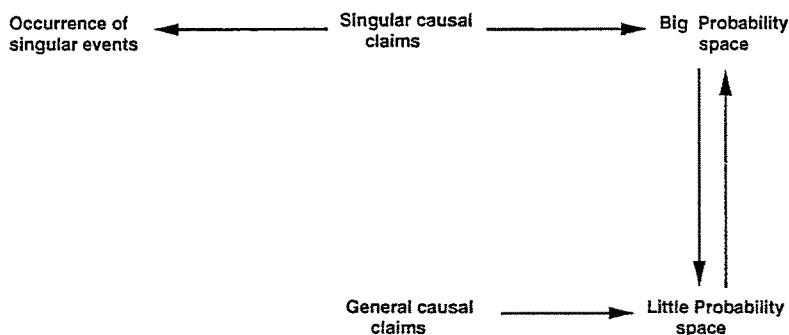


Figure 4. Singular and general causation.

ditional probability functions on spaces whose events are indexed to specific individuals (and also assert that the physical events represented by those mathematical events occurred), whereas claims of general causation describe such functions on spaces whose events are not so indexed. But the probability spaces described by claims such as 1 and 2 are very closely related, so that information about one of them automatically serves to describe the other (see Figure 4).

In the introduction, we mentioned two strategies for accounting for the relationship between singular and general causation. One strategy was to begin with singular causation, and define claims of general causation as generalizations over claims of singular causation. This strategy seems to be untenable on the general grounds that singular causal claims imply the occurrence of events of a certain sort, while general causal claims do not. The second strategy was to begin with general causation, and to treat claims of singular causation as describing instantiations of general causal laws. Is this the strategy that I have followed? Not exactly. Note that on the account developed in Section 7, the sorts of probability structures described by general causal claims will only exist when some sort of uniformity condition holds. This is entirely appropriate: if Mary and David are alike with respect to all other relevant factors, but their probabilities for developing lung cancer conditional upon smoking certain amounts differ wildly, then we are simply not able to talk about *the* causal relevance of smoking for lung

cancer. It follows that singular causation is not parasitic upon general causation, for the former can exist in the absence of the latter.<sup>32</sup> The strategy I have followed does not begin with one level of causation and then try to define the other; it treats both levels in parallel, and then tries to describe the relationship between them. I suspect the presupposition that one must proceed from the bottom up or from the top down is a vestige of determinism; once we move to the indeterministic context, neither approach looks very satisfactory. Fortunately, probability theory provides the resources for a new approach to this old question.

## APPENDIX

*Theorem:* Let  $f$  be a function from the real numbers to the unit interval that is measurable with respect to  $\mathcal{R}$ , the Borel sets in  $\mathbf{R}$  (these are the members of the smallest  $\sigma$ -field containing all of the open intervals). In other words, for every Borel set  $B$  in  $[0, 1]$ ,  $f^{-1}(B) = \{x \in \mathbf{R} : f(x) \in B\}$  is a Borel set in  $\mathbf{R}$ . Then there exists a probability space  $\langle \Omega, \mathcal{F}, P \rangle$  such that: there is a random variable  $X$  on  $\Omega$  that is measurable with respect to  $\mathcal{F}$ , and an event  $E$  in  $\mathcal{F}$ , such that  $f$  is a version of  $P(E|X)$ .<sup>33</sup>

*Proof:* Let  $\Omega$  be  $\mathbf{R} \times [0, 1]$ . If  $\omega \in \Omega$ , then  $\omega = (x, y)$  with  $x \in \mathbf{R}$  and  $y \in [0, 1]$ . Let  $\mathcal{F}$  be the restriction of  $\mathcal{R}^2$  to  $\Omega$ ; that is, let  $\mathcal{F}$  be the smallest  $\sigma$ -field containing all sets of the form  $B_1 \times B_2$ , where  $B_1$  is a Borel subset of  $\mathbf{R}$ , and  $B_2$  is a Borel subset of  $[0, 1]$ . Define  $X: \Omega \rightarrow \mathbf{R}$  by  $X((x, y)) = x$ . Let  $E = \{(x, y) : x \in \mathbf{R}, 0 \leq y \leq f(x)\}$ .  $\lambda$  will be Lebesgue measure on the  $[0, 1]$  interval (so that  $\lambda([a, b])$  will be  $b - a$ ). Now suppose that  $g$  is a density function on  $\mathbf{R}$ , so that  $\int_{\mathbf{R}} g dx = 1$ , and let  $\mu$  be the resulting measure on Borel sets, so that  $\mu(H) = \int_H g dx$ .  $P$  will be the product measure  $\mu \times \lambda$ ; in particular,  $P(B_1 \times B_2) = \mu(B_1)\lambda(B_2)$ . We will show first that  $E \in \mathcal{F}$ ; then we will show that  $f$  is a version of  $P(E|X)$ .

Let  $A_{ik} = \{(x, y) \in \Omega : y < (k+1)/2^i, k/2^i \leq f(x) < (k+1)/2^i\}$  for  $i = 1, 2, \dots$  and  $k = 0, 1, \dots, 2^i - 1$ . Then  $A_{ik} = f^{-1}([k/2^i, (k+1)/2^i]) \times [0, (k+1)/2^i]$ ; since each of these sets is a Borel set,  $A_{ik}$  is in  $\mathcal{F}$ . Let  $A_{i2^i} = f^{-1}(1) \times [0, 1]$ , so this set is in  $\mathcal{F}$  also. Now let  $E_i = \cup_{k=0,1,\dots,2^i} A_{ik}$ . Since  $E_i$  is a finite union of sets in  $\mathcal{F}$ ,  $E_i$  is in

$\mathcal{F}$ . We will now show that  $E = \cap_i E_i$ . It will follow that  $E$  will be a countable intersection of sets in  $\mathcal{F}$ , and therefore that  $E$  is in  $\mathcal{F}$ . The idea is that each  $E_i$  is a union of ‘rectangles’ that contains  $E$ , and that each successive  $E_i$  is made of finer rectangles, so that the sequence of  $E_i$  converges on  $E$ .

Suppose  $(x, y) \in E$ . Then, by the definition of  $E$ ,  $0 \leq y \leq f(x) \leq 1$ . If  $f(x) = 1$ , then  $(x, y) \in A_i 2^i \subseteq E_i$  for all  $i$ , and hence  $(x, y) \in \cap_i E_i$ . So suppose that  $f(x) < 1$ . We can represent  $f(x)$  in binary form as  $f(x) = \sum_{j=1,2,\dots} k_j 2^{-j}$ .<sup>34</sup> Pick an arbitrary  $i$ ; for this  $i$  we will have  $(\sum_{j=1,\dots,i} k_j 2^{i-j})/2^i \leq f(x) < ((\sum_{j=1,\dots,i} k_j 2^{i-j}) + 1)/2^i$ . Letting  $k = (\sum_{j=1,\dots,i} k_j 2^{i-j})$ , this shows that  $(x, y) \in A_{ik} \subseteq E_i$ . Since the choice of  $i$  was arbitrary,  $(x, y)$  is in every  $E_i$ , and hence in  $\cap_i E_i$ .

Suppose  $(x, y) \notin E$ . If  $y < 0$ , or  $y > 1$ , then it is obvious that  $(x, y)$  is not in  $\cap_i E_i$ . Then let us suppose that  $f(x) < y \leq 1$ ; say  $y = f(x) + \epsilon$ , for some  $\epsilon > 0$ . Choose  $i$  so that  $2^{-i} < \epsilon$ . For this  $i$ , let  $f(x) = \sum_{j=1,2,\dots} k_j 2^{-j}$ , and let  $k = (\sum_{j=1,\dots,i} k_j 2^{i-j})$  as above. Then  $k/2^i \leq f(x) < (k + 1)/2^i$ . Since  $2^{-i} < \epsilon$ ,  $y > (k + 1)/2^i$ , and so  $(x, y) \notin A_{ik}$ . And since  $k/2^i \leq f(x) < (k + 1)/2^i$ ,  $(x, y) \notin A_{ik'}$  for all  $k' \neq k$  as well. Thus  $(x, y) \notin E_i$  and  $(x, y) \notin \cap_i E_i$ . This shows that  $E = \cap_i E_i$ .

We now show that  $f$  is a version of  $P(E|X)$ . Note first that  $P(X \in H) = P(H \times [0, 1]) = \mu(H)\lambda([0, 1]) = \mu(H)$ . The first equality follows from the definition of  $X$  and the second from the definition of product measure. It follows that  $\mu$  is the distribution of  $X$ .  $f$  is measurable with respect to  $\mathcal{R}$  by hypothesis, so it remains only to show that

$$P(E \wedge X \in H) = \int_H f(x) d\mu(x).$$

Let  $E' = E \wedge X \in H$ . According to a variant of Fubini’s theorem,<sup>35</sup> we have

$$P(E') = \int_{\mathbf{R}} \lambda(\{y : (x, y) \in E'\}) d\mu(x).$$

Now  $E'$  can also be written  $\{(x, y) : x \in H, y \leq f(x)\}$ . For  $x$  in  $H$ ,  $\{y : (x, y) \in E'\} = \{y : 0 \leq y \leq f(x)\}$ , and so  $\lambda(\{y : (x, y) \in E'\}) = f(x)$ ; if  $x \notin H$ , then  $\{y : (x, y) \in E'\} = \phi$ , and  $\lambda(\{y : (x, y) \in E'\}) = 0$ . Thus, as a function of  $x$ ,  $\lambda(\{y : (x, y) \in E'\}) = \chi_H(x)f(x)$ . Thus

$$\begin{aligned} P(E') &= \int_{\mathbf{R}} \chi_H(x)f(x)d\mu(x) \\ &= \int_H f(x)d\mu(x) \end{aligned}$$

QED.

*Remark.* The idea is that  $f(x)$  is a common conditional probability function on a more complex probability space:  $f(x) = P^\circ(E_i|X_i = x)$ , where the values of these functions are independent of  $i$ . The theorem shows that  $f$  can be represented as a single conditional probability function  $f(x) = P(E|X = x)$  on an abstract probability space. The theorem can be easily generalized to the case where  $f$  is a measurable function on  $\mathbf{R}^n$ .

#### NOTES

\* Earlier versions of this paper were presented at the University of Georgia, Massachusetts Institute of Technology, University of Michigan, University of Pittsburgh, and Rice University. I am grateful to audience members at all of these institutions for many helpful comments.

<sup>1</sup> Although Humean it is not Hume's because it leaves out the psychological element so important in Hume's own account of causation.

<sup>2</sup> A  $\sigma$ -algebra is similar to a Boolean algebra, except that a  $\sigma$ -algebra is closed under countable unions and intersections, and not merely under finite unions and intersections.

<sup>3</sup> Readers should consult any text on probability (such as Billingsley, 1986) for definitions of these concepts. A brief overview is provided in the appendix of Hitchcock (1993).

<sup>4</sup> I do not presuppose that this partition is countably infinite.

<sup>5</sup> Although in Eells' account one can limit the range of cells by specifying a population relative to which a particular causal claim is being made.

<sup>6</sup> Dupré suggests that a cause must raise the probability of its effect in a 'fair sample', where fairness, requires 'lack of bias with respect to independent causally relevant factors' (Dupré, 1984, p. 173). This means that we should compare samples where the cells  $G_1, G_2, \dots$  are represented in the same proportions, giving rise to the following account:  $C$  is a cause of  $E$  if  $\sum_i P(E|C \wedge G_i) \times P(G_i) > \sum_i P(E|\sim C \wedge G_i) \times P(G_i)$ . (Note that this is not equivalent to the inequality  $P(E|C) > P(E|\sim C)$ .) This is equivalent to saying that  $P(E|C \wedge G_i) > P(E|\sim C \wedge G_i)$  in a weighted majority of the  $G_i$ , where each  $G_i$  is weighted by  $P(G_i)$  and  $|P(E|C \wedge G_i) - P(E|\sim C \wedge G_i)|$ .

<sup>7</sup> Humphreys (1989) offers a theory of singular causation in terms of single case probabilities that is very similar to that of Eells in its formal structure.

<sup>8</sup> I mean here smoking *precisely* one pack per day, and not smoking *at least* one pack per day. Some readers have suggested to me that in common usage 'smoking one pack of cigarettes per day' more strongly suggests the latter.

<sup>9</sup> I argue in Hitchcock (1993) that the taxonomy is much richer than the dichotomous one suggested here.

<sup>10</sup> I do not here offer a theory of events and event-types. The approach I favor is that of Bennett (1988), according to which both facts and events are named by nominalized sentences. According to Bennett, causation is primarily a relation between facts, and only derivatively a relation between events. I agree with Bennett on this point, but have retained the language of events here for its greater familiarity. Within this framework, I take it that an event-type or its factive counterpart would be named by a nominalized *predicate*.

<sup>11</sup> Indeed, it is not even true that the majority of smokers develop lung cancer. This creates problems for Carroll's (1991) proposal, which would require that the frequency of living cancer among smokers be high. (I am assuming here that the frequency of cases in which one person's smoking causes *another person* to develop lung cancer does not make up the difference.)

<sup>12</sup> A full assessment of this example hinges on the issue of context-unanimity, which we have vowed to avoid.

<sup>13</sup> Unless one is a strict frequentist in one's interpretation of probabilities.

<sup>14</sup> Wes Salmon has used this expression in discussion, although I have not found it in his writings. The expression comes from the dice game 'craps', in which a player sometimes has to roll a specific number on two dice in order to win. A player who needs a ten, for instance, and succeeds by rolling two fives, rather than the more likely combination of six and four, is said to have 'made it the hard way'.

<sup>15</sup> There is a second type of argument which is closely related. These arguments aim to show the inadequacy of probabilistic theories of causation by describing examples in which causes lower the probabilities of their effects. The most famous example of this sort is Deborah Rosen's story of the miraculous birdie, related in Suppes (1970);

Salmon (1984) discusses similar examples. These counterexamples always involve cases of singular causation, however, so at best they can show that the probabilistic theory is inadequate as a theory of singular causation. Since the probabilistic theory is plausible as a theory of general causation, these arguments can also be construed as supporting the independence of singular and general causation.

<sup>16</sup> I take some liberties in setting Good's story at this locale. Doyle's Holmes nearly met his end there at the hands of the evil Moriarty. (It was Doyle's intent to kill Holmes, but the public demanded that he be brought back, and thus the events at Reichenbach Fall were re-told so as to make Holmes' death only apparent.) Since Hans Reichenbach was the first to propose a probabilistic theory of causation, it seems only fitting that the story take place at a site that shares his name.

<sup>17</sup> These names are intended as abbreviations only; I do not mean to imply that they pick out the casually relevant features of the events in question.

<sup>18</sup> I realize that intuitions about 13 and 14 are not univocal, but bear with me; the account developed later will be able to explain the confusion of our intuitions.

<sup>19</sup> Note that the neo-Humean cannot say that we have not yet found the general causal law under which 14 is subsumed. The events named in 14 do instantiate the event-types named in 13, so if the neo-Humean view is correct, a relation of negative causation must hold at the singular level as well. This points to another problem that faces the neo-Humean view (but only if it is combined with a Davidsonian theory of events): if we allow probabilistic causal laws as well as deterministic ones, then the same pair of singular events may instantiate two competing causal laws. If smoking causes heart attacks and exercise prevents them, what are we to say about a smoker who exercises? This problem is similar to the problem of epistemic relativity faced by Hempel's (1965) Inductive-Statistical model of explanation.

<sup>20</sup> Cartwright's intention was only to show that causation could not be reduced to probabilities without remainder.

<sup>21</sup> Eells (1991) uses 'despite' to express negative causal relevance at the singular level, and this seems to accord reasonably well with normal English usage. Note that we cannot use 'prevent' in this context. Just as 'cause', when used at the singular level, implies that the two named events occurred, 'prevent' at the singular level implies that one of the named events occurred, and the other did not.

<sup>22</sup> And we are now able to understand why our intuitions were muddy to being with: claims 13 and 14 are ambiguous unless an alternative is specified.

<sup>23</sup> Let us assume that when Watson arrived, he and Moriarty had an equal chance to push the rock. Then the overall probability of Holmes' dying was 0.5.

<sup>24</sup> The theories of Good and Lewis are quite different. Lewis (1986) offers a probabilistic generalization of his counterfactual theory of causation. Good offers a measure of the extent to which one event caused another. What the two theories have in common is an emphasis on the role of causal chains.

<sup>25</sup> See Eells and Sober (1983) and Eells (1991, chapter 4) for extensive discussion of questions of transitivity of general causation.



<sup>26</sup> Menzies (1989) shows that this condition is not sufficient, and offers a refinement, but the details need not concern us here.

<sup>27</sup> For this reason, Salmon (1984) calls this approach the method of successive recon-  
ditionalization.

<sup>28</sup> Recall that 'event' is a measure-theoretic term for an element of a  $\sigma$ -field over which a probability function is defined. The coincidence of terminology is both fortuitous and dangerous, but I will not deviate from standard usage here.

<sup>29</sup> For a similar proof dealing with a hypothetical example, see Suppes and Zanotti (1981).

<sup>30</sup> One could do this by taking the restriction of  $P$  to the  $\sigma$ -field generated by  $E_i$  and sets of the form  $X_i \in H$  for some  $i$ , but this would raise doubts about whether the resulting probability space is really an abstract representation of the common conditional probability functions.

<sup>31</sup> Indeed, one is tempted to say that the uniformity condition holds automatically, since if David and Mary's chances for lung cancer still differ after holding fixed their level of smoking together with other background factors, one can argue that the chances are themselves causally relevant properties with respect to which they differ. Alternately, one who interprets probabilities as hypothetical frequencies within reference classes will not even be able to make sense of the possibility that John and Mary's probabilities for lung cancer differ when they are alike in all relevant respects. For now, however, I am contented to let the uniformity condition remain as an empirical hypothesis.

<sup>32</sup> As remarked in the previous footnote, however, this scenario cannot arise for one whose account of single-case probability is parasitic upon statistical frequencies. Within such a framework, the account developed above is a version of the neo-Humean strategy. Thus the question of how to interpret the probabilities is an important one, but it must wait for another day.

<sup>33</sup> See Breiman (1968) or another advanced text on probability for the definition of a generalized conditional probability function.

<sup>34</sup> If more than one representation is possible, choose the one that is last in lexical order. For example,  $1/2$  may be represented as  $0.1000 \dots$  or as  $0.0111 \dots$ ; we would use the former binary expansion.

<sup>35</sup> See Billingsley (1986, ch. 18).

#### REFERENCES

- Anscombe, G. E. M. (1981) 'Causality and Determination.' In: *Metaphysics and the Philosophy of Mind*, pp. 133–147. Oxford: Basil Blackwell.  
 Bennett, J. (1988) *Events and Their Names*. Indianapolis: Hackett.  
 Billingsley, P. (1986) *Probability and Measure* (second edition). New York: Wiley.  
 Breiman, L. (1968) *Probability*. Reading, MA: Addison-Wesley.

- Carroll, J. (1991) 'Property-level Causation?' *Philosophical Studies* 63: 245–270.
- Cartwright, N. (1979) 'Causal Laws and Effective Strategies.' *Nous* 13: 419–437.
- Davidson, D. (1980) 'Causal Relations.' In: *Essays on Actions and Events*, pp. 149–162. Oxford: Clarendon Press.
- Davis, W. (1988) 'Probabilistic Theories of Causation.' In: Fetzer (1988), pp. 133–160.
- De Finetti, B. [1937] (1964) 'Foresight: Its Logical Laws, Its Subjective Sources.' Translated by H. Kyburg in H. Kyburg and H. Smokler (eds) *Studies in Subjective Probability*, pp. 93–158. New York: Wiley.
- Dupré, J. (1984) 'Probabilistic Causality Emancipated.' In: French et al. (1984), pp. 169–175.
- Eells, E. (1988) 'Probabilistic Causal Interaction and Disjunctive Causal Factors.' In: Fetzer (1988), pp. 189–209.
- Eells, E. (1991) *Probabilistic Causality*. Cambridge, U.K.: Cambridge University Press.
- Eells, E. and Sober, E. (1983) 'Probabilistic Causality and the Problem of Transitivity.' *Philosophy of Science* 50: 35–57.
- Fetzer, J. (ed) (1988) *Probability and Causality: Essays in Honor of Wesley C. Salmon*. Dordrecht: Reidel.
- Fine, A. (1982) 'Hidden Variables, Joint Probability and the Bell Inequalities.' *Physical Review Letters* 48: 291–295.
- French, P. A., Uehling, T. E. and Wettstein, H. K. (eds) (1984) *Midwest Studies in Philosophy IX: Causation and Causal Theories*. Minneapolis: University of Minnesota Press.
- Good, I. J. (1961–2) 'A Causal Calculus I–II.' *British Journal for the Philosophy of Science* 11: 305–318; 12: 43–51.
- Hempel, C. G. (1965) 'Aspects of Scientific Explanation.' In: *Scientific Explanation and Other Essays in the Philosophy of Science*, pp. 331–496. New York: Free Press.
- Hitchcock, C. R. (1993) 'A Generalized Probabilistic Theory of Causal Relevance.' *Synthese* 97: 335–364.
- Hume, D. [1748] (1955) *An Enquiry Concerning Human Understanding*. Indianapolis: Bobbs-Merrill.
- Humphreys, P. (1989) *The Chances of Explanation: Causal Explanations in the Social, Medical, and Physical Sciences*. Princeton: Princeton University Press.
- Lewis, D. K. (1986) 'Causation,' plus postscripts. In: *Philosophical Papers, Volume II*, pp. 159–213. Oxford: Oxford University Press.
- Menzies, P. (1989) 'Probabilistic Causation and Causal Processes: A Critique of Lewis.' *Philosophy of Science* 56: 642–663.
- Salmon, W. C. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Skyrms, B. (1980) *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale University Press.
- Skyrms, B. (1984) 'EPR: Lessons for Metaphysics.' In: French et al. (1984), pp. 245–255.

- Sober, E. (1984) *The Nature of Selection*. Cambridge, Mass.: MIT Press.
- Sober, E. (1985) 'Two Concepts of Cause.' In: P. Asquith and P. Kitcher (eds) *PSA 1984, Vol. II*, pp. 405–424. East Lansing: Philosophy of Science Association.
- Suppes, P. (1970) *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.
- Suppes, P. and Zanotti, M. (1981) 'When Are Probabilistic Explanations Possible?' *Synthese* 48: 191–199.
- Van Fraassen, B. C. (1989) *Laws and Symmetry*. Oxford: Clarendon Press.

*Department of Philosophy*  
*Rice University*  
*Houston TX 77251*  
*USA*