A Probabilistic Analysis of Causation

Luke Glynn

ABSTRACT

The starting point in the development of probabilistic analyses of token causation has usually been the naive intuition that, in some relevant sense, a cause raises the probability of its effect. But there are well-known examples both of non-probability-raising causation and of probability-raising non-causation. Sophisticated extant probabilistic analyses treat many such cases correctly, but only at the cost of excluding the possibilities of *direct* non-probability-raising causation, failures of causal transitivity, action-at-a-distance, prevention, and causation by absence and omission. I show that an examination of the structure of these problem cases suggests a different treatment, one which avoids the costs of extant probabilistic analyses.

1 Introduction

Quantum mechanics, at least on standard 'collapse' interpretations (e.g. Copenhagen and GRW), seems to show that the fundamental dynamics of our world are probabilistic. Many of the special sciences also give probabilistic laws for events falling under their purview. Statistical Mechanics and Mendelian genetics explicitly formulate such laws, and probabilistic functional laws are encoded in the models of economists and meteorologists.

The fact that ours seems to be a probabilistic world–both at fundamental and non-fundamental levels–has done little to persuade philosophers, scientists, or laypeople that it is not a causal world. On the contrary, the apparent compatibility of causality with probabilistic indeterminism has motivated attempts (by philosophers) to develop probabilistic analyses of causation. Amongst those who have advanced such analyses are Good ([1961a], [1961b], [1962]), Reichenbach ([1971]), Suppes ([1970]), Lewis ([1986e]), Menzies ([1989]), Eells ([1991]), Mellor ([1995]) and Kvart ([1991], [1994a], [1994b], [1997], [2004a], [2004b]).

Luke Glynn

2 A Naive Probabilistic Analysis, Two Objections and a Refinement

The starting point in the development of probabilistic analyses of causation has usually been the

naïve intuition that, in some relevant sense, a cause raises the probability of its effect. The

standard way of cashing this out (it shall soon be seen that there are others) is in terms of an

inequality between conditional probabilities. This gives rise to the following naïve probabilistic

analysis of (positive) token causation.

Take any two distinct, actual events c and e.[1] Let C and E be binary variables that take

the values 1 or 0 according respectively to whether or not c and e occur.[2] Then a naïve

probabilistic analysis says that c is a cause of e iff $C = 1$ and $E = 1$ (that is, both c and e actually

occur) and inequality (1) holds:

$$P(E = 1|C = 1) > P(E = 1|C = 0) \tag{1}$$

This inequality says, in effect, that the probability of e's occurrence conditional upon c's

occurrence is greater than the probability of e's occurrence conditional upon c's non-occurrence.

In what follows, I shall sometimes speak of $C = 1$ raising the probability of $E = 1$, or even

of C's raising the probability of E (or C's causing E, or C's occurring). Since C and E are really

binary variables representing the occurrence or non-occurrence of token events, I should be

understood in these cases as meaning that the event whose occurrence is represented by $C = 1$

raised the probability of the event whose occurrence is represented by $E = 1$ (or that the former

caused the latter, or that the former occurred, and so on).

A Probabilistic Analysis of Causation

There are well-known problems for a naive probabilistic analysis of causation in terms of the obtaining of inequality (1). As one might expect, counterexamples come from two directions. On the one hand, examples are given of causes that fail to raise the probability of their effects, demonstrating that straightforward probability-raising is not *necessary* for causation. On the other hand, there are examples of probability-raising non-causes, demonstrating that straightforward probability-raising is not *sufficient* for causation either.

One type of non-causal probability-raising is that which obtains between *independent effects of a common cause*. So, for example, let C (= 1) represent a fall in the reading of a certain barometer, and let E (= 1) represent the occurrence of a subsequent storm. Then inequality (1) holds, and the naive analysis yields the incorrect result that C is a *cause* of E.

Another type of probability-raising non-causation arises where an *effect raises the probability of its cause*. Indeed, it is a straightforward implication of the probability calculus that, wherever inequality (1) holds, so also does inequality (2):

$$P(C = 1|E = 1) > P(C = 1|E = 0) \tag{2}$$

Wherever a cause raises the probability of its effect, the effect also raises the probability of its cause. The naive analysis therefore has the disastrous implication that each effect is a cause of its causes.

The problems multiply: the naive analysis yields the result that any case that, by its lights, is one of causation is a case of bi-directional causation. So any instance of probability-raising non-causation (such as the case of the falling barometric reading and the storm) not only

becomes one of causation, but one of bi-directional causation. The naive analysis as it stands is hopeless.

Advocates of probabilistic analyses are sensitive to these problems. The standard response is to hold fixed certain background conditions in evaluating the probabilistic relationship between C and E.[3] Suppose that $B_1 \dots B_n$ are variables representing the relevant background conditions (it is not required that each of these variables be binary–they could represent continuous quantities, for example). Let $\boldsymbol{B}$ be the set $\{B_1 \dots B_n\}$, and let $\bar{b}_1 \dots \bar{b}_n$ be the actual values taken by the members of $\boldsymbol{B}$. Finally, let $\bar{\boldsymbol{B}}$ be the proposition that $B_1 = \bar{b}_1, \dots,$ and $B_n = \bar{b}_n$. Then it might be held that C is a cause of E iff:

$$P(E = 1 | C = 1 . \bar{\boldsymbol{B}}) > P(E = 1 | C = 0 . \bar{\boldsymbol{B}}) \tag{3}$$

That is, C is a cause of E iff C raises the probability of E once the relevant background is held fixed.

Of course everything now turns upon what counts as relevant background. If it is specified to include the other *causes* of E,[4] then since common causes screen off their independent effects from one another,[5] we will avoid generating spurious causal relations between independent effects of a common cause (but not between an effect and its cause, since a cause needn't be probabilistically independent of its effect conditional upon the causes of the cause). But nor will it be reductive, because of the appeal to causal facts in the specification of what must be held fixed.

An alternative suggestion is the following: suppose t_C is the time at which C occurred, then the proposition $B_i = \bar{b}_i$ concerns the relevant background iff relative to t_C, $B_i = \bar{b}_i$ is an

historical proposition (that is, B_i is a variable representing the obtaining or not of some state of affairs prior to t_C). This specification of the background to be held fixed makes reference not to causal facts, but to temporal facts.[6] Yet since any common causes of C and E will have occurred by t_C they will constitute part of the fixed background. Likewise if E is not an effect of C, but rather a *cause* of C, then E itself will already have occurred by t_C and will constitute part of the fixed background. Either way, inequality (3) will not hold.[7]

This suggestion works because of the correspondence between the direction of causation and the direction of time. If it is possible for this correspondence to break down–that is, if simultaneous or backwards-in-time causation are possible–some other specification of the relevant background conditions to be held fixed will have to be given. But there is no reason to suppose that, within the context of the conditional probability approach to causation, this implicit analysis of the direction of causation in terms of the direction of time could not be replaced with any more adequate analysis that might be discovered (one proposal shall be mentioned shortly).

A potential difficulty with the present suggestion is that which arises if the history up until (just before) time t_C *determines* that $C = 1$. If this is the case, then $P(C = 0|\overline{B}) = 0$ and the orthodox (Kolmogorov [1933]) axiomatization of the probability calculus leaves the RHS of inequality (3) undefined. The inequality therefore fails to hold and we get the result that C is not a cause of E (for any choice of E)–a potential case of non-probability-raising causation (see Lewis [1986e], pp. 178-9). In response I note only that this alleged problem is lessened or eliminated altogether if (a) the incompatibility between determinism and non-trivial objective chance is rejected (Eagle [forthcoming]; Frigg and Hoefer [forthcoming]; Glynn [2010]; Hoefer [2007]; Loewer [2001]); *and/or* (b) the Kolmogorovian analysis of conditional probability as a

ratio of unconditional probabilities is rejected (Hájek [2003a], [2003b], [2007]). I am in fact sympathetic to both position (a) and position (b).

But perhaps the problem could be avoided altogether by cashing out the relevant probability-raising relation not in terms of an inequality between conditional probabilities (which, as standardly understood, go undefined when the probability of the proposition conditioned upon is equal to zero), but instead in terms of counterfactuals whose consequents concern the unconditional probability of the putative effect (which have well-defined truth-values even when the probability of the antecedent is zero). This is the approach of Lewis ([1986e]) and Menzies ([1989]).[8] The idea is that, rather than explicating probability-raising between $C = 1$ and $E = 1$ in terms of inequality (3), it should instead be explicated as follows. Suppose that $C = 1$ and that the unconditional probability, $P(E = 1)$, is equal to x. Then C raised the probability of E in the relevant sense iff the following counterfactual is true: 'If it had been the case that $C = 0$, then $P(E = 1)$ would have been less than x.' (Lewis in fact requires that the counterfactual probability of $E = 1$ be lower than x by 'a large factor'.)

In addition to avoiding the supposed problem of probability 1 causes, a *prima facie* benefit of this counterfactual approach is that it seems to obviate the need for assuming correspondence between the causal and temporal orders, thus making room for the possibility of backwards-in-time causation (Lewis [1986b], pp. 50-1). Whereas the conditional probability approach (as developed above) explicitly holds historical background fixed by conditioning upon it, the counterfactual approach holds relevant background fixed implicitly in virtue of the non-backtracking nature of the relevant counterfactuals. Lewis's ([1986b], [1986c]) semantics are intended to secure this non-backtracking property, not by brute stipulation, but (in the indeterministic case) by grounding it in a contingent asymmetry of *quasi-miracles* (Lewis,

[1986c], p. 61)–via his similarity metric over possible worlds. It is the contingence of this asymmetry that makes room for the possibility of backwards causation.

Yet one might reasonably harbour reservations about this counterfactual approach to probabilistic causation. One might, for example, be suspicious of Lewis's method of 'reverse-engineering'[9] a similarity metric over possible worlds from just those non-backtracking counterfactuals required to come out true if the counterfactual analysis is to succeed.[10] One reason for suspicion is that there *do* seem to be some true backtracking counterfactuals (see e.g. Hall [2000], pp. 218-9). These backtrackers seem to be made true by causal facts ('if the bomb had exploded, the fuse would have been lit' sounds true because the lighting of the fuse is causally necessary for the bomb's explosion). Perhaps we should take a general lesson from this and expect that an adequate semantics for the foretrackers must also make reference to causal facts (thus making trouble for the counterfactual analyst's pretentions to reductivity). This worry is particularly difficult to allay because of another fault of Lewis's similarity metric: it is altogether too vague to allow us readily to derive testable predictions about the truth values of particular counterfactual conditionals (cf. Hitchcock [2001b], p. 378).

It is also worth observing that one of the supposed advantages of the counterfactual over the conditional probability approach–that the former is compatible with backwards-in-time causation–does not result from any essential feature of either approach. Rather, it is a consequence of historical accident in the way the two approaches have been developed. It is clearly open to someone who prefers the use of conditional probabilities to analyse causal order directly in terms of whichever contingent asymmetry of our world is supposed to break the symmetry of counterfactual dependence (according to Lewis, an asymmetry of quasi-miracles),

thus reaping the benefit of logical consistency with backwards-in-time causation without the detour via chancy counterfactuals and their possible-worlds semantics.

In any case, though I shall here use the more traditional conditional probability approach together with the assumption that the temporal and causal orders coincide, I think that the main points that follow could be captured using a counterfactual notion of probability-raising and/or some alternative analysis of causal direction. Consequently, what follows should be of interest to those who don't share my tastes on these matters.

3 Non-Probability-Raising Causation

In the previous section, it was seen that a simple modification of the naive probabilistic analysis (conditioning upon historical background[11]) allows us to deal with two types of probability-raising non-causation: independent effects of a common cause and effects that raise the probability of their causes.

It also helps deal with some examples of non-probability-raising causation. Consider Rosen's ([1978], pp. 607-8) example (discussed by Suppes [1970], p. 41) of a golfer who badly slices an approach shot, with the result that it hits a tree and rebounds into the hole for a spectacular birdie. The striking of the tree by the ball seems intuitively to be a cause of the birdie, but surely the probability of the birdie given that the ball hits the tree is lower than in the absence of the tree-impact (since in the absence of the tree-impact there would presumably have been some probability of the golfer's having hit the ball truly and its travelling a normal trajectory toward the hole). Not so if we condition upon history up until a time just before the impact. By this time, the ball has already been sliced and is going well wide of the hole. *Given*

that this is the case, the tree-impact actually raises the probability of the birdie because it changes the trajectory of the ball (cf. Salmon [1980], p. 69, [1984], pp. 199-200).

Yet there are examples both of non-probability-raising causation and of probability-raising non-causation that are not handled by holding fixed historical background. Let us start with an examination of three cases of the former. Each is structurally different, and their various structures seem to exhaust those found in the literature (with an exception to be discussed in §8.2, below). The first is due to Hesslow ([1976], p. 291), the second is my own, and the third is due to Humphreys ([1989], pp. 41-2).[12]

Example 1: Thrombosis

Studies have shown that consumption of contraceptive pills can cause thrombosis. But pregnancy is a relatively potent cause of thrombosis and consumption of contraceptive pills reduces the risk of pregnancy. Suppose that Jane engages in unprotected sex but takes contraceptive pills which prevent her from becoming pregnant. Sometime later she suffers thrombosis.

Because of the negative relevance of birth control pills to pregnancy, it might be the case that overall Jane's consumption of the pills fails to raise the probability of her suffering thrombosis:[13]

$$P(Thrombosis = 1|Pills = 1) \leq P(Thrombosis = 1|Pills = 0) \qquad (4)$$

But it was acknowledged that the consumption of contraceptive pills *can* cause thrombosis. And suppose that this is such a case: Jane's consuming the pills causes her to suffer thrombosis (our evidence might be the existence of a complete biochemical process connecting the two events.) We therefore have a case of causation without probability-raising.

Example 2: Bridge Collapse

Billy and Suzy are contemplating whether to cross a rickety bridge over a stream. Billy adopts the following policy: he'll wait and see what Suzy does; if Suzy decides not to cross the bridge, Billy will cross it. If on the other hand Suzy decides to cross the bridge, Billy will flip a coin and cross the bridge just in case the coin lands heads.

Billy is heavier than Suzy; there is a moderate chance that the bridge will collapse under Suzy's weight alone, a high chance that it will collapse under Billy's weight alone and a very high chance that it will collapse under their combined weight. In fact Suzy decides to cross the bridge ($SX = 1$), Billy tosses the coin, the coin lands heads, Billy follows Suzy onto the bridge and the bridge collapses.

It seems that Suzy's crossing is a (partial) cause of the collapse. Nevertheless, because of the negative probabilistic relevance of Suzy's crossing to Billy's crossing (a more efficacious potential cause), the probabilities could well be such that inequality (5) holds:

$$P(Collapse = 1|SX = 1) \leq P(Collapse = 1|SX = 0) \tag{5}$$

If so, then although Suzy's crossing was a cause of the bridge's collapse, it failed to raise its probability.

Example 3: Medicine

Patient has a potentially fatal condition. There is one known drug that can treat it. This drug is expensive and has unpleasant side-effects. Doctor has just three courses of action available to her: she can give Patient a high dose, a low dose or no dose at all. The probability of Patient's recovery is 0.9 given a high dose, 0.4 given a low dose, and 0.1 given no dose. Doctor is equally disposed to follow each of the three courses of action: she does each with a probability 1/3. In fact, Doctor administers a low dose (*Low* = 1), and Patient recovers (*Recovery* = 1).

From the probabilities specified in the example, it is straightforwardly calculated that the following inequality obtains:

$$P(Recovery = 1|Low = 1) = 0.4 \, < 0.5 = P(Recovery = 1|Low = 0) \tag{6}$$

So Doctor's administering a low dose lowers the probability of Patient's recovery. Nevertheless, it is perhaps plausible to regard it as a cause.

Attempts have been made to produce a sophisticated probabilistic analysis that can accommodate cases of non-probability-raising causation, such as the three just described. For instance, Good ([1961a], [1961b]), Menzies ([1989], p. 656) and Lewis ([1986e], p. 179) analyse causation not in terms of probability-raising, but in terms of the ancestral of that relation. Their analyses allow that, where *C does not* raise the probability of *E*, *C* may nevertheless be a cause of *E* provided that there is a sequence $\langle C, D_1, \ldots, E \rangle$ such that each member of this sequence raises the probability of its immediate successor.

It seems that this proposal may well allow adequate treatment of the three examples described. Although *Pills* doesn't straightforwardly raise the probability of *Thrombosis*, there may be some *Intermediate* on the biochemical process connecting *Pills* to *Thrombosis* such that *Pills* raises the probability of *Intermediate* and *Intermediate* raises the probability of *Thrombosis*.[14] Similarly, although Suzy's crossing does not straightforwardly raise the probability of *Collapse*, it does raise the probability of *both Billy and Suzy crossing together*, which in turn straightforwardly raises the probability of *Collapse*. Again, although the low dose does not raise the probability of *Recovery*, it raises the probability of there being some of the active agent in Patient's blood stream, and this raises the probability of *Recovery* (since the comparison is with the alternative where there is *no* active agent present).

But there are difficulties with this solution. For one thing, it is not clear that such a sequence will *always* exist: there might be cases of 'direct' non-probability-raising causation. Salmon ([1980], p. 65) gives an example of such a phenomenon which will be discussed further in §8.2. In addition, since ancestral relations are transitive, analysing causation in terms of the ancestral of the probability-raising relation has the effect of–to quote Hitchcock ([2001a], p. 275)–'rendering causation transitive by definition.' But since there are well-known examples of

apparent failures of causal transitivity,[15] it seems that the resulting accounts will be too liberal. Transitivity shall be discussed further in §8.3.

This proposed solution to the problem of non-probability-raising causation is what Salmon ([1980], p. 64, [1984], p. 195) calls *the method of interpolated causal links*. He distinguishes two other potential responses ([1980], pp. 64, 68-70, [1984], pp. 194-201), which he dubs *the method of more detailed specification of events* and *the method of successive reconditionalization*. The latter solution combines the requirement that historical background be conditioned upon so as to exclude Rosen-type examples (a proposal that has already been adopted) with a weakening of the naive analysis so that the ancestral of probability-raising is sufficient, and probability-raising not necessary, for causation. Consequently, just like the method of interpolated causal links, it runs into difficulties with direct non-probability-raising causation and failures of causal transitivity.

The method of more detailed specification of events, on the other hand, is a distinct solution and is that which is adopted by Rosen herself ([1978], p. 608). The idea is that, by giving a more detailed specification of an apparently non-probability-raising cause *C*, it might be revealed after all to be a probability-raiser of its effect *E*. So, for example, whilst Jane's consumption of birth control pills does not seem to raise the probability of thrombosis, when we specify that this was a consumption of birth control pills by someone with such-and-such a physiology (where that physiology in combination with birth control pills is especially conducive to thrombosis), it may turn out that this event raised the probability of thrombosis after all.[16]

The problem is that there is no obvious justification for the assumption that the empirical details will turn out as Rosen supposes (indeed they don't in the example to be discussed in §8.2). Salmon ([1984], pp. 194-5) points out that such an assumption is on a par with the view–

which 'amount[s] to no more than a declaration of faith' (Salmon [1980], p. 50; cf. Anscombe [1971])–that causal interactions could be shown to be deterministic, if only they were specified closely enough. In any case, it is not such unknown details that lead us to make the judgements we do in the three examples. And it would surely be preferable to have an account of non-probability-raising causation that reconstructs our intuitive judgements from the facts that ground those judgements.[17]

Because of the inadequacy of traditional responses, I prefer to look elsewhere for a solution to the problem of non-probability-raising causes. In order to successfully do so, it will be necessary to examine the structure of these cases in a bit more detail. One convenient method for doing so is with the use of graphs.[18]

4. Graphical Representation of Cases of Non-Probability-Raising Causation

A graph[19] is an ordered pair $\langle V, E \rangle$ where V is a set of vertices and E a set of edges. The elements of E are pairs of vertices. In the *directed graphs* that shall be used here, these pairs are *ordered*. The ordered pair $\langle V_1, V_2 \rangle$ represents the directed edge $V_1 \rightarrow V_2$. Where there is a directed edge from V_1 to V_2, V_1 is said to be a parent of V_2, and V_2 a child of V_1. A *directed path* from V_1 to V_n in a graph G is a sequence of vertices beginning with V_1 and ending with V_n, such that for each pair of vertices V_i, V_j, such that V_j succeeds V_i in the sequence, $\langle V_i, V_j \rangle \in E$. An *ancestor* of a vertex V is any vertex W such that there is a directed path from W to V. A *descendant* of a vertex V is any vertex W such that there is a directed path from V to W. An *acyclic* path is one that contains no vertex more than once and a *directed acyclic graph* is a directed graph (i.e. a graph containing only directed edges) that contains no directed cyclic paths.

I shall make use only of directed acyclic graphs here, and since I shall not make use of the notion of an undirected path, I shall sometimes just use the term 'path' (or 'route') as short for 'directed path'. The vertices in the graphs used will always represent variables. The variables I shall use will mostly be binary, taking value 1 or 0 according, respectively, to whether some event (or event alteration[20]) occurs or fails to occur. But multi-valued variables shall sometimes be used and it would be possible to use variables to represent continuous quantities such as air pressure or the reading of an analogue barometer.

The graphs used here shall be assumed to satisfy the Markov and Minimality Conditions (but not Faithfulness) for the probability distributions they represent (these being objective chance distributions resulting from conditioning upon historical background). The Markov condition says that, for every variable W in V, the value taken by W is probabilistically independent of the values taken by its non-descendants in V given the values taken by its parents. Minimality says that no edge can be removed from the graph without the resulting subgraph violating the Markov condition (in other words, each edge represents some conditional dependence relation).

Call variable V a *temporal antecedent* of variable W just in case V represents the occurrence or non-occurrence (or, in the non-binary case, the occurrence-in-some-degree) of an event or alteration that, if it did occur, would occur prior to that represented by W. Then the Markov and Minimality conditions will be satisfied by any graph such that if V_i, $V_j \in V$, then $\langle V_i, V_j \rangle \in E$ just in case $V_i \in T$ (the set of variables in V that are also temporal antecedents of V_j) *and* there exist values v for each variable in $T \backslash \{V_i\}$ such that, fixing the variables in $T \backslash \{V_i\}$ at the values v, the value taken by V_j probabilistically depends upon that taken by V_i (over some range of possible values for V_i and V_j).

Given this rule for drawing directed edges, the structure of Hesslow's thrombosis example can be represented by means of the following graph (see Hitchcock [2001b], p. 364 for a similar representation):
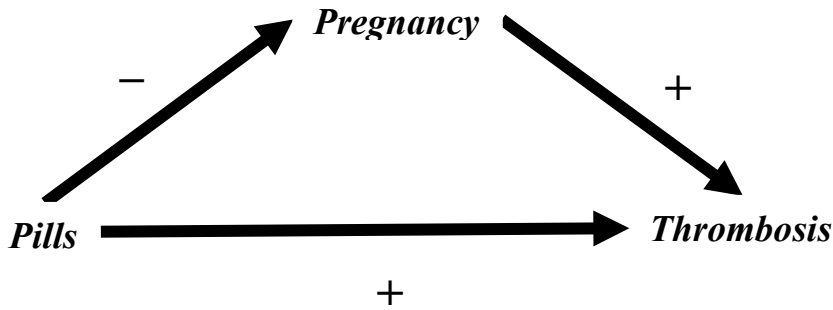

Figure 1

$Pills = 1$; $Pregnancy = 0$; $Thrombosis = 1$

The directed edge from *Pills* to *Pregnancy* indicates the probabilistic dependence of the value of the latter on that of the former. The directed edge from *Pregnancy* to *Thrombosis* indicates that the value of *Thrombosis* depends probabilistically upon the value of *Pregnancy* (for at least one value of *Pills*). The consequence of the existence of these two directed edges is the existence of an indirect path from *Pills* to *Thrombosis* (via *Pregnancy*) in the graph.

There is also a *direct* path from *Pills* to *Pregnancy*. This is because there is a value of *Pregnancy* such that, holding this value fixed, the value of *Thrombosis* depends probabilistically on the value of *Pills* (in fact this is true for both values of *Pregnancy*). That is to say *Pills* may have a probabilistic impact upon *Thrombosis* over and above that which it has in virtue of its probabilistic impact on *Pregnancy*.

The graph itself does not convey information about the actual values taken by the variables its vertices represent, nor about the nature of the probabilistic impact represented by its directed edges. I have therefore supplemented it by writing the actual values of variables

underneath and by annotating the edges with '+' or '−' labels. The latter annotation is possible only because the probabilistic relations happen to be unambiguous. If it were the case, for instance, that *Thrombosis* depended positively on *Pills* given *Pregnancy* = 1, but negatively given *Pregnancy* = 0, then the edge from *Pills* to *Thrombosis* would not be amenable to such labelling.

Note that the direct route from *Pills* to *Thrombosis* only indicates that there exists *at least one* value of *Pregnancy* such that, holding this value fixed, *Thrombosis* depends probabilistically on *Pills*. But it is the *actual* value of *Pregnancy* (= 0) in which we are particularly interested, since our concern is with the *actual* contribution made by *Pills* to *Thrombosis*. Holding *Pregnancy* fixed at its actual value, we can factor out the actual contribution of *Pills* to *Thrombosis* along the indirect route and so isolate the contribution along the direct route. We find that this is non-null:

$$P(Thromb. = 1|Pills = 1. Preg. = 0) > P(Tromb. = 1|Pills = 0. Preg. = 0) \qquad (7)$$

Because *Thrombosis* depends probabilistically upon *Pills* holding *Pregnancy* fixed at its actual value, we can say–borrowing some terminology from Hitchcock ([2001a], p. 286)–that the direct route is *active* (it would have been *inactive* if there were only *non-actual* values of *Pregnancy* for which *Thrombosis* depended on *Pills*). Using some more Hitchcock ([2001b], p. 362) terminology, we can say that *Pills* therefore has a *component effect* upon the value of *Thrombosis* along the direct route.[21] Because of the sign of the contribution made, this component effect is *positive*.

By contrast, *Pills* has a negative component effect on *Thrombosis* along the indirect route

running via *Pregnancy*. This is because *Pills* is negatively relevant to *Pregnancy* which, in turn

is positively relevant to *Thrombosis*. The value of *this* component effect would be isolated if we

interpolated an appropriate variable on the route from *Pills* to *Thrombosis* (this would have to be

a variable representing an event on the biochemical process connecting *Pills* to *Thrombosis* that

is not also on the process that goes via *Pregnancy*[22]), and held this fixed at its actual value.

The *net effect* (Hitchcock, ibid.) of *Pills* on *Thrombosis* is a function of these component

effects. *Pills* has a non-positive net effect upon *Thrombosis* (that is to say the former fails to raise

the probability of the latter overall) because the negative component effect along the indirect

path is at least as strong as the positive component effect along the direct path. In spite of this,

we judge that *Pills* was a cause of *Thrombosis* and this seems to be because of its positive

component effect.

The structure of *Bridge Collapse* can be represented by a graph similar to that used for
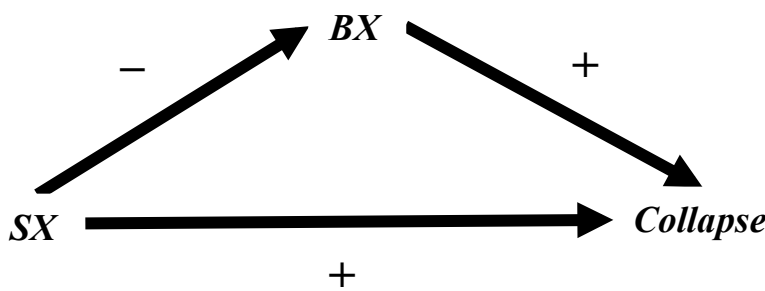
Thrombosis:

Figure 2

SX = 1; *BX* = 1; *Collapse* = 1

Again there are two paths (or 'routes') from cause to effect. One is a direct route, along which *SX* has a positive component effect. The other is an indirect route, running via *BX* (a variable representing Billy's crossing), along which *SX* has a negative component effect. This latter component effect is negative because the value of *SX* is negatively relevant to that of *BX* which is positively relevant to the value of *Collapse*. The case is one of non-probability-raising causation because the negative influence along the indirect route cancels out the positive influence along the direct route. It does so because of the strong positive relevance of *BX* to *Collapse*.

The only difference between this and *Thrombosis* is that in this case the intermediate variable on the indirect route takes the value 1 *despite* the negative relevance of *SX* to *BX* (the relation between the value of these two variables is, after all, only probabilistic and not deterministic). But, again, conditioning upon the value of the intermediate variable on the indirect route isolates the positive component effect along the direct route:

$$P(Collapse = 1|SX = 1. BX = 1) > P(Collapse = 1|SX = 0. BX = 1) \qquad (8)$$

Our intuitions about token causation seem to track this positive component effect: because of its positive component effect on *Collapse*, we judge that *SX* was a cause of *Collapse in spite* of its non-positive net (and negative component) effect.

Both *Thrombosis* and *Bridge Collapse* involve just *two* paths between cause and effect. But non-probability-raising causation can also occur in cases where there are *more* than two paths. A three-path example can be generated from *Bridge Collapse* by just adding another person, Amy, who like Billy will cross if Suzy doesn't and, if Suzy does, will flip a coin and

cross if the coin lands heads. (An *n*-path case can be generated by just adding *n* - 3 additional coin-flippers to the scenario alongside Billy and Amy.) This case can be represented by the following graph:
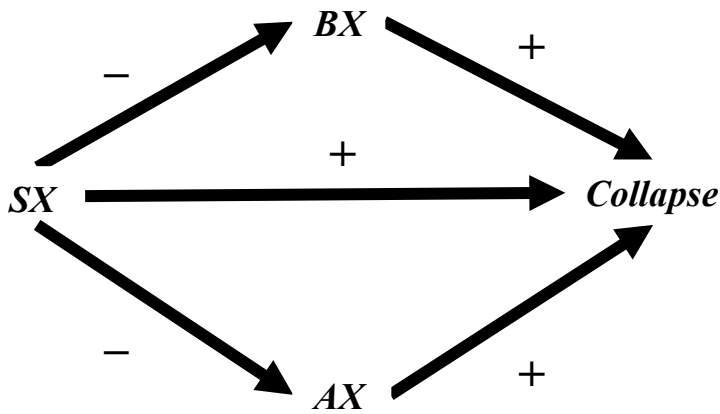

Figure 3

$SX = 1; BX = 1; AX = 1; Collapse = 1$

SX has a positive component effect on *Collapse* along the direct route, but a negative component effect on its value along each of the indirect routes, and an overall non-positive net effect. Suppose that its negative component effect along *each* of the indirect routes alone is enough to cancel out the positive component effect along the direct route (Billy and Amy are both much heavier than Suzy). Then holding fixed merely the value of the intermediate variable on *one* of these routes won't be enough to reveal the positive component effect of *SX* on *Collapse*. But conditioning upon the values of intermediate variables on *both* of these routes *will* reveal a relation of positive relevance:

$$P(Collapse = 1|SX = 1. AX = 1. BX = 1) > P(Collapse = 1|SX = 0. AX = 1. BX = 1) \quad (9)$$

(This would have worked just as well had either or both *AX* or *BX* taken value 0.)

The suggestion, then, is that all cases of non-probability-raising causation at least involve positive component effect. It is for this reason that they are regarded as cases of genuine, positive causation. In other words, our intuitions about (positive) token causation *track* positive component effect. Non-probability-raising causes fail to raise the probability of their effects only because of the existence of cancelling negative component effects along other routes.

To state the suggestion a bit more precisely: *C* is a positive token cause of *E* only if *C* has a positive component effect upon *E*. Where this is so, there will exist a (possibly empty) set **S** of variables (containing a variable on each of the routes from *C* to *E* that transmits a negative component effect) such that, when we hold fixed the value of each variable *S*∈**S** at its *actual value* $\bar{s}$, *C* raises the probability of *E*. In other words *C* is a positive token cause of *E* only if there is a set **S** of variables such that the following inequality holds (where $\overline{\textbf{\textit{S}}}$ is the proposition that each variable *S*∈**S** takes its actual value $\bar{s}$):

$$P(E = 1 | C = 1.\overline{\textbf{\textit{S}}}) > P(E = 1 | C = 0.\overline{\textbf{\textit{S}}}) \tag{10}$$

Call the set **S** a *Revealer of Positive Relevance* (*RPR*) for *C* and *E*.

Note that the parenthetical reference to *routes* from *C* to *E* in the passage immediately preceding inequality (10) is merely heuristic: the notion of an *RPR* is defined purely probabilistically and is not itself a graph-theoretic notion. By identifying positive component effect with the existence of an *RPR*, the former notion is also rendered non-graph-theoretic. This is important because of the obvious point that graphs are merely representational devices

22

(differing representations of the very same cases–employing more or fewer variables, for example–are possible), and it would be undesirable for an account of causation to relativize its diagnosis of a case to a choice of representation.[23]

Some further (non-graph-theoretic) restrictions must be placed on what variables the set S can be allowed to take as members if the probability-raising relation thereby revealed is to be potentially causal. First, S must include only variables representing events occurring no later than t_E.[24] To see why this is necessary, consider the following example:[25]

Example 4: Flood

Suppose a particular water main has a 0.02 chance of bursting during a certain interval of time. A nearby levee has an independent 0.01 chance of bursting during that same interval. If either bursts, the local neighbourhood will be flooded. If neither bursts, it will not. In fact the water main bursts (*Main* = 1), the levee holds (*Levee* = 0) and the neighbourhood floods (*Flood* = 1). Sometime later the levee engineer wins a professional accolade for her work (*Award* = 1).

Given the chances specified for *Main* = 1 and *Levee* = 1, together with the stipulation of their independence (and the stipulation that *Flood* = 1 just in case *Main* = 1 or *Levee* = 1), inequality (11) can be derived by straightforward application of the probability calculus:

$$P(Levee = 0|Main = 1. Flood = 1) = 0.99 > 0 = P(Levee = 0|Main = 0. Flood = 1) \qquad (11)$$

Holding fixed the occurrence of the flood, the chance that the levee holds conditional upon the water main's bursting is higher than the chance that the levee holds conditional upon the water main's holding (the fact that the putative effect event is here represented by the taking of value 0 by a variable makes no difference to the analysis). So it seems that the singleton containing only *Flood* acts as an *RPR* for *Main* = 1 and *Levee* = 0. Nevertheless, it is clearly *not* the case that the burst water main was a *cause* of the levee's holding. The example can be represented graphically as follows:[26]
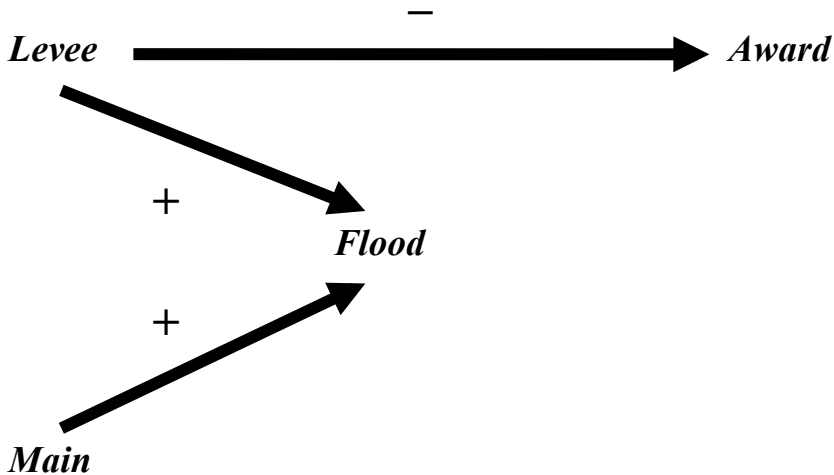
Levee ──────────────── **Award**

−

+

Flood

+

Main   **Figure 4**

Note that, by the time that *Flood* comes to pass, the relevant bursting of the levee has either happened or it hasn't (likewise with the bursting of the water main). So by disallowing probabilistic contributions revealed by conditioning upon events occurring later than the putative effect (in this case *Levee* = 0) from counting as causal, it is ensured that the analysis will not deliver the incorrect result that *Main* = 1 was a cause of *Levee* = 0.

In general, two alternative potential causes of a common effect will be probabilistically dependent conditional upon that effect. And we can exclude such probabilistic contributions

from counting as causal by disallowing those contributions revealed by conditioning upon (variables representing) events occurring later than the putative effect (thus once again invoking the assumption, relied upon in §2, that the causal and temporal orders coincide).

But don't further effects of these independent causes pose a problem? The holding of the levee is perhaps a cause of the levee engineer's winning the accolade (*Award*). But, holding fixed the flooding of the neighbourhood (which occurred prior to *Award*), *Main* raises the probability of *Award* (assuming that *Award* = 1 just in case *Levee* = 0, the probability that *Award* = 1 conditional upon *Flood* = 1 and *Main* = 1 is 0.99, whilst conditional upon *Flood* = 1 and *Main* = 0 it is 0). But surely the burst water main isn't a cause of the engineer's success![27]

The singleton {*Flood*} is indeed an *RPR* for *Main* and *Award*, but *Main* isn't a cause of *Award*. This is a case of positive component effect without causation, demonstrating that although positive component effect (as opposed to the stronger requirement of positive net effect, or straightforward probability-raising) may be a *necessary* condition for causation, it is not *sufficient*. It shall be seen in §§5-6 below that examples of probability-raising non-causation also involve positive component effect. They turn out to be cases in which the positive component effect is neutralised. And so it is in the present case.

To anticipate the discussion of §6: *Levee*'s taking value 0 neutralizes the positive component effect of *Main* on *Award* because (i) it *screens off Main* from *Award* in a stable manner (this notion of 'stability' will be explicated in §6) and (ii) there does not exist an *RPR* for *Main* and *Levee* = 0 (as has been seen, the singleton {*Flood*} is not, because *Flood* occurs after the relevant holding of the levee and so is excluded by our temporal restriction from counting as an *RPR* for the pair). Since the existence of an *RPR* is a necessary condition for causation, *Main* is therefore not a cause of *Levee* = 0 and so *Levee* = 0 isn't part of a causal chain from *Main* to

Award, despite (stably) screening off *Main* from *Award*. *Levee*'s taking value 0 therefore neutralizes the positive component effect of *Main* on *Award* and so *Main* is not a cause of *Award*.

A second sort of restriction that must be placed on a set **S** of variables that is a putative *RPR* for *C* and *E* is one that restricts its members to those representing reasonably natural events. Consider, for instance, the unnatural *disjunctive* event consisting of *C*'s *non-occurrence or E's occurrence*. A binary variable *V* representing this disjunctive event is one that takes value 1 if *C* fails to occur *or* if *E* occurs, and 0 otherwise. Evidently, for *any* choice of *C* and *E*, conditioning upon (the actual value taken by the sole element of) a singleton set containing only *V* would make *C* positively relevant to *E*. Likewise, where *C* raises the probability of *D* and *F* raises the probability of *E* (and *F* occurs), then conditioning upon (the actual values taken by members of) a singleton set containing a variable *V′* that takes value 1 either if *D* fails to occur *or* if *F* occurs, and 0 otherwise, can make *C* positively relevant to *E*. Evidently such unnatural variables aren't of the sort to reveal a causal relevance of *C* to *E* (cf. Yablo [2004], p. 122).

The notion of positive component effect has now been given a fully reductive analysis: *C* has a positive component effect upon *E* iff there exists a set **S** of variables representing the occurrence or non-occurrence (or occurrence-in-some-degree) no later than *E* of (fairly natural) events or states of affairs such that fixing each member of **S** at its actual value reveals a relation of positive relevance of *C* to *E*. In order to know what should be included in **S** we merely need to know facts about probability distributions and temporal relations, and not about causation.

The suggestion that positive token causation involves *positive component effect* is not in itself new.[28] It is one that has been developed by Hitchcock ([2001a]) in particular. Yablo

([2002], [2004]) and Dowe ([2004]) give related accounts of causation in terms of *de facto dependence* and *intrinsic probability-raising along a causal path*, respectively.

The central proposal has, however, been developed in a different way here than it is by Hitchcock, Yablo and Dowe. Whilst I have given a reductive probabilistic analysis of positive component effect in terms of the notion of an *RPR*, Hitchcock and Yablo each give counterfactual analyses that are designed specifically for determinism.[29] Dowe gives an account that, though probabilistic, appeals also to his notion of a causal process (which is not itself analysed probabilistically–see Dowe [2000]). Moreover, none of these accounts (nor even a probabilistic version of Hitchcock's counterfactual approach, given in his [2004b]), contain much by way of suggestion as to how to address the problem of probability-raising non-causation (a problem that has no analogue in the deterministic case).[30] Hitchcock in particular indicates that more work needs to be done to develop a probabilistic analysis that handles this problem ([2004a], esp. pp. 416-7, cf. [2001a], p. 275, [2001b], p. 372). This is precisely the challenge I seek to meet in what follows, and it is in this that the principal novelty of the account lies. The notion of positive component effect, and the corresponding probabilistic notion of an *RPR*, will play very important roles.

Of extant probabilistic analyses of causation, that developed by Kvart (in a series of articles including his [1991], [1994a], [1994b], [1997], [2004a], [2004b]) is perhaps the most similar in spirit to the one presented here. Kvart defines a notion of *ex post facto probability increase*, which is related to positive component effect. The idea is that in cases of non-probability-raising causation there may 'be an actual intermediate event *D* that yields probability increase when held fixed' ([2004a], p. 360; notation modified for consistency). In his ([1994b], pp. 206-7), Kvart also distinguishes the relations of *overall positive causal impact* from *some*

positive causal impact. *C* has *some positive causal impact* on *E* if it has positive causal impact that is propagated through certain *routes* or *threads* of causal impact. This echoes the notion of positive component effect. By contrast, *overall positive impact* is a function of the positive and negative impacts along all the various threads from *C* to *E*, echoing the notion of positive net effect. In his ([1997]) Kvart argues that causation is to be analysed in terms of *some positive causal impact*, where the latter is analysed in terms of the notion of *ex post facto probability increase*.

My quibbles with Kvart's diagnosis of non-probability-raising causation are relatively few and superficial. Whereas he thinks the relevant notion of *ex post facto probability increase* is one that is revealed by holding fixed *an actual intermediate event* between *C* and *E*, I insist (for reasons that should be fairly clear from the above presentation of my account) that to be sure of revealing potentially causal hidden probabilistic dependencies it is necessary to hold fixed the actual values of a *set* of variables, which represent the occurrence of actual events *or absences* occurring *no later than E*. I shall not rehearse my reasons for this since, as shall be seen below, my main objections to Kvart's account concern his treatment of probability-raising non-causes.

Before turning to that problem we must first consider the structure of the third of our examples of non-probability-raising causation, *Medicine*. Does this admit of the same treatment as the first two? That is, does Doctor's administering a low dose of the drug have a positive component effect upon Patient's recovery? The example would conform to the structure of *Thrombosis* if the following were an accurate representation (where *Low* = 1 or 0 according to whether or not Doctor administers a low dose, *High* = 1 or 0 according to whether Doctor administers a high dose, and *Recovery* = 1 or 0 according to whether Patient recovers):

Figure 5
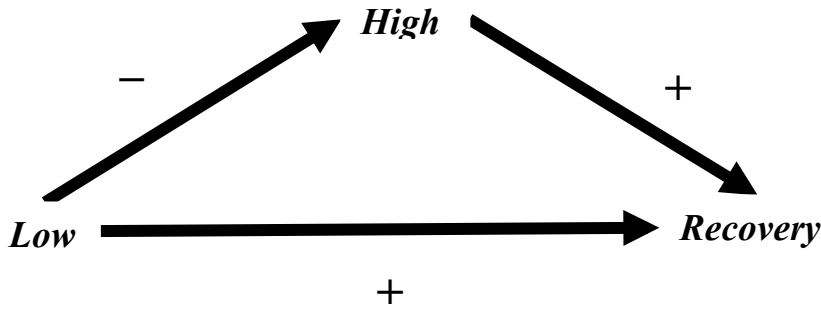
If *Medicine* has this structure, then clearly *Low* has a positive component effect on *Recovery* along the direct route and this can be isolated by holding fixed *High* at its actual value (namely 0). The singleton containing just *High* therefore acts as our *RPR*.

There is, however, a disanalogy between *Thrombosis* and *Medicine*. Whilst in *Thrombosis*, *Pills* and *Pregnancy* represented distinct events, it is not so clear that *Low* and *High* represent genuinely distinct events (the exclusion of the high dose by the low dose does not seem a mere *causal* exclusion).[31] So, if we wish to ensure that distinct variables represent potentially causally related events, we should use a single variable to represent both (cf. Hitchcock [2001a], p. 293, [2007a], pp. 502-3).

Let us then introduce a new variable, *Dose*, that takes values 2, 1, or 0 according, respectively, to whether Doctor administers a high, low, or zero dose. The structure of the example might then be represented as follows (where the '+' indicates a positive relation between the value of *Dose* and the value of *Recovery*):

Figure 6
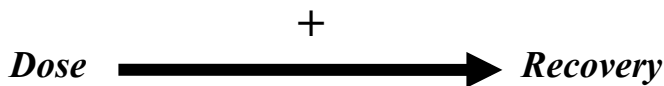
Dose = 1; *Recovery* = 1

This graph doesn't display two distinct routes, one transmitting a positive component effect, the other negative. Consequently, it does not make apparent what ought to be conditioned upon in order to reveal a relation of positive relevance between *Dose* = 1 and *Recovery* = 1.

There are two things to say about this case. First, the fact that *High* and *Low* seemingly do not represent distinct events in no way *precludes* us from modelling the example using the variables *High*, *Low*, and *Recovery*. We can model it like that so long as we are clear about what the resultant edges *do*, and what they *do not* represent. In particular, because of the failure of distinctness, the directed edge from *Low* to *High* cannot be taken to indicate a potential causal relation. If we do use these variables, then it is quite clear what must be held fixed to reveal a positive component effect of *Low* on *Recovery*.

The second thing to say is that, though the variable *High* is not employed in the Figure 6 representation of *Medicine*, nevertheless the singleton containing just *High* still acts as an *RPR* for *Dose* = 1 and *Recovery* = 1. *High* = 0 is equivalent to ~*Dose* = 2. Both express the proposition that *it is not the case that* Doctor administers the high dose. And *given* that ~*Dose* = 2, *Dose* = 1 raises the probability of *Recovery* = 1 (since the alternative is *Dose* = 0). Conditioning upon *High* = 0 (or equivalently ~*Dose* = 2) reveals a positive component effect of *Dose* = 1 on *Recovery* = 1, though Figure 6 (unlike Figure 5) doesn't represent a route along which that positive component effect is transmitted.

One might worry that *High* = 0 (or ~*Dose* = 2) corresponds to the unnaturally disjunctive state of affairs, *Dose* = 0 ∨ *Dose* = 1 (cf. Hitchcock [1993], pp. 340-2), and we earlier restricted the variables that could figure in *RPR*s to those representing reasonably natural states of affairs. But the doctor's failure to administer a high dose (*High* = 0) *is* a reasonably natural state of

affairs (certainly it is not as *un*natural as the state of affairs consisting in the doctor's failure to administer a low dose *or* the patient's recovery, the sort we wished to exclude from being represented by a variable in an *RPR*). The fact that it corresponds to the disjunct *Dose* = 0 ∨ *Dose* = 1 just reflects the commonplace that *absences* are typically multiply realisable by positive states of affairs. It does not follow from this that there are no reasonably natural absences or negative states of affairs.

One further worry about the present example is the following. By identifying positive token causation with positive component effect, we get the unequivocal result that *Dose* = 1 is a cause of *Recovery*. Yet, in the original presentation of *Medicine*, I said only that it is *perhaps* plausible that Doctor's administering a low dose was a cause of Patient's recovery. Intuition seems equivocal: it doesn't, for example, sound obviously false to say that in fact Patient recovered *despite* Doctor administering the low dose.

I don't think that the two locutions ('*C* is a cause of *E*' and '*E* despite *C*') are actually incompatible, though because of their contrasting explanatory import it sounds odd to assert them together. In Hesslow's example, we straightforwardly judge *Pills* to be a positive, token cause of *Thrombosis*, but it is nevertheless true that *Thrombosis* occurred *despite Pills*. The distinction between component and net effect is useful in part because it makes clear how both claims can be true together: *Pills* is a positive, token cause of *Thrombosis* because of the positive component effect of the former on the latter; but it is true that *Thrombosis* occurred *despite Pills* because of the negative net (and component) effect of the latter on the former (cf. Hitchcock [2001b], pp. 365-6).

The story is just the same in *Medicine*. *Dose* = 1 has a positive component but negative net (and component) effect on *Recovery*. It is consequently true both that the former is a cause of

the latter, and (though it sounds odd to say it at the same time) that the latter occurred in spite of the former. There exist two salient alternatives to the low dose (the high dose and the zero dose). Focus upon one (the zero dose) leads one to notice the positive component effect and to favour the 'is a cause of' locution; focus on the other (the high dose) leads one to notice the negative component effect and to favour the 'despite' locution (cf. Hitchcock [1993], p. 347, [1996b], p. 271). The salience of the alternatives and symmetry of the example makes it particularly difficult to settle on one locution over the other. The case is different in *Thrombosis* (where we more readily settle on the 'is a cause of' locution) because it lacks an analogous symmetry of salient alternatives to *Pills*.

In *Medicine*, the salience of one alternative can be raised at the expense of the other, with the result that one locution is correspondingly favoured. One way of doing this is by use of contrastive stress: one would not readily assent to the claim that 'the Doctor's administering a *low* dose was a cause of patient's recovery' (cf. Hitchcock [1996a], esp. pp. 408-14, [1996b], pp. 275-7, [2003], pp. 15-7). Another might be the mere act of asserting 'Patient recovered *despite* the low dose' as opposed to 'the low dose *was a cause of* recovery'. In the presence of cooperative conversational partners, this speech-act may in itself be sufficient to bring about a conversational context in which the high-dose alternative is salient, so that the claim becomes appropriate (cf. Lewis [1979], esp. pp. 346-7).

Further factors may complicate the picture still more: approbation or disapprobation for the doctor's actions may respectively incline us to use 'despite' or 'was a cause of'. What is important, though, is that none of these factors enter into the *metaphysical* story about causation, they merely govern which of various (and, strictly speaking, compatible) causal locutions (each loaded with explanatory, moral, and other connotations) we favour in a given context. This

observation is nothing new, but has been made by (among others) Lewis ([1986d], p. 162), Hall ([2000], p. 208) and Hitchcock ([2001b], esp. p. 384, [2003]).

Pragmatic issues shall be discussed again in §8.3. But, having argued in this section that positive component effect of *C* on *E* is necessary for (positive, token) causation, I now wish to examine the objection that it is not sufficient. This is because it might be that each positive component effect of *C* on *E* is *neutralized*. Examples of probability-raising non-causation (where there is not only positive component effect but positive net effect) illustrate, but are really only a special case. The point shall be seen to generalise: irrespective of whether *C* has a positive net effect on *E* (and so is an overall probability-raiser of *E*), *C* will not be a cause of *E* if each positive component effect is neutralized. First let us focus upon the special case.

5 Probability-Raising Non-Causation

The following three examples of probability-raising non-causation seem to be representative of the structures of those to be found in the literature. The first is my own, the second is due to Hitchcock ([2004a], p. 411) and the third to Schaffer ([2000a], p. 41).[32]

Example 5: Cricket

Tom and I are playing cricket. Tom hits the ball in the direction of the window
(*Hit*). I catch the ball, thus preventing the ball from impacting upon the window
(*Impact*). Coincidentally, a stone thrown by James strikes the window a moment
later and the window breaks (*Break*).

In the example, *Hit* raised the probability of *Break* (at least if we assume that there was some chance that I would fail to catch the ball), and *Break* did indeed occur, but clearly *Hit* was not a cause of *Break*. Intuitively, the reason for this is that the causal chain from *Hit* to *Break* was cut by my catching the ball.

One might doubt whether the breaking of the window that has its probability raised by *Hit* is identical with the breaking of the window that actually occurs (due to the striking of the window with a stone). But note that it can be hypothesized that the former would have been exactly the same as the latter in time and manner, so as to make such doubt not only difficult to maintain, but unwarranted even on extreme standards of event fragility.[33]

Analogous doubts seem more warranted in the case of a second example of probability-raising non-causation, one that is not characterized by the existence of a cut causal chain.

Example 6: Cancer

> 'Barney smokes, and he also spends a lot of time in the sun. These two proclivities are not connected; for example, Barney is not forced to go outside in order to smoke. Barney's smoking increases the probability that he will get lung cancer. By increasing his probability of getting lung cancer, Barney's smoking increases the overall probability that he will suffer from some form of cancer, and analogously for his exposure to the sun. In fact, Barney develops skin cancer. A fortiori, Barney develops cancer of some form or other.' (Hitchcock [2004a], p. 411)

Smoking raised the probability of Barney's developing cancer, but did not cause him to develop cancer (since the cancer he got was not of the sort that has its probability raised by smoking).

This example will be structurally different from *Cricket* provided we stipulate that the causal chain from Barney's smoking to his developing (lung) cancer is not cut at any stage (apart from at the very last stage, by his failure to get lung cancer): we can stipulate, for instance, that all the appropriate carcinogens made contact in the right way with the relevant lung tissue cells, yet (by chance) none of the cells became cancerous.

There is another sort of case that does not (at least not essentially) involve cut causal chains. The problem here, however, is not that the effect itself is of the wrong sort to be caused by the probability-raising non-cause, but that it has the wrong sort of accompaniments. Examples of this sort have been identified by Schaffer ([2000a]), who called them cases of 'overlapping'.

Example 7: Decay

'An atom of U-238 and an atom of Ra-226 are placed in a box at t0 (assume for simplicity that the box is otherwise empty). At t1 the box contains an atom of Th-234, an alpha particle, and (still) an atom of Ra-226. The relevant physical laws are: (1) an atom of U-238 has a certain chance per unit interval of producing Th-234 and an alpha particle, (2) an atom of Ra-226 has a certain chance per unit interval of producing Rn-222 and an alpha particle, and (3) these chances are independent. Now the presence of Ra-226 is *not a cause* of there being an alpha particle

(rather the U-238 produced the alpha particle independently), but is by

law a probability-raiser of it' (Schaffer [2000a], p. 41)

The presence, at t_1, of the Th-234 atom and the (continued) presence of the Ra-226 atom

(together with the relevant laws) serve to convince us that, despite being a probability-raiser, the

presence of the Ra-226 atom at t_0 was not a cause of the presence of the alpha particle at t_1. This

is not a case of chain-cutting, since such particle emissions do not involve intermediate processes

(cf. Schaffer, ibid.). Nor is it the nature of the effect itself (the presence of the alpha particle) that

convinces us that the Ra-226 atom did not cause it (since that effect would have been no

different if caused by the Ra-226 atom). Rather, it is the fact that the alpha particle is not

accompanied by a Rn-222 atom (but *is* accompanied by a Th-234 atom and an Ra-226 atom).

Attempts have been made in the literature to produce a sophisticated probabilistic

analysis that excludes probability-raising non-causes from counting as genuine causes. One

attempt–adopted by Good ([1961a], [1961b]) and Menzies ([1989], p. 656)–is to incorporate the

requirement that there be a *continuous* chain connecting cause to effect. Because cut causal

chains are (often, at least) spatio-temporally discontinuous, this requirement (often) succeeds in

excluding those cases of probability-raising non-causation that result from the presence of cut

causal chains. It will succeed in *Cricket*, for example, because there is no continuous chain

connecting *Hit* and *Break*.

But the requirement of a continuous chain has its costs. One is its definitional exclusion

of the possibility of action at a spatio-temporal distance.[34] A second is that it is far from clear

that cases of prevention and causation by absence and omission involve spatio-temporally

continuous chains (cf. Hall [2004], pp. 243, 249, Hitchcock [2004a], p. 411).[35] So whilst cases of

probability-raising non-causation often result from cut causal chains, such chains ought not to be characterised in terms of a lack of spatio-temporal continuity on pain of rendering the resulting analysis too restrictive. The analysis will also be too liberal, since the requirement of a continuous causal chain doesn't help us with those cases of probability-raising non-causation, such as *Cancer* and *Decay*, that don't involve cutting.

Other responses to the problem of probability-raising non-causation might be considered. The *method of more detailed specification of events*, a proposed solution to the problem of non-probability-raising causation, can also be deployed to help with the problem of probability-raising non-causation. In some cases a more precise specification of the non-cause might reveal it as a non-probability-raiser. More commonly though a more precise specification of the effect-event will reveal the non-cause to be a non-probability-raiser of it. So, whilst Barney's smoking raised the probability of his developing cancer, it did not raise the probability of his developing *skin* cancer.

But this solution does not work in all cases (cf. Schaffer [2001], pp. 81-2). There is no further specification of the t_1 presence of the alpha particle or of the t_0 presence of the Ra-226 atom that will reveal a failure of probability-raising between the two events (aside, perhaps, from extrinsic characterisations such as the-presence-at-t_1-of-an-alpha-particle-accompanied-by-a-Th-234-atom). And in *Cricket*, the success of this treatment involves suppositions about the details of the case (e.g. that *Hit* doesn't at all raise the probability of *Break* precisely specified) that can be stipulated away and don't seem part of our reason for making the causal judgements that we do (cf. Hitchcock [2004a], pp. 412-3).

Why *do* we judge that these cases of probability-raising don't involve causation? Hitchcock (ibid., p. 416) observes that in each case this is the result of the existence of a *marker*:[36]

'[We know that *Hit* was not a cause of *Break*] because the two events were not connected by an appropriate type of spatiotemporal process. [On the other hand] we know that Barney's smoking did not cause his cancer because Barney developed skin cancer, and that is not the sort of cancer that smoking causes. [Again,] we know that the presence of one atom but not the other caused the [presence of the alpha particle], because atoms cause [alpha particles to be present] by decaying, and atoms that decay are transformed from an atom of one type to another. In each case, there is a marker that distinguishes the genuine cause from the spurious probability-raiser.'

Yet Hitchcock despairs of the possibility of exploiting the existence of these markers to come up with a general solution to the problem of probability-raising non-causation:

'[A]s metaphysicians, we are interested in providing a general theory of causation. The markers described above are heterogeneous in nature [...]. For a theory of causation to exploit these markers, something more must be said about what they have in common in virtue of which they are causal markers. It will not do to simply say that the actual cause is the one that is marked as such.' (ibid.)

The account to be given below shall specify, in probabilistic terms, exactly what it is that the various sorts of marker have in common in virtue of which they are markers. To anticipate: each of these markers acts as *a positive component effect neutralizer*, a notion that shall be given a probabilistic definition. In order to assist us in seeing how this notion can be defined, it will be helpful to represent our examples of probability-raising non-causation graphically.

6 Graphical Representation of Cases of Probability-Raising Non-Causation

It was observed in the previous section that probability-raising non-causation often arises in cases of cut causal chains. In the *Cricket* example, *Hit* raised the probability of *Break* and *Break* in fact occurred, but the causal chain connecting the two events was cut by my catching the ball. The causal relevance of *Hit* to *Break* was neutralised by the failure of a complete causal chain to exist. The example can be represented using the following graph:
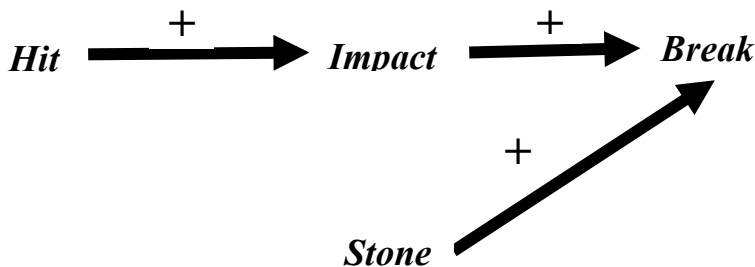
Figure 7

Hit has a positive component effect on *Break* along the path via *Impact*. Moreover, there is no negative component effect of *Hit* on *Break*, and so *Hit* straightforwardly raises the probability of (that is, has a *positive net effect* upon) *Break*.

But although *Hit* raises the probability of *Break*, it is not a cause of *Break* because *Impact* = 0. Since *Hit* was only positively relevant to *Break* because of its positive relevance to *Impact*, once *Impact* takes value 0, *Hit* is rendered irrelevant to *Break*. The component effect of *Hit* on *Break* is thus neutralized by *Impact's* taking value 0.

How can we characterise such a neutralising event in probabilistic terms? Well one thing *Impact* does is screen off *Hit* from *Break*. *Given that Impact* = 0, *Hit* was probabilistically irrelevant to *Break*:

$$P(Break = 1|Hit = 1. Impact = 0) = P(Break = 1|Hit = 0. Impact = 0) \qquad (12)$$

But this is no good as a general criterion for neutralising events. It is not only the *failures* of links on a causal chain (such as that represented by *Impact's* taking value 0), but also *links* themselves that screen off potential causes from their putative effects (*Impact's* taking value 1 would equally have screened off *Hit* from *Break*, yet the corresponding event–the ball's hitting the window– would have been a *link* on the causal chain.)

What, then, is the difference between a *link* and the *failure* of a link on a causal chain? It would be incorrect to say that a link must always be a positive event, and the failure an absence: in cases of causation by absence and omission, absences often constitute links, and positive events their failures (my failure to water the plant caused it to die by causing an absence of water in the soil). Rather, intuitively speaking, where *D* is a link on a complete chain running from *C* to *E*, *C* is a cause of *D* and *D* is a cause of *E* (irrespective of whether *D* represents an event or an absence). On the other hand, where *D* is the *failure* of a link, it is *either* the case that *C* is not a cause of *D or* that *D* is not a cause of *E*.[37]

For example, the ball's failure to hit the window is intuitively the failure of a link on the causal chain from Tom's hitting the ball to the window's breaking and Tom's hitting the ball was not a cause of the ball's failure to hit the window, nor indeed was the ball's failure to hit the window a cause of the window's breaking. By contrast, suppose that the ball had hit the window and the window had broken. Then Tom's hitting the ball would have been a cause of the ball's hitting the window, which in turn would have been a cause of the window's breaking.

This observation is not on its own much help in furnishing us with a non-causal criterion for a causal chain's having been cut. However, as has been observed, positive component effect (which was reductively analysed in terms of the existence of an *RPR*) is necessary for causation. And note that *Hit* does not have a positive component effect upon *Impact* = 0, nor does *Impact* = 0 have a positive component effect upon *Break*. On the other hand, had it been the case that *Impact* = 1, *Hit* would have had a positive component effect upon *Impact* = 1 and *Impact* = 1 a positive component effect upon *Break*. Since positive component effect is necessary for causation, and since links stand in causal relations to both endpoints of a chain but failures don't, perhaps this can serve as our probabilistic criterion for distinguishing links from their failures.

But, as has been indicated, the existence of positive component effect is not *sufficient* for causation, since there might be a neutralizing event. So perhaps an event or absence *D* could be the failure of a link on a chain from *C* to *E* in spite of *C*'s having positive component effect upon *D* and *D* on *E*, for it may be the case that *C* isn't a cause of *D* or *D* of *E* if a there is a neutralizing event or absence for either of *these* event pairs. I think that this worry is misplaced. Consider the following example.

Example 8: Cricket II

Tom hits the ball, I catch the ball, the window breaks (because James hits it

with a stone) and the burglar alarm goes off (because the cat set it off a moment

before the window broke).

In this example, *Hit* has a positive component effect upon *Break*, and *Break* has a positive

component effect on *Alarm* (in fact, the net effect is positive in each case). But it is neither the

case that *Hit* was a cause of *Break* nor that *Break* was a cause of *Alarm*, since both positive

component effects are neutralized (the first by the ball's failure to hit the window, and the second

by the failure of the shock waves from the window to reach the security system's detector prior

to the activation of the alarm.)

So the window's breaking fails to be a link on a complete causal process running from

Hit to *Alarm*. Nevertheless, the window's breaking is not itself *the failure of a link* on that causal

process. After all, a complete process would have included the window's breaking. The obvious

failures of links are events upon which *Hit* does not have a positive component effect, or which

do not have a positive component effect upon *Alarm*, e.g., the ball's failure to hit the window and

the failure of the shock waves to reach the detector prior to the alarm's sounding. So I maintain

that it is both necessary and sufficient for *D*'s constituting the failure of a link on a causal

process from *C* to *E* that either *C* has no positive component effect upon *D* or *D* has no positive

component effect upon *E*.

So we now have the following proposal: a neutralizing event *D* on a path along which *C*

has a positive component effect upon *E* is such that (a) *D* screens off *C* from *E*, and (b) either *C*

has no positive component effect upon D or D has no positive component effect on E (i.e. there exists no RPR for at least one event pair).

This analysis won't quite do. Where there exist paths from C to E that don't run via D, conditioning upon D needn't screen off C from E. Conditional upon D, C might be positively or negatively relevant to E, depending on the nature and strength of the other paths. Consider the following case:
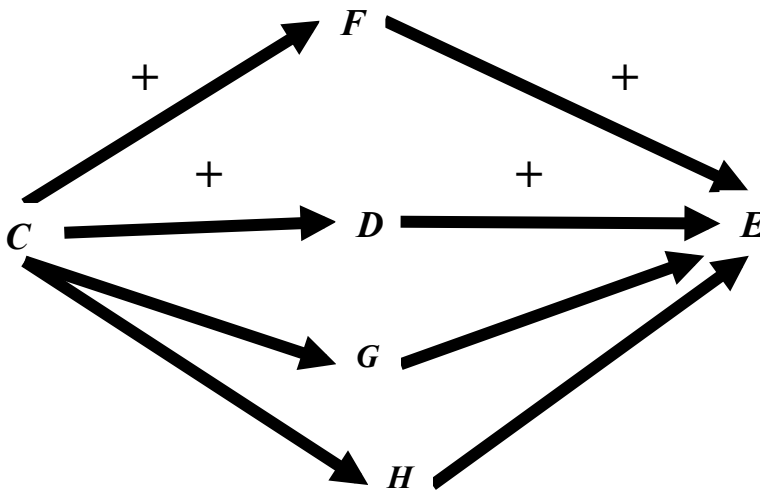
Figure 8

$C = 1; D = 0; F = 0; \ldots ; E = 1$

Suppose $D = 0$ and that this is such as to neutralise any causal relevance of C to E along the route via D. Nevertheless, because of the existence of the paths via F, G, and H, holding fixed $D = 0$ needn't render C probabilistically irrelevant to E.

For concreteness, suppose that in the structure represented by the above graph, positive component effect is transmitted along the routes via D and F and that negative component effect or no component effect is transmitted along those via G and H (if the latter, the route in question is *inactive*). Suppose, moreover, that the positive component effect transmitted along the two positive routes is neutralized by each of D and F taking value 0. Then holding fixed *both* D and

F at their actual values will render *C* either negatively relevant or irrelevant to *E* (depending on whether the routes via *G* and *H* transmit negative component effect or *no* component effect). That is, suppose that $A = \{D, F\}$, and let $\overline{A}$ be the proposition that each of the variables in A takes the value 0. Then:

$$P(E = 1|C = 1.\overline{A}) \leq P(E = 1|C = 0.\overline{A}) \tag{13}$$

Indeed, since the routes via *D* and *F* are the only positive routes, there are no further variables that we can condition upon to reveal a positive relevance relation between *C* and *E*. (If the route via *G* had been a positive route, conditioning upon the actual value of *H* in addition to $\overline{A}$, would reveal a positive relevance relation).

The general proposal then is this: where T contains variables whose actual values represent neutralising events for *each* of the positive routes from *C* to *E*, then conditioning upon the actual values of all variables in T will lead to a stable elimination of positive relevance of *C* to *E*. That is, if T is conditioned upon, there will be no further set U of variables (containing only variables corresponding to reasonably natural events occurring no later than t_E) such that conditioning upon their actual values reveals a positive relevance relation. That is, there is no set U such that:

$$P(E = 1|C = 1.\overline{T}.\overline{U}) > P(E = 1|C = 0.\overline{T}.\overline{U}) \tag{14}$$

Because of this, we might call T a *Stable Positive Relevance Eliminator Set* (*SPRES*) for *C* and *E*.

The corrected general probabilistic criterion for its being the case that all positive routes from C to E are neutralized can now be stated. All positive routes are neutralized if and only if (a) there exists an *SPRES*, T, for C and E (containing only variables that represent reasonably natural events and that occur no later than t_E) such that (b) *for each variable D_i in T, either C* does not have a positive component effect upon D_i or D_i does not have a positive component effect upon E. Where this is so, we might call T a *failure set* for C and E. Where there exists a *failure set* for C and E, there is no unneutralized positive component effect of C upon E, and so C fails to be a cause of E.[38]

We can now easily deal with cases of probability-raising non-causation that involve cut causal chains: in *Cricket*, there is only one positive route between *Hit* and *Break*, namely that which runs via *Impact*. But *Impact* = 0, and so this route is neutralized. So take the singleton I containing just the variable *Impact*. Let $\bar{I}$ be the proposition that this binary variable takes its actual value (= 0). When $\bar{I}$ is conditioned upon (effectively conditioning upon the fact that the ball fails to hit the window), *Hit* is no longer positively relevant to *Break* (presumably it is *irrelevant*):

$$P(Break = 1|Hit = 1.\bar{I}) \leq P(Break = 1|Hit = 0.\bar{I}) \tag{15}$$

Conditioning upon $\bar{I}$ eliminates the positive relevance of *Hit* to *Break* in a stable manner: once $\bar{I}$ is conditioned upon, there is nothing else that we can condition upon to recreate a positive probabilistic relation between *Hit* and *Break*. In particular, since there are no further routes from *Hit* to *Break*, we cannot condition upon further negative routes to reveal the existence of still further positive routes. So I is an *SPRES* for *Hit* and *Break,* and since it is neither the case that

Hit has a positive component effect upon *Impact* = 0, nor that *Impact* = 0 has a positive component effect upon *Break* (in each case there is only one negative route between the two), *I* counts as a *failure set* for *Hit* and *Break*, and we get the correct result that *Hit* was not a cause of *Break*.

But this was only *one* of our original examples of probability-raising non-causation. The others were chosen because they seemed to exhibit a different structure–one that didn't involve a cut causal chain. And yet hasn't the notion of a *failure set* been designed just to exclude cases involving cut causal chains?

In fact, the notion of a failure set is more generally applicable. There exists a failure set in each of our two other examples of probability-raising non-causation (which together with *Cricket* are exhaustive of the structures to be found in the literature). Consider *Cancer*. Barney's smoking (*Smoke*) raises the probability of his developing cancer (*Cancer*), but does not cause it. Take the set *L* containing the variable *Lung* which takes actual value 0 corresponding to Barney's failure to develop lung cancer (we will also need to include in *L* the variables *Throat*, *Mouth*, etc. representing the various other types of smoking-induced cancer). Conditioning upon the actual values of all the variables in *L* serves to stably screen off *Smoke* from *Cancer*. Moreover, since *Lung* = 0 does not have a positive component effect on *Cancer* = 1,[39] *L* serves as a failure set for *Smoke* and *Cancer*, and we get the correct result that *Smoke* does not cause *Cancer*, despite raising its probability.

Consider our final example, *Decay*. The presence of the Ra-226 atom in the box at t_0 (*Radium*) raised the probability of the presence of an alpha particle at t_1 (*Alpha*), but did not cause it (at t_1, there is no Rn-222 particle present, but there is still a Ra-226 particle present and there is a Th-234 particle present). Take the singleton *R* containing just the variable *Radon*,

which takes value 1 if there is an Rn-222 particle present at t_1 and 0 otherwise. Conditioning

upon the actual values of all the variables in **R** serves to stably screen off *Radium* from *Alpha*.

Moreover, *Radium* doesn't have a positive component effect upon *Radon* = 0, nor does *Radon* =

0 have a positive component effect upon *Alpha*. **R** therefore serves as a failure set for *Radium*

and *Alpha*, and we get the correct result that *Radium* doesn't cause *Alpha*, despite raising its

probability.[40]

So the probabilistically defined notion of a failure set serves to exclude all three examples

of probability-raising non-causation from counting as genuine cases of causation. And it does so

on the grounds on which we judge them to be cases of non-causation. That is, it does so in virtue

of the *markers* that allow us distinguish them from cases of genuine causation: the failure of the

ball to hit the window; Barney's not developing *lung* cancer (but rather *skin* cancer); the absence

of a Rn-222 particle at t_1 (together with the presence of a Ra-226 particle). The present account

says what these things have in common in virtue of which they are markers: they each have the

probabilistic properties that are characteristic of elements of a failure set.

Although *Cancer* and *Decay* are not cases of cut causal chains, they are nevertheless

cases of neutralization of positive component effect. The neutralization doesn't occur in virtue of

a cut in the causal chain, but rather in virtue of the nature of the effect itself (in *Cancer*) or what

accompanies or fails to accompany the effect (in *Decay*).

As has been observed, probability-raising non-causation is really just a special case of the

phenomenon of a potential cause *C* having at least one positive component effect upon *E*, but

where each positive component effect is neutralized (where there are cancelling negative

component effects, this positive component effect won't necessarily show up as overall

probability raising–or positive net effect). Not only are we now equipped to deal with

probability-raising non-causation, but the notion of a failure set allows us to deal with the more general phenomenon of which this is an instance.

7 Completing the Probabilistic Analysis of Causation

Putting the whole of the preceding together, we arrive at a probabilistic analysis of positive token causation. Positive token causation consists in *unneutralized positive component effect*. This has been given a reductive probabilistic analysis in terms of the existence of an *RPR* together with the non-existence of a *failure set*.

The analysis overcomes the well-known objections that have been brought to bear against the naïve probabilistic analysis of causation described in §2. The requirement of positive component effect rather than straightforward probability-raising, or positive *net effect*, is weak enough that the analysis is satisfied in cases of non-probability-raising causation. The requirement that it not be the case that each positive component effect is *neutralized* (in which case a failure set will exist) makes the analysis strong enough to prevent it from being fulfilled in cases of probability-raising non-causation.

Moreover, unlike extant sophisticated probabilistic analyses, the unneutralized positive component effect analysis overcomes the objections to the naïve probabilistic analysis in a manner that does not render it incompatible with causation by absence or omission, prevention, the possibility of action at a spatio-temporal distance, *direct* non-probability-raising causation, or failures of causal transitivity. This shall be demonstrated in the remainder of this paper.

Luke Glynn

## 8 Problem Cases for Extant Probabilistic Analyses

8.1 Causation by Omission

It may well be true in a particular instance that the farmer's omission to water his crop (*Water* = 0) was a cause of the crop's failure (*Crop* = 0). But probabilistic analyses, such as those of Good and Menzies, that require the existence of a continuous causal process connecting cause to effect have difficulty accommodating this fact. For whilst it might be maintained that there *is* an intermediate process connecting the farmer's omission and the failure of the crop–one that involves further absences, such as a deficiency of moisture in the soil, a lack of water being taken up by the roots of the crop, and an insufficiency of water available for metabolic processes–it is doubtful that this process is *continuous*. It is not at all clear, for instance, that the farmer's failure to water the crop has a spatial, or a precise temporal, location. But without such a location, it does not seem that it can be spatio-temporally contiguous with subsequent stages of the process.[41]

The unneutralized positive component effect analysis does not require that cause and effect be connected by a spatio-temporally continuous process. There is no implication that a lack of spatio-temporal continuity of a connecting process is itself sufficient to neutralize a positive component effect. Because of this, the analysis can readily accommodate cases of causation by omission.

The structure of the present case can be represented as follows:

Figure 9

$$Water \xrightarrow{\;+\;} Moist. \xrightarrow{\;+\;} Meta. \xrightarrow{\;+\;} Crop$$
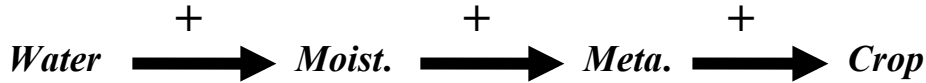
Water = 0; *Moist.* = 0; *Meta.* = 0; *Crop* = 0

The farmer's watering the crop (*Water*) raises the probability of there being enough moisture in the soil (*Moist.*) for the crop to metabolise sufficiently (*Meta.*) for it to survive (*Crop*). Consequently, *Water* has a positive component effect upon *Crop* (indeed, *Water* has a positive net effect on *Crop*). Correspondingly, *Water* = 0 has a positive component (and indeed net) effect upon *Crop* = 0. Since, in fact, *Moist.* = 0 and *Meta.* = 0 (and so on for any other relevant variables we care to interpolate), the positive component effect of *Water* = 0 on *Crop* = 0 is unneutralized,[42] and so we get the correct result that *Water* = 0 is a cause of *Crop* = 0 (irrespective of whether the intermediate causal process can be considered spatio-temporally contiguous).

I have considered a case of causation by omission. Causation by absence and prevention (or causation *of* absence) are analogous in terms of the difficulties they pose for extant probabilistic analyses and in terms of the treatment that they receive by the present account.[43] And, because it does not require the existence of a spatio-temporally continuous connecting process, the unneutralized positive component effect analysis also avoids definitional exclusion of the possibility of unmediated action-at-a-distance.

8.2 Direct Non-Probability-Raising Causation

The following is a slightly modified version of an example that Salmon gives of 'direct' non-probability-raising causation ([1980], p. 65, [1984], pp. 200-1):

Example 9: Direct Causation

Suppose that an unstable atom occupies a state which may be called the *fourth energy level*. There are several different ways by which it might decay to the *zeroeth* or ground level. Let $P(m \rightarrow n)$ be the probability that an atom in the mth level will make a direct transition to the nth level. And suppose the probabilities are as follows:

$P(4 \rightarrow 3) = 0.4 \quad P(3 \rightarrow 1) = 0.75 \quad P(2 \rightarrow 1) = 0.25$

$P(4 \rightarrow 2) = 0.4 \quad P(3 \rightarrow 0) = 0.25 \quad P(2 \rightarrow 0) = 0.75$

$P(4 \rightarrow 0) = 0.2$

The probability of the atom's occupation of the first energy level conditional upon its occupying the second is 0.25. The probability of its occupying the first conditional upon its non-occupation of the second is 0.5.[44] Its occupation of the second therefore lowers the probability of its occupation of the first. It might nevertheless seem plausible, if it occupies the fourth, then the second, then the first, that its occupation of the second is amongst the causes of its occupation of

the first. Salmon ([1980], p. 65) says that '[a]lthough this example is admittedly fictitious, one finds cases of this general sort in examining the term schemes of actual atoms.'

This example is one of *direct* non-probability-raising causation because there is apparently no intermediate causal process between the atom's occupation of the second energy level and its occupation of the first. As Salmon ([1980], p. 65) says 'we cannot, so to speak, "track" the atom in its transitions from one energy level to another' and, therefore, 'it appears that there is no way, even in principle, of filling in intermediate "links"' of a causal process. Consequently, analyses such as those of Good, Lewis and Menzies, that seek to deal with non-probability-raising causation by replacing the requirement of probability-raising with the requirement that cause and effect stand in the ancestral of that relation will not succeed in yielding the correct result in this case. Nor will Kvart's, since he requires the existence of an *actual intermediate event* that can be conditioned upon to reveal a hidden probability-raising relationship. And finally, it doesn't seem that Rosen's proposal will work here either, since there seems not to be any more detailed way in which we can specify the events involved so as to reveal hidden probability-raising.

The analysis advanced in this paper, by contrast, *does* yield the correct result. It requires neither probability-raising nor the ancestral of probability-raising. It instead requires the existence of unneutralized positive component effect. But occupation of the second level (*Second*) does have a positive component effect on occupation of the first (*First*). This is revealed by the fact that, if we condition upon the fact that the atom didn't occupy the third level (*Third* = 0), *Second* raised the probability of *First* (since, *given Third* = 0, the only alternative to

Second was the atom's decaying directly from the fourth to the ground level). The singleton containing only *Third* therefore acts as an *RPR* for *Second* and *First*.

The positive component effect revealed by this *RPR* isn't neutralized, for (unlike *Cricket*) there is no intermediate process to be cut in this case, nor (like *Cancer*) can *First* be more precisely specified so as to be revealed as the wrong sort of event to be caused by *Second*, nor (like *Decay*) are there any accompaniments to *First* which make us judge that *Second* is not its cause.

The structure of this case is rather like that of *Medicine*, and one could similarly debate how best to graphically represent it. Either we could represent it as *Medicine* is represented in Figure 5 (with *Low* swapped for *Second*, *High* for *Third* and *Recover* for *First*) or as in Figure 6 (with *Recover* again swapped for *First*, and *Dose* swapped for *Level*, which may take 3, 2, or 0 for values according respectively as the atom occupies levels 3 or 2 or decays directly to the ground state). But, just as in the case of *Medicine*, it does not ultimately matter which representation we go for, since either way the singleton containing only *Third* acts as an *RPR* for *Second* and *First* and we get the result that *Second* is a cause of *First*. Again, just as in *Medicine*, one might feel apparently conflicting temptations to say that *Second* was a cause of *First*, or that *First* occurred *in spite of Second*. But, as was observed in connection with that earlier example, there is no genuine inconsistency here, though the two locutions carry rather different explanatory connotations.

8.3 Failures of Transitivity

Examples of apparent transitivity failure have differing structures. The two considered below seem representative of those found in the literature. The first is due to McDermott, the second is given by Hitchcock who attributes it to Ned Hall:[45]

Example 10: Dog Bite

'My dog bites off my right forefinger. Next day I have occasion to detonate a bomb. I do it the only way I can, by pressing the button with my left forefinger; if the dog-bite had not occurred, I would have pressed the button with my right forefinger. The bomb duly explodes.'
(McDermott [1995], p. 531)

Intuitively, the dog's biting Michael's right forefinger (*Bite*) is a cause of his pressing the button with his left forefinger, and the latter is a cause of the explosion. Yet *Bite* is not a cause of *Explosion*.

Example 11: Boulder

'A boulder is dislodged, and begins rolling ominously toward Hiker. Before it reaches him, Hiker sees the boulder and ducks. The boulder

sails harmlessly over his head with nary a centimeter to spare. Hiker

survives his ordeal.' (Hitchcock [2001a], p. 276)

Intuitively, the falling of the boulder (*Fall*) is a cause of Hiker's ducking (*Duck*), and *Duck* is a

cause of *Survival*. Yet *Fall* is not a cause of *Survival*.

In the case of *Dog Bite*, the unneutralized positive component effect analysis, unlike the

probabilistic analyses of Good, Menzies and Lewis, yields the intuitively correct result. Let *Press*

be a variable that takes value 2, 1, or 0 depending on whether Michael presses the button with his

left hand, his right hand, or not at all. *Bite* has a positive component and net effect on *Press* = 2,

which in turn has a positive component and net effect on *Explosion*. The null set acts as an *RPR*

for both event pairs.[46] Since there is nothing to neutralize the positive contribution of *Bite* to

Press = 2, nor of *Press* = 2 to *Explosion*, there is no *failure set* for either event-pair, and we get

the correct result that *Bite* was a cause of *Press* = 2, and *Press* = 2 a cause of *Explosion*.

But the analysis *does not* yield the result that *Bite* was a cause of *Explosion*, since the

former does not have a positive component effect upon the latter. To see this, observe that the

structure of the case can be represented as follows:[47]

Figure 10

Bite ⟶ **Press** ⟶ **Explosion**

Bite = 1: *Press* = 2: *Explosion* = 1

(The edges in this graph don't admit of '+' or '−' annotations, since there is no simple positive or

negative correlation between values of the variables represented.) The only route from *Bite* to

Explosion runs via *Press* and then on to *Explosion*. It is this very path, and in particular its early

stages, along which *Bite* transmits its threat to *Explosion*. *Bite* consequently has a negative

component effect upon *Explosion* along this path. Nor is there a distinct path along which *Bite*

has a positive component effect. The unneutralized positive component effect analysis therefore

yields the intuitively correct result that *Bite* is *not* a cause of *Explosion* and that this is a case in

which causation fails to be transitive.[48] The route from *Bite* to *Explosion* is characterised by the

transmission of positive component effects along each of its stages, but no positive component

effect is transmitted along the entire route.

Boulder has a different structure to *Dog Bite* (cf. Hitchcock [2001a], pp. 290-1, 295-6,

[2007b], pp. 78-80). In *Dog Bite* there is a single route with positive component effect

transmitted along its stages, but not along the entire route. By contrast, in *Boulder* there are two

routes: on the one hand there is a route going via *Duck* (since the value of *Duck* depends

positively on the value of *Fall* and, given *Fall*, the value of *Survival* depends positively upon that

of *Duck*). On the other hand, there is a route bypassing *Duck* since there is a value of *Duck*

(namely 0) such that, holding *Duck* fixed at that value, the value of *Survival* depends (negatively)

upon the value of *Fall*. Hitchcock has pointed out that it is possible to isolate the component

effect transmitted along the route via *Duck* by interpolating a variable along the route bypassing

Duck (one that doesn't also lie on the route via *Duck*). He observes ([2001a], p. 296):

> 'There will be a point on the boulder's trajectory–let us say one meter from
>
> Hiker's head–such that by the time the boulder reaches that point, it is too late
>
> for Hiker to duck if he has not done so already.'

Let *Metre* be a variable representing the presence or absence of a boulder one metre from Hiker's head. The structure of the example might then be represented as follows:
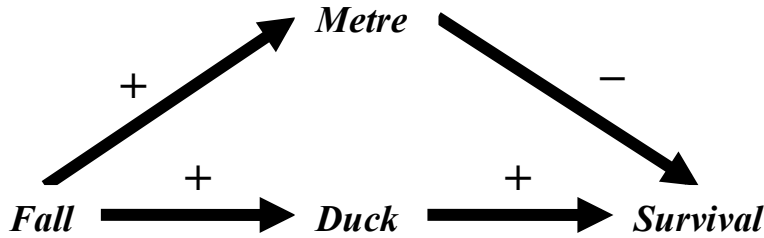

Figure 11

Fall = 1; *Metre* = 1; *Duck* = 1; *Survival* = 1

Holding fixed *Metre* = 1, *Fall* raises the probability of *Survival*. That is, *given* the presence of the boulder one metre from Hiker's head, *Fall* raises the probability of *Survival*, since it raises the probability that Hiker will see the boulder in time and duck (see Hitchcock [2001a], p. 297). So the singleton containing just *Metre* acts as an *RPR* for *Fall* and *Survival*.

Since the causal chain running from *Fall* to *Survival* via *Duck* is complete, there is nothing to neutralize the positive component effect. The unneutralized positive component effect analysis therefore yields the result that *Fall* is a cause of *Survival* and therefore that, unlike *Dog Bite*, *Boulder* is not a genuine case of transitivity failure.

This result is somewhat surprising, though I don't think it incorrect. The very same reasoning that helped us, two paragraphs ago, to identify the positive component effect of *Fall* on *Survival* also inclines us to regard the former as a cause of the latter (cf. Hitchcock [2001a], p. 297). There it was said that *given* the presence of the boulder one metre from Hiker's head, *Fall* raises the probability of *Survival*, since it raises the probability that Hiker will see the boulder in time and duck. I think it might equally be said that because of the presence of the boulder one metre from Hiker's head, *Fall* was a cause of *Survival* because it raised the probability of Hiker's

seeing the boulder in time and ducking. By mentioning the presence of the boulder one metre

from Hiker's head (a state of affairs that was not made salient in the original presentation of the

example), intuition is alerted to the existence of a positive component effect that was initially

difficult to detect and consequently judges *Fall* to be a cause of *Survival* after all.

But the initial difficulty of detecting a positive component effect is not the only reason

why, from the initial presentation of *Boulder*, we judge the case to be one of transitivity failure.

A second reason can be seen by comparing the structure of *Boulder* to that of *Thrombosis*. Both

Hall ([2007], pp. 121-3) and Hitchcock ([2007a], pp. 516-8) have observed that there is a close

analogy between these two structures.[49] In fact, they can be represented in *exactly* the same

manner if we make two alterations to our original graph of *Thrombosis*. These involve (a)

interpolating a variable on the direct route from *Pills* to *Thrombosis* that represents some

intermediate on the biochemical process by which *Pills* brings about *Thrombosis* that is not also

on the indirect route via *Pregnancy* (this variable will correspond to *Duck* in *Boulder*);[50] (b)

replacing the variable *Pregnancy* with the variable ¬*Pregnancy*, which takes value 1 if

pregnancy fails to occur and 0 otherwise (the sign associated with the edge from *Pills* to

¬*Pregnancy* will consequently be '+', and that from ¬*Pregnancy* to *Thrombosis* will now be '−

'):

Figure 12
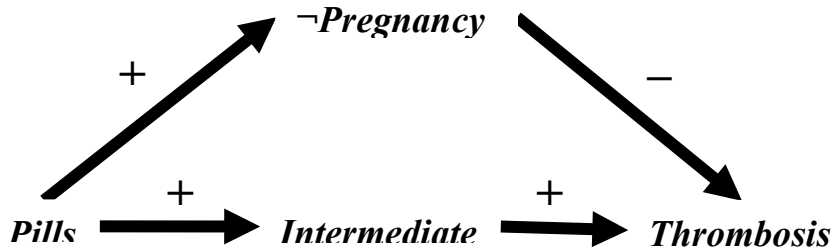
Pills = 1; ¬Pregnancy = 1; Intermediate = 1; Thrombosis = 1

Despite the close structural similarity, we readily judge *Pills* to be a cause of *Thrombosis*, but do not readily judge *Fall* to be a cause of *Survival*. What, then, is the disanalogy?[51] Why is it that, in the case of *Thrombosis*, intuition latches on to the positive component effect of *Pills* on *Thrombosis*, whilst in *Boulder*, it fails to latch on to the positive component effect of *Fall* on *Survival* (but seems rather to focus upon the negative component and net effect)?

A first important disanalogy is that, as already noted, the positive component effect is difficult to detect in *Boulder*. The positive component effect of *Pills* on *Thrombosis* is more readily discernible. *Pregnancy* is an obvious variable to control when looking for such a component effect. It is one that is referred to in the original presentation of the example, and one that we are told has an independent effect upon *Thrombosis*. That there is a route on which *Pregnancy* does not lie, and that transmits a positive component effect is also suggested in the original presentation of the example by talk of a 'complete biochemical process connecting *Pills* and *Thrombosis*'.

By contrast, in *Boulder*, the detection of separate positive and negative routes is hindered by the fact, first, that the presence of the boulder at a time too late for Hiker to duck is not a feature made salient in the initial presentation of the example and, second, that it seems rather

odd to hold fixed this presence whilst varying whether the boulder fell: how could there be a boulder there, if the boulder didn't fall?[52] Indeed, as Hitchcock points out ([2001a], p. 298), any story that would make salient the possibility of there being a boulder one metre from Hiker's head even without the boulder's falling would make the claim that *Fall* was a cause of *Survival* more plausible.

'Perhaps one could tell a story that would lead us to take this possibility seriously–perhaps Hiker has inadvertently walked in front of a boulder launcher that is carefully camouflaged against the hillside. But in just such a case, we should take the original causal claim seriously: by causing Hiker to duck in plenty of time, the fall of the boulder down the hillside does indeed save Hiker's life.'

This relates to the second reason that we more readily regard *Pills* as a cause of *Thrombosis* than we do *Fall* as a cause of *Survival*. This is that in *Boulder* (as in structurally similar examples of transitivity failure), the positive component effect of *Fall* on *Survival* comes about by way of *Fall*'s counteracting or, to use Hall's ([2000], p. 199; [2007], pp. 120-1) terminology, 'short-circuiting' a threat to *Survival* that *Fall* itself initiated.[53] The threat to *Survival* initiated by *Fall* is transmitted along the route via *Metre*, but it is counteracted by that via *Duck*.

Now it is also true that the positive component effect of *Pills* on *Thrombosis* results from its positive relevance to *Intermediate* which helps to counteract a threat (via ¬*Pregnancy*) to *Thrombosis* that *Pills* itself initiated. But the difference between the two cases is that the threat to

Thrombosis that *Pills* counteracts is not *entirely* initiated by *Pills*. There was a background threat to *Thrombosis*, since the probability of *Thrombosis* given *Pregnancy* is a good deal less than one. Introduce a similar background threat to *Survival*, and we are much more inclined to say that *Fall* is a cause of *Survival* (cf. Hall [2007], pp. 121-2). This is part of what is going on in the case described in the above passage from Hitchcock. To make the point even clearer, consider another variant on the example.

Example 12: Boulder II

This time suppose that, prior to the Boulder's fall, Hiker is suffering from a dangerously low supply of blood to his head. Hiker's ducking to avoid the boulder was in fact just what was needed to get his circulation back to normal.[54]

In this case there is a background threat to *Survival*–a threat posed by Hiker's poor circulation– that is not itself initiated by *Fall*. It is consequently much more natural to speak of *Fall* as a cause of *Survival*. I suspect, moreover, that the greater we make the background threat (the more dangerous Hiker's poor circulation is made to be), the more natural it will be to speak in this way (even if we don't make it so great as to yield an overall positive net effect of *Fall* on *Survival*).

What happens if we go in the other direction, and eliminate the background threat to *Thrombosis*? Consider the following example.

Example 13: Thrombosis II

Imagine that there is an animal species amongst which the females invariably

die from thrombosis in child birth. Scientists decide that the most ethical way to

test newly-developed birth control pills would be on members of this species.

The birth control pills are highly reliable at achieving their purpose, but

sometimes the creatures on which they are tested succumb to thrombosis.

Suppose that birth control pills are administered to one of these creatures and it

suffers thrombosis.

Is it plausible to say, in this case, that the consumption of birth control pills was a cause of the

creature's suffering thrombosis? It seems to me much less so–after all, the creature would

certainly have suffered thrombosis had the birth control pills not been administered. The example

now bears close resemblance to the 'switching' structures described by Hall ([2000], p. 205,

[2007], pp. 117-9), in which intuition yields the result that the switching event is *not* a cause of

the relevant effect.

So, in addition to the fact that positive component effect is difficult to identify in the

original presentation of *Boulder*, we are disinclined to describe *Fall* as a cause of *Survival*

because the positive component effect of the former on the latter arises exclusively from *Fall*'s

counteracting (or short-circuiting) a threat to *Survival* that *Fall* itself initiated.

The influence of this second factor is not at all difficult to explain. It just reflects the

explanatory connotations of the 'is a cause of' locution. These connotations sometimes render its

use misleading even if strictly correct.[55] Examples with the structure of *Boulder* are cases in

point: it is paradoxical to explain an event *E* in terms of an event *C* where there was no threat to

E other than that resulting from *C* itself. In such instances, saying that *C is a cause of E* will tend to mislead even though strictly true (the compatible '*E despite C*' locution is more natural). I therefore submit that examples with the structure of *Boulder*, unlike those with the structure of *Dog Bite*, are not genuine cases of transitivity failure.

9 Conclusion

In this paper, an analysis of positive token causation in terms of unneutralized positive component effect has been advanced. The latter notion has been given a fully reductive analysis in terms of the existence of an *RPR* and the non-existence of a *failure set*, both of which notions are defined in purely probabilistic terms. Unlike *naive* probabilistic analyses in terms of straightforward probability-raising, this analysis gives the correct diagnosis of cases of non-probability-raising causation and cases of probability-raising non-causation. Unlike extant *sophisticated* probabilistic analyses, it is able to correctly diagnose the full range of such cases to be found in the literature, and achieves these diagnoses in a manner that does not render it inconsistent with causation by absence and omission, prevention, direct non-probability-raising causation, failures of causal transitivity, or the possibility of action-at-a-distance. It therefore represents an improvement over these analyses.

I have here focused upon problems that specifically afflict extant *probabilistic* analyses of causation, showing how a probabilistic analysis can be developed that overcomes them. I have not attempted a demonstration of the ability of the resulting account to handle certain problem cases that pose no *special* problem for the probabilistic approach. The more exotic varieties of pre-emption, such as trumping, fall into this category.

Mundane varieties of pre-emption are readily handled by the present account. In fact *Thrombosis* is a case of probabilistic early pre-emption: Jane's consumption of the birth control pills is the pre-empting cause of her suffering thrombosis, her engaging in unprotected sex a pre-empted alternative (it initiates a process that if uninterrupted would with some probability have led to pregnancy and thrombosis). The pre-empting cause (Jane's consumption of the pills) is a non-probability-raiser of thrombosis. The pre-empted alternative (Jane's engaging in unprotected sex) a probability-raising non-cause.[56] Menzies has noted ([1989], pp. 645-7, [1996], pp. 88-9) that this is commonly the case in examples of probabilistic pre-emption.

It has already been seen that the present account correctly diagnoses Jane's pill-consumption as a cause: it counts as such because, despite being a non-probability-raiser of thrombosis, it has a positive component effect upon thrombosis. It should be equally clear that Jane's engaging in unprotected sex will be correctly diagnosed as a non-cause, despite being a probability-raiser. This is because its positive component effect upon thrombosis is neutralised by Jane's failure to become pregnant.

Late pre-emption cases are treated in just the same way. Take Lewis's ([2004], p. 82) example of Billy and Suzy throwing rocks at a bottle. Suzy throws slightly earlier or slightly harder, so that her rock arrives first and the bottle shatters. Billy's rock arrives on the scene a moment later. Even if Billy is so reliable a shot that Suzy's throw fails to raise the probability of the bottle's shattering, her throw nevertheless has a positive component effect revealed by holding fixed the failure of Billy's rock to hit the bottle. The present account consequently gives the correct result that Suzy's throw is a cause. Since the failure of Billy's rock to hit the bottle (at least when taken together with the fact that Suzy's rock did hit) neutralises any positive component effect of his throw on the bottle's shattering, the present account also correctly treats

Billy's throw as a non-cause, even though it may be a probability-raiser (if there is some chance that Suzy will miss and Billy will hit).

Now consider the more difficult case of trumping pre-emption. Schaffer ([2000b], p. 175) describes an example in which Sergeant and Major stand before Corporal and simultaneously shout 'Charge!' and Corporal charges. Major's order trumps Sergeant's, but Sergeant's order spoils the probabilistic dependence of Corporal's action upon Major's order. Nevertheless, perhaps a positive component effect can be recovered by conditioning upon the fact that Major issues a *ranking* order (this suggestion is inspired by the treatment of trumping pre-emption given by Yablo [2004], p. 134). *Given* that Major issues a ranking order, his shouting 'Charge!' raises the probability of Sergeant's charging (since the alternative is his issuing some other command).

I have some qualms about the fact appealed to in order to recover positive component effect in this case.[57] If these worries are well-founded, then perhaps the present account requires further refinement in order to provide a fully satisfactory treatment of trumping. But solving the problem of trumping pre-emption is not the burden of this paper, since trumping is certainly not a problem that *specifically* afflicts probabilistic analyses of causation. In any case, the present account is compatible the best extant proposals for dealing with trumping,[58] and we might reasonably hope that any still better solution might equally be adapted. For now, though, I will be content to have at least demonstrated that we have significant grounds for optimism about the prospects for a successful probabilistic analysis, which is desirable for reasons described at the outset.

Fachbereich Philosophie

Universität Konstanz

78457 Konstanz

Germany

luke.glynn@uni-konstanz.de

---

[1] For simplicity I make the standard assumption that *events* are the relata of the causal relation. I believe, however, that most of the central points that follow are compatible with opposing views on this matter (see e.g. Mellor [1995] and Paul [2000]).

---

[2] The approach can be extended to cover causes and effects that are more naturally represented by multi-valued variables. The use of ternary variables to represent cause-events shall be illustrated at the end of §4 and in §8.2.

[3] Each of the authors mentioned at the end of §1 implements this strategy in some form or other.

[4] As suggested by Cartwright ([1979], esp. pp. 420-3), Skyrms ([1980], pp. 103-9), and Eells ([1991], p. 330).

[5] For a defence of this claim, see Papineau ([1991], pp. 406-8).

[6] This is the approach of Reichenbach ([1971], p. 204), Suppes ([1970], p. 23), and Kvart ([2004a], p. 359).

[7] Holding fixed historical background in evaluating the probabilistic relationship between C and E also ensures that the probabilistic dependence of E upon C is not of the sort–identified by Sober ([1987], [2001])–that may arise between two causally independent quantities both of which increase monotonically (or at least with high but independent probabilities) over time. Take Sober's own example of a positive correlation in a time series between Venetian Sea levels and British bread prices. As he observes, the values taken by these quantitative variables at any given time 'are independent of each other once one conditionalizes on the separate causes affecting each' ([2001], p. 340). Holding historical background fixed, as well as serving to hold fixed any *common* causes of C and E, also serves to hold fixed any *independent* causes, thus rendering them probabilistically independent (in the absence of any causal relation between them). Exactly the same is true with regard to Sober's example of causally independent quantities that are correlated not just in their *levels* but also in their *changes* (ibid., pp. 335-9).

[8] Mellor ([1995]) also makes use of counterfactual conditionals in attempting to explicate a suitable notion of probability-raising. However, the consequents of the counterfactuals that he

appeals to do not concern the unconditional probability of the putative effect, but rather the probability that it *gets* from the putative cause or its absence (see Edgington [1997], p. 415). Since causes don't *get* probabilities from their effects, nor do independent effects of a common cause *get* probabilities from one another, Mellor takes his account to be in no danger of generating spurious cases of backwards causation or causation between independent effects of a common cause (see Mellor op. cit., esp. pp. 62, 224-9). Unlike Lewis, Mellor therefore does not rely upon the non-backtracking nature of the relevant counterfactuals in order to avert such a danger. On the other hand, as Edgington (ibid., pp. 415-6) observes, by cashing out probability-raising in terms of the apparently causal notion of one thing's *getting* probability from another, Mellor seems to introduce a circularity into his probabilistic account of causation. For this reason I focus instead upon the Lewis-Menzies counterfactual approach in the main text.

[9] A phrase used by Collins, Hall and Paul ([2004], p. 6).

[10] In fact Elga ([2001]) has argued that, at least in deterministic worlds, Lewis's similarity metric doesn't even succeed in excluding backtrackers. This is especially ironic since it was particularly with respect to deterministic worlds that the counterfactual approach was supposed to enjoy an advantage over the conditional probability approach (for reasons outlined three paragraphs ago in the main text).

[11] Or (what we might call) 'quasi-causal' background if we want to maintain formal neutrality over whether or not causal order is to be analysed in terms of temporal order ('quasi-causal order' being a place-holder for whatever figures in the analysans of one's preferred analysis of causal order). I shall make no attempt to retain this formal neutrality in what follows.

[12] Examples given by Good ([1961a], p. 318), Eells ([1991], pp. 281-2) and Hitchcock ([2001b], pp. 366-9) can be assimilated to the first case, whilst one given by Hitchcock ([1996a], pp. 401-

3) and a variant of Rosen's golfer example discussed in Hitchcock ([2004a], pp. 404-5; see also Salmon [1984], pp. 199-200) have the same structure as the third. The example to be discussed separately in §8.2 (due to Salmon [1980], p. 65, [1984], pp. 200-1) is also a variant on this third case. The second example illustrates a structure that is just an obvious variant on the first sort of case.

[13] Here and in what follows I suppress the proposition $\overline{B}$ for notational clarity. Strictly speaking, this should appear in the conditions of each of the conditional probabilities given in the remainder of this paper.

[14] Although given that birth control pills work by mimicking the hormonal effects of pregnancy, the empirical supposition of such an *Intermediate* is at least somewhat dubious. I thank an anonymous referee of this journal for pointing this out.

[15] Including those given by McDermott ([1995], pp. 531-3), Hall ([2000], pp. 200-1, [2004], pp. 246-8), and Hitchcock ([2001a], pp. 276-7).

[16] The *method of more detailed specification of events* is here considered as it applies to *causes*. But in some cases a more detailed specification of the *effect* event might reveal a hidden probability-raising relation. This latter strategy–akin to one considered and rejected by Lewis ([1986e], pp. 204-5; cf. Menzies [1989], pp. 649-50)–is open to exactly the same objections and to more besides (for an additional objection see Lewis, ibid. pp. 198-9).

[17] Hitchcock ([2004a], pp. 412-3) makes similar points about an analogous proposal for dealing with the problem of probability-raising non-causation, to be discussed in §5 below.

[18] There are others, including neuron diagrams. For arguments that graphical representation (at least when accompanied by detailed information about the associated probability distribution or pattern of counterfactual dependence) is superior, see Hitchcock ([2007b]).

[19] Here I follow the presentation of graph theory given in Spirtes, Glymour and Scheines ([2000]). In particular, I use graphs to represent features of probability distributions rather than patterns of counterfactual dependence or corresponding structural equations as, for example, do Pearl ([2000]) and Hitchcock ([2001a]).

[20] An alteration of an event (as defined by Lewis [2004], p. 88) is a very fragile version of the event in question or a very fragile alternative to it.

[21] In what follows, I deploy the terminology of *component effect* somewhat differently to Hitchcock. The main difference is that I shall end up giving it a more-or-less stipulative definition in probabilistic rather than graph-theoretic terms. Consequently the notion of C's having a component effect on E is not here relativized to a graphical representation of a probability distribution nor need we always speak of a component effect as being propagated along some or other 'route'.

[22] Consequently there may be difficulty in isolating this negative component effect for the reason outlined in footnote 14 above.

[23] Though see Hitchcock ([2001a]) for an endorsement of such relativity.

[24] More precisely: each variable $S \in \mathbf{S}$ must be such that its value depends just upon whether or not some event or state of affairs v occurs or obtains at a time no later than t_E.

[25] I thank an anonymous referee of this journal for drawing my attention to the need to address such examples.

[26] In the graph, *Flood* is what Spirtes, Glymour and Scheines ([2000], p. 10) call an 'unshielded collider'.

[27] It was a stipulation of the example that whether or not the levee bursts is independent of whether or not the water main burst. We can stipulate in particular that the flooding of the

neighbourhood that resulted from the burst water main didn't subject the levee to any additional strain (and so didn't enhance the impressiveness of the levee's holding to the award-committee). This ensures the accuracy of Figure 4 (which lacks a directed path from *Flood* to *Award*) as a representation of the structure of the example.

[28] In saying that positive token causation 'involves' positive component effect, I mean only that positive component effect (as opposed to positive *net* effect, or straightforward probability-raising) is *necessary* for positive token causation. As already noted, I do not mean to claim that it is *sufficient*. One example of positive component effect without causation has already been given, and several more will be discussed in §5. In §6 I will seek to show how the requirement of positive component effect may be supplemented in order to arrive at a full-blown probabilistic analysis of causation.

In contrast to the approach taken here, Hitchcock ([2001a]) analyses positive component effect in counterfactual (rather than probabilistic) terms and takes the notion, so analysed, to be both necessary *and* sufficient for deterministic causation. The same cannot be maintained in the probabilistic context. Roughly speaking, this reflects the fact that counterfactual dependence between distinct events can plausibly be taken as sufficient for causation, whilst probabilistic dependence cannot.

[29] Hitchcock ([2001b], pp. 363, 374, [2004b]) is clearly aware that the notion *can* be analysed in probabilistic terms, though he does not consider the possibility that such an analysis might be *reductive*. Instead, he contrasts non-reductive probabilistic analyses with potentially reductive counterfactual analyses ([2001b], pp. 371, 377-378, 389-390, 393-5). However, because he worries about the possibility of giving a non-causal semantics for non-backtracking counterfactuals, he is also sceptical about the possibility of a reductive counterfactual analysis

([2001b], pp. 378, 393, [2004b], p. 139). Another difference between my account and Hitchcock's is that (as already observed in footnote 23) Hitchcock relativizes the notion of positive component effect (and consequently token causation) to a mode of representation.

[30] The helpfulness of Dowe's requirement of a connecting process in dealing with this problem is mitigated by the resulting difficulties his account has in handling causation by absence and omission, prevention, and the possibility of action-at-a-distance. In any case (as shall be seen in §5) not all cases of probability-raising non-causation involve an incomplete connecting process.

[31] Consequently the directed edge from *Low* to *High* cannot be taken as indicative of a potentially causal probabilistic dependence between the events represented by these variables. Indeed the rule, given at the beginning of this section, for including directed edges in graphs justifies the inclusion of an edge from *High* to *Low* just as much as it justifies the inclusion of that from *Low* to *High*. This would yield a *bi-directed edge* (cf. Spirtes, Glymour and Scheines [2000], p. 6). By contrast, the rule does not justify the inclusion of an edge from *Pregnancy* to *Pills* in the graphical representation of Hesslow's example.

[32] Examples given by Menzies ([1989], pp. 645-7), Edgington ([1997], p. 419), Lewis ([2004], pp. 79-80) and Hitchcock ([2004a], p. 410) assimilate to the first sort of case, whilst an example given by Hitchcock (ibid., p. 415) and the various other examples of 'overlapping' given by Schaffer ([2000a]), have the same structure as the third case. Lewis ([1986e], pp. 193-4) gives an example with the same structure as the second.

[33] For discussion of this point, see Lewis ([1986e], pp. 204-5) and Menzies ([1989], pp. 649-50).

[34] This is one of the reasons for Menzies' later abandonment of his analysis ([1996], p. 94).

[35] Causation by omission is discussed further in §8.1 below.

[36] A similar point is made by Halpern and Pearl ([2005], pp. 862-3).

---

[37] It needn't be that both disjuncts are true. It might be that C is a cause of D, but D isn't a cause of E. Suppose Jane takes birth control pills, fails to suffer thrombosis and dies from some other (unrelated) cause. Jane's failure to suffer thrombosis is a failure of a link on the causal chain from her taking birth control pills to her death. Nevertheless (because of their prevention of pregnancy) the consumption of birth control pills can be considered a cause of her failure to suffer thrombosis (but her failure to suffer thrombosis is hardly a cause of her death.) Or it could be that D is a cause of E, but C isn't a cause of D. Suppose that Billy is about to throw a stone at a bottle. Suzy resolves to hit the bottle with a sledgehammer just in case Billy's stone does not hit the bottle. Billy throws and misses and Suzy hits the bottle, breaking it. The failure of Billy's stone to hit the bottle constitutes the failure of a link on the causal chain from Billy's throw to the bottle's breaking. Nevertheless the failure of the stone to hit the bottle can perhaps be considered a cause of its breaking (because it caused the hit with the much heavier sledgehammer). But Billy's throw didn't cause the failure of the stone to hit the bottle (the stone certainly wouldn't have hit the bottle had Billy not thrown it).

One might wonder whether C could be a cause of D and D of E and yet D nevertheless be a failure of a link on a causal chain. In particular, one might wonder whether cases of transitivity failure have such a structure. Not so. Cases of transitivity failure are puzzling because there is no causation despite the existence of a complete causal chain. The problem there is that the chain fails to transmit a positive component effect, or so I shall argue in §8.3 below.

[38] The notion of a *failure set* containing a neutralizing event for each positive component effect of C on E bears some similarities to Kvart's ([2004a], pp. 366-9) notion of a 'causal relevance neutralizer'. Both are intended to probabilistically characterize cut causal chains. However, Kvart defines a 'causal relevance neutralizer' as an intermediate *event* that acts as a *stable*

screener for *C* and *E*, and that *is not caused* by *C*. By contrast, I allow a) that neutralization may be due to *absences* as well as positive events; b) that (at least where positive causation is concerned) *stable weak decrease* and not *stable screening* is the central probabilistic characteristic of neutralization (for reasons given in the main text); c) that in multi-route cases, only a *set* of events and absences, and not a single event or absence, will act as a *stable weak decreaser*; d) that neutralization can be characterized without circular reference to the notion of *cause* (though Kvart argues that the circularity in his account can be avoided at the cost of introducing 'non-vicious' infinite regress).

[39] Perhaps *Tar* (a variable representing the presence of tar in Barney's lungs) acts as an *RPR* for *Smoke* and *Lung* = 0, so that the former has a positive component effect on the latter. *Given* that *Tar* = 1, perhaps *Smoke* raises the probability of *Lung* = 0 (since it raises the probability that *Throat* = 1 and so raises the probability that Barney won't survive long enough to develop lung cancer). (A structurally similar example is discussed in §8.3.) Nevertheless, there can be no RPR for *Lung* = 0 and *Cancer* = 1 because $P(Cancer = 1 | Lung = 1) = 1$.

[40] The singleton $\boldsymbol{Q}$ containing just *Radium1*, which takes value 1 just in case there is an Ra-226 atom in the box at t_1, also acts as a failures set (since it stably screens off *Radium* from *Alpha* and there is no positive component effect of *Radium1* on *Alpha*).

[41] Hall ([2004], pp. 243, 249) observes that even if absences and omissions do have precise spatio-temporal locations there is the additional problem that in certain cases of causation by absence or omission it is not clear that the sequence of absences or omissions initiated by the cause will intersect spatio-temporally with the effect.

[42] If, by contrast, it had rained heavily so that *Moist.* = 1 in spite of *Water* = 0, the positive component effect *would* have been neutralized.

[43] The account is liberal when it comes to admitting omissions and absences as causes and effects. For reasons best expressed by Lewis ([1986d], p. 162) and Hall ([2000], p. 208) I think it quite right that it should be so.

[44] If it doesn't occupy the second, then there is a probability ⅔ of its occupying the third and a probability ⅓ of its decaying directly to the ground level without occupying the first). The probability of its occupying the first given its occupation of the third is 0.75. Multiplying ⅔ by 0.75, one gets 0.5.

[45] Examples given by Yablo ([2002], pp. 134-5) and Hitchcock ([2003], p. 10, [2007a], p. 517), as well as one that Hall ([2000], p. 201) attributes to Hartry Field are assimilable to the second (in that each involves positive component effect). An example of Hall's ([2000], p. 201), and one that Hall (ibid.) attributes to Kvart assimilate to the first (since none of these involve positive component effect). Some slight variants on the second are mentioned in footnote 53 below.

[46] Lest it be doubted that *Press* = 2 has a positive net effect on *Explosion* recall that it is a stipulation of the example that, when Michael presses the button, his doing so with his left forefinger is the only way he can do it. But even if it were not, *Press* = 2 would still have had a positive component effect upon *Explosion*, revealed by conditioning upon the fact that he didn't press the button with any other part of his body, that he issued no order to press the button to an underling, etc.

[47] Hitchcock ([2001a], pp. 290-1) represents the example in the same manner. Indeed, he discusses the structure of the example in some detail (ibid., pp. 290-5, [2001b], pp. 387-8) and also concludes that it is a case in which positive component effect fails to be transitive.

[48] Paul ([2000]) has argued that the appearance of transitivity failure in cases like this disappears if we allow that the relata of the causal relation are property instances or *aspects* and not (or not

only) events. Without going into the details here, it seems that if one wanted to attempt to maintain causal transitivity by replacing events with property-instances as the (primary) causal relata, then one could do so consistently with preserving the essence of the analysis presented in this paper (note, though, that Hall ([2000], p. 205) has observed that *Dog Bite* can be straightforwardly modified to render it immune to Paul's treatment).

[49] I thank an anonymous referee for this journal for pressing this point.

[50] Let us assume, for the sake of argument, that there is such an intermediate (thus setting aside the worries expressed in footnote 14).

[51] Hall ([2007], pp. 121-32) and Hitchcock ([2007a], pp. 516-22) give diagnoses of the disanalogy that are different from that which is presented here. Their treatments both depend upon a distinction between 'default' and 'deviant' states of affairs (see Hitchcock [2009], for a critique of Hall's version.) These alternative treatments are broadly compatible with the analysis of causation that has been given in this paper except that, if the default/deviant distinction must be cashed out in causal terms (cf. Hall [2007], p. 125; Hitchcock [2007a], p. 506), the claim of reductivity would have to be abandoned.

[52] Cf. Hitchcock ([2001a], p. 297). Correspondingly, the probability of *Survival* = 1 conditional upon *Metre* = 1 and *Fall* = 0 will be undefined on the Kolmogorov axiomatization unless there is a non-zero probability (albeit astronomically small) of *Metre* = 1 despite *Fall* = 0.

[53] Similar diagnoses are given by Yablo ([2002], pp. 134-5, [2004], p. 124); Hall ([2000], p. 202; [2007], pp. 120-1); and Hitchcock ([2003], p. 11; [2007a], p. 521). This diagnosis also applies to the structurally slightly different examples of transitivity failure presented in Menzies ([2004], pp. 825-6) and Hitchcock ([2007a], pp. 519-20).

[54] Yablo ([2004], p. 128) gives a similar example.

[55] Compare: 'Your failure to water my cheese plant was a cause of its death'. This comes out true on any analysis of causation in terms of counterfactual or probabilistic dependence (and rightly so–see footnote 43). Yet, unless you were specially responsible for watering my cheese plant, the sentence is misleading since your omission does not play a central role in any adequate explanation of the cheese plant's death.

[56] In spite of Jane's having taken the birth control pills, her engaging in unprotected sex will still raise the probability of her suffering thrombosis provided that the pills aren't 100% reliable, that is, provided there is some chance of her becoming pregnant despite taking them.

[57] In particular, I'm worried that Major's issuing a ranking order may be an inappropriately unnatural fact (cf. Yablo, ibid.). I'm also worried that if it is stipulated that Major is capable only of shouting 'Charge!', the probability of Corporal's charging conditional upon Major's issuing a ranking order but not shouting 'Charge!' will be undefined.

[58] It has already been noted that Yablo's proposal is readily adapted, and so too is that suggested by Lewis ([2004]) (cf. Hitchcock [2001a], esp. p. 289).

References

Anscombe, G. E. M. [1971]: *Causality and Determination*, Cambridge: Cambridge University Press.

Cartwright, N. [1979]: 'Causal Laws and Effective Strategies', *Noûs*, **13**, pp. 419-37.

Collins, J., Hall, N. and Paul, L. A. (*eds*.) [2004]: *Causation and Counterfactuals*, Cambridge, MA: MIT Press.

Dowe, P. [2000]: *Physical Causation*, Cambridge: Cambridge University Press.

Dowe, P. [2004]: 'Chance-Lowering Causes', in P. Dowe and P. Noordhof (*eds*.), *Cause and Chance: Causation in an Indeterministic World*, London: Routledge, pp. 28-38.

Eagle, A. [forthcoming]: 'Deterministic Chance', forthcoming in *Noûs*.

Edgington, D. [1997]: 'Mellor on Chance and Causation', *British Journal for the Philosophy of Science*, **48**, pp. 411-33.

Eells, E. [1991]: *Probabilistic Causality*, Cambridge: Cambridge University Press.

Elga, A. [2001]: 'Statistical Mechanics and the Asymmetry of Counterfactual Dependence', *Philosophy of Science*, **68**, Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers, pp. S313-24.

Frigg, R. and Hoefer, C. [forthcoming]: 'Determinism and Chance from a Humean Perspective', in D. Dieks, W. Gonzalez, S. Hartmann, M. Weber, F. Stadler and T. Uebel (*eds*.): *The Present Situation in the Philosophy of Science*, Berlin and New York: Springer.

Glynn, L. [2010]: 'Deterministic Chance', forthcoming in *British Journal for the Philosophy of Science*, doi:10.1093/bjps/axp020.

Good, I. J. [1961a]: 'A Causal Calculus (I)', *British Journal for the Philosophy of Science*, **11**, pp. 305-18.

Good, I. J. [1961b]: 'A Causal Calculus (II)', *British Journal for the Philosophy of Science*, **12**, pp. 43-51.

Good, I. J. [1962]: 'Errata and Corrigenda', *British Journal for the Philosophy of Science*, **13**, p. 88.

Hájek, A. [2003a]: 'Conditional Probability Is the Very Guide of Life' in H. Kyburg Jr. and M. Thalos (*eds*.) *Probability is the Very Guide of Life: the Philosophical Uses of Chance*, Chicago: Open Court, pp. 183-203.

Hájek, A. [2003b]: 'What Conditional Probability Could Not Be', *Synthese*, **137**, pp. 273-323.

Hájek, A. [2007]: 'The Reference Class Problem is Your Problem Too', *Synthese*, **156**, pp. 563-85.

Hall, N. [2000]: 'Causation and the Price of Transitivity', *Journal of Philosophy*, **97**, pp. 198-222.

Hall, N. [2004]: 'Two Concepts of Causation', in Collins, Hall and Paul (*eds*.) ([2004]), pp. 225-76.

Hall, N. [2007]: 'Structural Equations and Causation', *Philosophical Studies*, **132**, pp. 109-36.

Halpern, J. Y. and Pearl, J. [2005]: 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', *British Journal for the Philosophy of Science*, **56**, pp. 843-87.

Hesslow, G. [1976]: 'Two Notes on the Probabilistic Approach to Causality', *Philosophy of Science*, **43**, pp. 290-2.

Hitchcock, C. [1993]: 'A Generalized Probabilistic Theory of Causal Relevance', *Synthese*, **97**, pp. 335-64.

Hitchcock, C. [1996a]: 'The Role of Contrast in Causal and Explanatory Claims', *Synthese*, **107**, pp. 395-419.

Hitchcock, C. [1996b]: 'Farewell to Binary Causation', *Canadian Journal of Philosophy*, **26**, pp. 267-82.

Hitchcock, C. [2001a]: 'The Intransitivity of Causation Revealed in Equations and Graphs', *Journal of Philosophy*, **98**, pp. 273-99.

Hitchcock, C. [2001b]: 'A Tale of Two Effects', *Philosophical Review*, **110**, pp. 361-96.

Hitchcock, C. [2003]: 'Of Humean Bondage', *British Journal for the Philosophy of Science*, **54**, pp. 1-25.

Hitchcock, C. [2004a]: 'Do All and Only Causes Raise the Probabilities of Effects?' in Collins, Hall and Paul (*eds*.) ([2004]), pp. 403-17.

Hitchcock, C. [2004b]: 'Routes, Processes and Chance-Lowering Causes' in P. Dowe and P. Noordhof (*eds*.), *Cause and Chance: Causation in an Indeterministic World*, London: Routledge, pp. 138-51.

Hitchcock, C. [2007a]: 'Prevention, Preemption, and the Principle of Sufficient Reason', *Philosophical Review*, **116**, pp. 495-532.

Hitchcock, C. [2007b]: 'What's Wrong with Neuron Diagrams?', in J. Campbell, M. O'Rourke and H. S. Silverstein (*eds*.), *Causation and Explanation*, Cambridge MA: MIT Press, pp. 69–92.

Hitchcock, C. [2009]: 'Structural Equations and Causation: Six Counterexamples', *Philosophical Studies*, **144**, pp. 391-401.

Humphreys, P. [1989]: *The Chances of Explanation: Causal Explanation in the Social, Medical, and Physical Sciences*, Princeton, NJ: Princeton University Press.

Hoefer, C. [2007]: 'The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance', *Mind*, **116**, pp. 549-96.

Kolmogorov, A. N. [1933]: *Grundbegriffe der Wahrscheinlichkeitrechnung*, Berlin: Springer. Translated as Kolmogorov, A. N. [1950]: *Foundations of the Theory of Probability*, N. Morrison (*ed.*), New York: Chelsea Publishing Company.

Kvart, I. [1991]: 'Transitivity and Preemption of Causal Relevance', *Philosophical Studies*, **64**, pp. 125-60.

Kvart, I. [1994a]: 'Causal Independence', *Philosophy of Science*, **61**, pp. 96-114.

Kvart, I. [1994b]: 'Overall Positive Causal Impact', *Canadian Journal of Philosophy*, **24**, pp. 205-27.

Kvart, I. [1997]: 'Cause and Some Positive Causal Impact', *Philosophical Perspectives*, **11**, pp. 401-32.

Kvart, I. [2004a]: 'Causation: Probabilistic and Counterfactual Analyses' in Collins, Hall and Paul (*eds*.) ([2004]), pp. 359-86.

Kvart, I. [2004b]: 'Probabilistic Cause, Edge Conditions, Late Preemption and Discrete Cases' in

    P. Dowe and P. Noordhof (*eds*.), *Cause and Chance: Causation in an Indeterministic World*,

    London: Routledge, pp. 163-87.

Lewis, D. [1979]: 'Scorekeeping in a Language Game', *Journal of Philosophical Logic*, **8**, pp.

    339-59.

Lewis, D. [1986a]: *Philosophical Papers*, Vol. II, Oxford: Oxford University Press.

Lewis, D. [1986b]: 'Counterfactual Dependence and Time's Arrow' in Lewis ([1986a]), pp. 32-

    52.

Lewis, D. [1986c]: 'Postscripts to "Counterfactual Dependence and Time's Arrow"' in Lewis

    ([1986a]), pp. 52-66.

Lewis, D. [1986d]: 'Causation' in Lewis ([1986a]), pp. 159-72.

Lewis, D. [1986e]: 'Postscripts to "Causation"' in Lewis ([1986a]), pp. 172-213.

Lewis, D. [2004]: 'Causation as Influence', in Collins, Hall and Paul (*eds*.) ([2004]), pp. 75-106.

Loewer, B. [2001]: 'Determinism and Chance', *Studies in History and Philosophy of Modern Physics*, **32**, pp. 609-20.

McDermott, M. [1995]: 'Redundant Causation', *British Journal for the Philosophy of Science*, **46**, pp. 523-44.

Mellor, D. [1995]: *The Facts of Causation*, London: Routledge.

Menzies, P. [1989]: 'Probabilistic Causation and Causal Processes: A Critique of Lewis', *Philosophy of Science*, **56**, pp. 642-63.

Menzies, P. [1996]: 'Probabilistic Causation and the Pre-Emption Problem', *Mind*, **105**, pp. 85-117.

Menzies, P. [2004]: 'Causal Models, Token Causation, and Processes', *Philosophy of Science*, **71**, pp. 820-32.

Papineau, D. [1991]: 'Correlations and Causes', *British Journal for the Philosophy of Science*, **42**, pp. 397-412.

Paul, L. A. [2000]: 'Aspect Causation', *Journal of Philosophy*, **97**, pp. 235-56.

Pearl, J. [2000]: *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Reichenbach, H. [1971]: *The Direction of Time*, M. Reichenbach (*ed.*)*,* Berkeley, CA: University of California Press. First published in 1956.

Rosen, D. A. [1978]: 'In Defense of a Probabilistic Theory of Causality', *Philosophy of Science*, **45**, pp. 604-13.

Salmon, W. C. [1980]: 'Probabilistic Causality', *Pacific Philosophical Quarterly*, **61**, pp. 50-74.

Salmon, W. C. [1984]: *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ: Princeton University Press.

Schaffer, J. [2000a]: 'Overlappings: Probability-Raising Without Causation', *Australasian Journal of Philosophy*, **78**, pp. 40-6.

Schaffer, J. [2000b]: 'Trumping Preemption', *Journal of Philosophy*, **97**, pp. 165-81.

Schaffer, J. [2001]: 'Causes as Probability Raisers of Processes', *Journal of Philosophy*, **98**, pp. 75-92.

Skyrms, B. [1980]: *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*, New Haven, CT: Yale University Press.

Sober, E. [1987]: 'The Principle of the Common Cause', in J. Fetzer (*ed.*), *Probability and Causality: Essays in Honor of Wesley C. Salmon*, Dordrecht: Reidel, pp. 211-28.

Sober, E. [2001]: 'Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause', *British Journal for the Philosophy of Science*, **52**, pp. 331-46.

Spirtes, P., Glymour, C. and Scheines, R. [2000]: *Causation, Prediction, and Search*, Cambridge, MA: MIT Press, 2nd Ed.

Suppes, P. [1970]: *A Probabilistic Theory of Causality*, *Acta Philosophica Fennica*, **24**, Amsterdam: North-Holland Publishing Company.

Yablo, S. [2002]: 'De Facto Dependence', *Journal of Philosophy*, **99**, pp. 130-48.

Yablo, S. [2004]: 'Advertisement for a Sketch of an Outline of a Prototheory of Causation', in Collins, Hall, and Paul (*eds.*) ([2004]), pp. 119-37.