Brit. J. Phil. Sci. **57** (2006), 93–119

Causation: An Alternative

Wolfgang Spohn

ABSTRACT

The paper builds on the basically Humean idea that A is a cause of B iff A and B both occur, A precedes B, and A raises the metaphysical or epistemic status of B given the obtaining circumstances. It argues that in pursuit of a theory of deterministic causation this 'status raising' is best explicated not in regularity or counterfactual terms, but in terms of ranking functions. On this basis, it constructs a rigorous theory of deterministic causation that successfully deals with cases of overdetermination and pre-emption. It finally indicates how the account's profound epistemic relativization induced by ranking theory can be undone.

1 Introduction

Counterfactual analyses of causation have run through many epicycles.[1] It is time to look at a genuine alternative, which in principle has been available for 20 years.[2]

The logic of counterfactuals, as fully developed in (Lewis [1973a]), was an indispensable means for such counterfactual analyses. This is why they developed into a constructive research programme only with (Lewis [1973b]). It has been carried on throughout the years. But it seems it has received greatest attention just recently.[3] Reading all the papers leaves one in bewilderment: too many examples not adequately dealt with or only by increasingly

[1] The major cycles have been produced by David Lewis himself. See Lewis ([1973b], [1986], [2000]). Hints to further cycles may be found there.
[2] It is first presented in (Spohn [unpublished]).
[3] See, e.g. the April issue of the *Journal of Philosophy* **97** (2000), or the collection by Collins *et al.* ([2004]). See also the many references therein, mostly referring to papers since 1995.

doi:10.1093/bjps/axi151

Advance Access published on December 22, 2005

imperspicuous clauses, and too few satisfying theories. The prospects of the programme have clouded. What to do?

First, we must step back from the glaring equation 'A is a cause of B' = 'if A had not happened, B would not have, either', which has been stated for centuries, but could not be theoretically exploited until the advent of the logic of counterfactuals. We should rather look at the other classic formula that a cause is a necessary and/or sufficient condition for its effect under the obtaining circumstances. I shall develop this thought in Section 4. The formula may appear worn out, and all the familiar explications of 'necessary and/or sufficient condition' turned out to be unfit. But there is a thoroughly epistemological notion of 'condition', or 'reason', as I shall say, which is based on so-called ranking theory and which *does the job*. This will be elaborated in Section 3. Before this, Section 2 will briefly address the nature of causal relata as far as needed. An important general lesson of these sections will be that the theory of deterministic causation can be built in *perfect* analogy to the theory of probabilistic causation (as developed from Suppes [1970] up to Spirtes *et al.* [1993], Shafer [1996] and Pearl [2000]).

We will be rewarded in Section 5 where we shall see how to account for some of the most stubborn problem cases in an entirely straightforward way. How can this be? The reader will already have sensed the catch. Obviously, I am going to develop a thoroughly epistemological theory of causation, and this is simply not what we want and what the counterfactual analysts were after. This point is set straight in Section 6 where I shall suggest how to objectivize my account so far relativized to an epistemic subject. This will double the reward; we can explain then why the problem cases are so stubborn for objective theories of causation. In short, there *is* a workable alternative. Let us see how it works.

2 Variables, propositions, time

I have to be quite brief concerning the ontology of causal relata. Ordinary language is an unreliable guide in these matters. We often speak of events, or facts, or states, or even changes as causes and effects (for some early linguistic observations cf. e.g. Vendler [1967]), and then we might enter philosophical argument to clarify the subtle differences between these entities. However, this argument tends to be endless. As far as I see, we face here a largely tactical choice. Purely philosophical discussions of causation tend to get entangled into the notion of an event (as is amply exemplified by the work of David Lewis). By contrast, I observe that state-space terminology prevails among authors concerned rather with scientific applications and less with ontological subtleties. This terminological split is unfortunate, since by taking one side one presents oneself in an unfamiliar way to the other side. The split is all the more unfortunate as the approaches are, I think, intertranslatable to a large extent (and the issues where translatability may fail will not become relevant in this paper). I prefer state-space to event terminology (because this is the one I have grown up with and because it facilitates rigorous theorizing). Formally, of course, I shall use just some set-theoretic construction that is largely open to philosophical interpretation. It is useful to explicitly introduce the construction:

It starts with a set U of variables, a frame; members of U are denoted by x, y, z, etc. subsets of U by X, Y, Z, etc. All definitions to follow, in particular the notion of causation I am going to explicate, will be relative to this frame. This may be cause for concern which I shall take up in Section 6. Each variable can realize in this or that way, i.e. take one of several possible values (and may indeed be conceived simply as the set of its possible values). A small world[4] w is a function that tells how each variable realizes, i.e. assigns to each variable one of its possible values. W denotes the set of small worlds.

Typically, a variable consists of an object, a time, and a family of properties (say, colour, charge, marital status, income, etc.); and a realization of such a variable consists in the object's having at that time a certain property of that family (a certain colour, charge, marital status, income, etc.). This entails that variables are here considered to be specific, not generic.[5] Let me give a slightly extended example. Meteorologists are interested in generic variables like temperature, air pressure, humidity, wind, etc., which can take various values (the latter, for instance, a velocity vector). But these generic variables realize at certain times and places; only then do we have specific variables. Thus, the temperature at noon of 1 January 2004, in Konstanz is a specific variable that may take any value on the Celsius scale and actually took 2°C. For each of the generic meteorological variables there are hence as many specific variables as there are spatiotemporal locations considered by the meteorologist. A small meteorological world, then, is a weather course, that is, a specification of all specific meteorological variables, or of all generic variables for all the locations considered. Those in pursuit of causal laws or correlations tend to consider generic variables. Here, however, we are interested in singular causation which I, as well as counterfactual analyses, take to be primary. Hence, our causal investigation will focus on specific variables.

As usual, propositions are sets of small worlds, i.e. subsets of W; I use A, B, C, etc. to denote them. I could also call them states of affairs. But this is

[4] This allusion to Savage ([1954], Section 5.5) is to emphasize that we are dealing here with restricted well-defined model worlds and not yet with grand Lewisian possible worlds.

[5] This remark is directed to the state-space camp where the point is often unclear. The event camp is not concerned; if one translates events into state-space terminology one automatically ends up with what I call specific variables.

only to say that the subtle difference between the ontological connotation of 'state of affairs' and the epistemological connotation of 'proposition' is not my topic here, though it hides deep problems for any theory of causation.

Let us say that A is a proposition about the set $X \subseteq U$ of variables if it does not say anything about the other variables in $U \setminus X$, i.e. if for any small world w in A all other small worlds agreeing with w within X are also in A. For instance, a proposition about temperatures only consists of small (meteorological) worlds which realize air pressure, humidity, etc., for all locations considered in any way whatsoever. The set of propositions about X is denoted by $\boldsymbol{P}(X)$. Hence, $\boldsymbol{P}(U)$ is the set of all propositions considered. $\boldsymbol{P}(x)$ is short for $\boldsymbol{P}(\{x\})$. Indeed, propositions about single variables, i.e. in $\boldsymbol{P}(x)$ for some $x \in U$, are my candidates for causal relata.

The difference from event ontology is not so large as it may appear. For instance, events in the sense of (Kim [1973]) are the very same as my causal relata, i.e. my propositions about single variables. What Lewis ([2000]) ends up with may be translated into the language of variables, too. He is reluctant to decide how fragile (in his sense) events really are. In any case, an event has very fragile versions (which represent the same event), and it has very fragile alternatives, which may be either different or sufficiently similar. In the latter case, they are alterations of the event just as its versions are. Lewis then explains causation between events in terms of counterfactual dependence between their alterations; the details are not yet relevant, though. Now, the versions of an event and all its alternatives make up for a very fine-grained variable in my sense taking very many possible values, and each version or alternative of the event is a possible realization of the variable. The event itself realizes if a value from some subset of the set of all values realizes. How small this subset is depends on how fragile the event is taken to be. The alterations form a somewhat larger subset. In any case, what Lewis considers as causal relata also qualify as causal relata in my sense.

Since variables are specific, they have a natural temporal order. $x < y$ means that the variable x realizes before the variable y, and if $A \in \boldsymbol{P}(x)$ and $B \in \boldsymbol{P}(y)$, $A < B$ is to say that A precedes B. I shall facilitate matters by assuming that the variables are indeed linearly (not only weakly) ordered in time. Thereby I avoid nasty questions about simultaneous causation and neglect the even less perspicuous case of causation among temporally extended variables that possibly overlap one another. These complications are not our concern.

Finally, I will make the simplifying assumption that the frame U and the set W of small worlds it generates are finite; this entails, in particular, that temporal order is discrete. Loosening this assumption is a primarily mathematical (though philosophically not insignificant) task.

3 Induction first

Let us turn to causation after these preliminaries, and let us, as announced, start from the classic formula abundantly found in the literature: A is a cause of B iff A and B both occur, if A precedes B, and if A is a necessary and/or sufficient condition for B under the obtaining circumstances.[6]

The requirement of the cause preceding the effect is often doubted in the philosophical literature, for reasons I do not understand well. I take this requirement simply for granted. The only implicit argument I shall give is that the theory of causation I shall propose would not work at all without it. Hence, I will leave it open whether this is an argument for temporal precedence or against this theory.

What do 'the obtaining circumstances' refer to? Let us postpone this question to the next section. We should first note that a cause must not be a redundant condition for its effect given the circumstances.[7] If A is, say, a sufficient condition for B given circumstances C, this means that B is necessary given A and C, but not given C alone, i.e. that, given C, A raises the modal status of B from impossibility or contingency to necessity. Likewise, in case A is a necessary condition for B.

Hence, my favourite variant of the classic formula is, generally, this: *A is a cause of B iff A and B both occur, if A precedes B, and if A raises the metaphysical or epistemic status of B given the obtaining circumstances*. This makes explicit the relevance of A. It also adds the basic ambiguity in the notion of a condition between a metaphysical and an epistemic reading, which will acquire great importance later on. And it is even general enough to cover probabilistic causation as well where the statuses are probabilistic ones.

Note that counterfactual analyses are a special case of this general formula. They take the statuses metaphysically as counterfactual necessity and possibility. The temporal precedence is entailed by the constant reminder that all counterfactuals involved in the analysis must be read in a non-backtracking way. And the reference to the obtaining circumstances is always implicit in the antecedent of a counterfactual. However, they are only a special case; stepping back from them means widening the view and seeing what else might fall under the general formula.

Well, what else might fall under it? The traditional Humean view is that the talk of necessary and/or sufficient conditions should be explained in terms of nomological or lawful implication, where laws in turn are taken as mere

[6] This neglects Hume's contiguity condition, which is inexpressible in the framework introduced above, since it leaves out (or implicit) all spatial relations between variables.

[7] This is what Reichenbach's screening-off is about in the probabilistic case and Mackie's INUS conditions in the deterministic case.

regularities. However, I take it that all regularity accounts of causation have failed.[8]

Thus, we are back at Hume's famous question: what more is causal necessity than mere regularity? Hume should not be reduced to the answer: nothing. He was rather peculiarly ambiguous. More prominent in his writings is an associationist theory of causation, according to which the causal relation between two events is constituted by their being associated in our minds. Association, in turn, is explained as the transfer of liveliness and firmness, the marks by which Hume characterizes belief. Thus, we may say in more modern terms that, if A precedes B (and is contiguous to it), A is a cause of B for Hume iff B may be inductively inferred from A (and vice versa). At the same time, this entails a fundamental subjective relativization of the notion of causation.[9]

Here, I fully endorse this subjectivist turn. I shall not try to adduce principled reasons for doing so. My argument rather lies in Sections 5 and 6: this turn is successful where counterfactual and other objectivistic analyses are not, and there is still a way to escape from subjectivism. However, the equivalence of causation and induction was much too quick. We have to underpin the account of causation we are heading for by an elaborate theory of inductive inference. This is the crucial task for the rest of this section.

What might we expect of a theory of inductive inference? The task of induction is to project, from the total evidence we have received, all our beliefs transcending the evidence. The task of belief dynamics is to tell which posterior belief state to assume on the basis of the prior belief state and the evidence received in between. It is next to obvious that these two tasks are essentially equivalent (for details see Spohn [2000]). Hence, what we expect of a theory of induction is no more and no less than an account of doxastic states that specifies not only their static, but also their dynamic laws (understood as laws of rationality).

The form of these laws depends, of course, on the chosen representation of doxastic states. The best elaborated representation is certainly the probabilistic one, for which we have well-argued static and dynamic laws (cf. e.g. Skyrms [1990], ch. 5). But that would lead us to a theory of probabilistic causation.

In pursuit of deterministic causation, we should hence focus on plain belief or acceptance that admits, as it were, only of three grades: each proposition is

8 Including John Mackie's account in terms of INUS conditions. Indeed, contrary to his views in ([1965]) he concludes in ([1974], p. 86) that conditionality cannot be understood in terms of the regularity theory.

9 I believe that the associationist theory is conceptually more basic in Hume. But regularities shape our associations and explain why our associations run rather this way than that way. In this way, the associationist theory may eventually reduce to the regularity theory. It is obvious, though, that Hume's ambiguity between causation as a philosophical relation (regularity) and as a natural relation (association) has provoked many exegetic efforts.

held true, undecided, or held false. The obvious idea is to represent plain belief simply by the set of propositions held true, and the obvious static law for such belief sets is that they be consistent and deductively closed.[10] However, there are no general dynamic laws for doxastic states thus represented. Representing plain belief by extremal probabilities is of no avail, since all laws for changing subjective probabilities fail with the extremal ones.[11] Hence, a different representation is needed in order to account for the dynamics of plain belief.

To cut a long story short, I am still convinced that this is best achieved by the theory of ranking functions.[12] This conviction rests on the fact that ranking theory offers a good solution to the problem of iterated belief revision, and thus a *general* dynamics of plain belief, whereas the discussion of this problem in the belief revision literature has not produced a serious rival in my view (cf. Hansson [1998] or Rott [2003]). So, the next thing to do is to briefly introduce and explain this theory of ranking functions.

The basic concept is very simple:

Definition 1 *κ is a ranking function iff it is a function from the set W of small worlds into the set of non-negative integers such that $\kappa^{-1}(0) \neq \emptyset$. It is extended to propositions by defining $\kappa(A) = min\{\kappa(w) \mid w \in A\}$ for $A \neq \emptyset$ and $\kappa(\emptyset) = \infty$.*

A ranking function κ is to be interpreted as a ranking of disbelief. If $\kappa(w) = 0$, w is not disbelieved and might be the actual small world according to κ. This is why I require that $\kappa(w) = 0$ for some small world w. If $\kappa(w) = n > 0$, then w is disbelieved with rank n. The rank of a proposition is the minimum of the ranks of its members; thus, a proposition is no more and no less disbelieved than the most plausible worlds realizing it. $\kappa(A) = 0$ says that A is not disbelieved, but not that A is believed; rather, belief in A is expressed by disbelief in $\bar{A}$, i.e., $\kappa(\bar{A}) > 0$ or $\kappa^{-1}(0) \subseteq A$. In other words, all and only the supersets of $\kappa^{-1}(0)$ are believed in κ; they thus form a consistent and deductively closed belief set.

If we were only to represent belief, we would have to distinguish only an inner sphere of not disbelieved worlds having rank 0 and an outer shell of the remaining disbelieved worlds having rank > 0. But as we shall immediately see, more shells are needed in order to cope with the dynamics of belief. The picture of shells or spheres reminds of the entrenchment orderings used in belief revision theory or indeed of the similarity spheres used by Lewis for the

[10] This is at least what doxastic logic standardly assumes. There are well-known objections, but no standard way at all to meet them. So I prefer to keep within the mainstream.

[11] Popper measures are often thought to overcome the relevant restrictions of standard probability theory. But they do not go far enough; see (Spohn [1986], [1988]).

[12] Proposed in (Spohn [1988]) under the label ''ordinal conditional functions''. Their first appearance, though, is in (Spohn [unpublished], ch. 5).

semantics of counterfactuals. However, in both pictures the spheres or shells are only ordered. Ranks go beyond by numbering the shells; the arithmetic of ranks will turn out to be crucial.

Two simple, but important properties of ranking functions follow immediately: the *law of negation* that for all $A \subseteq W$ either $\kappa(A) = 0$ or $\kappa(\bar{A}) = 0$ or both, and the *law of disjunction* that for all A, $B \subseteq W$, $\kappa(A \cup B) = \min\{\kappa(A), \kappa(B)\}$.

So far, only disbelief comes in degrees. But degrees of disbelief are tantamount to degrees of belief. It is easy to represent both degrees in one notion:

Definition 2 *β is the belief function associated with the ranking function κ iff for each $A \subseteq W$, $\beta(A) = \kappa(\bar{A}) - \kappa(A)$ (due to the law of negation, at least one of the two terms is 0). β is a belief function iff it is associated with some ranking function.*

Thus, $\beta(\bar{A}) = -\beta(A)$, and A is believed to be true, false, or neither according to β (or κ) depending on whether $\beta(A) > 0$, < 0, or $= 0$. Belief functions may be the more intuitive notion; therefore, I often prefer to use them. However, they are a derived notion; laws and theorems are more easily stated in terms of ranking functions.

The ranks reveal their power when we turn to the dynamics of plain belief. The central notion is given by

Definition 3 *Let κ be a ranking function, and A some non-empty proposition. Then the rank of $w \in W$ given or conditional on A is defined as*

$$\kappa(w \mid A) = \begin{cases} \kappa(w) - \kappa(A) & \text{for } w \in A \\ \infty & \text{for } w \notin A \end{cases}.$$

Similarly, the rank of $B \subseteq W$ given or conditional on A is defined as $\kappa(B \mid A) = \min\{\kappa(w \mid A) \mid w \in B\} = \kappa(A \cap B) - \kappa(A)$. I also call the function $\kappa(. \mid A)$ the A-part of κ. If β is the belief function associated with κ, we finally set $\beta(B \mid A) = \kappa(\bar{B} \mid A) - \kappa(B \mid A)$.

Definition 3 is tantamount to the *law of conjunction* which states that $\kappa(A \cap B) = \kappa(A) + \kappa(B \mid A)$ for all propositions $A \neq \emptyset$ and B. The definition and the law essentially refer to the arithmetic of ranks; a mere ordering (of ranks, of entrenchment, or of similarity spheres) would not do. Indeed, in the relevant literature one finds quite often the proposal and elaboration of a theoretical structure that is more or less equivalent to the above laws of negation and disjunction. The divergence starts with the law of conjunction, which may thus be viewed as the distinctive feature of ranking theory. It has important consequences:

First, it is obvious that a ranking function κ is uniquely determined by its A-part $\kappa(. \mid A)$ its $\bar{A}$-part $\kappa(. \mid \bar{A})$, and the degree $\beta(A)$ of belief in A.

This suggests a simple model for doxastic changes: As is well known, probabilistic belief change is modelled on the assumption that the probabilities conditional on the proposition (or its negation) about which one receives information remain unchanged.[13] Similarly, we can assume here that, if the received information directly concerns only the proposition A (and its negation), only the ranks of A and $\bar{A}$ are changed—such that, say, the posterior rank of A is 0 and that of $\bar{A}$ is n so that A becomes believed with degree $n-$, whereas all the ranks conditional on A and on $\bar{A}$ remain unchanged. Thereby, the doxastic change results in a fully determinate posterior ranking function which one may call the *A,n-conditionalization* of the prior one.

The picture of shells or spheres may again be helpful. If A is not disbelieved in the prior state the effect of A,n-conditionalization is just to add A to the old beliefs (and to draw all logical consequences). This would be absurd, though, if A would be priorly disbelieved. In this case, the effect of A,n-conditionalization is to move to the innermost shell compatible with A; its intersection with A (and all the logical consequences thereof) then constitutes the posterior belief set. In order to allow for a differentiated revision behavior, more than one shell around the inner sphere is needed. So far, all accounts working with in this picture agree. However, for a full and iterated belief dynamics one must not only say what the posterior beliefs are, but also how the systems of spheres get rearranged in revision. This issue is precisely answered by the arithmetical method of A,n-conditionalization, but it presents great difficulties for other approaches. These remarks may suffice for indicating that ranking theory successfully provides a completely general dynamics of belief.[14]

Secondly, this account of conditionalization immediately leads to the crucial notion of doxastic dependence and independence: two propositions are independent iff conditionalization with respect to one does not affect the doxastic status of the other. More generally, two sets of variables are independent iff conditionalization with respect to any proposition about the one set does not affect the doxastic status of any proposition about the other. Or formally:

Definition 4 *Let β be the belief function associated with the ranking function κ. Then A and B are independent given $C \neq \emptyset$ relative to β (or κ) iff $\beta(B \mid A \cap C) = \beta(B \mid \bar{A} \cap C)$, i.e. iff $\kappa(A' \cap B' \mid C) = \kappa(A' \mid C) + \kappa(B' \mid C)$ for all $A' \in \{A, \bar{A}\}$, $B' \in \{B, \bar{B}\}$; unconditional independence results for $C = W$. Moreover, if X, Y, $Z \subseteq U$ are three sets of variables, X and Y are independent given Z relative to β*

[13] This is true of simple conditionalization as well as of generalized conditionalization proposed by Jeffrey ([1965], ch. 11).

[14] For more details, see (Spohn [1988], Section 5). The present paper will use only the precise definition of conditional ranks.

(or κ) iff for all $A \in \boldsymbol{P}(X)$, all $B \in \boldsymbol{P}(Y)$, and all realizations C of Z [or atoms or logically strongest consistent propositions in $\boldsymbol{P}(Z)$] A and B are independent given C w.r.t. β (or κ); unconditional independence results for $Z = \emptyset$.

Unconditional and conditional ranking independence conforms to the same laws as probabilistic independence.[15] This entails in particular that the whole powerful theory of Bayesian nets (cf. Pearl [1988], ch. 3, or, e.g. Jensen [1996]), which rests on these laws, can immediately be transferred to ranking functions.[16] Indeed, it may have become clear in the meantime that ranking functions, though their appearance is quite different, behave very much like probability measures.[17] So, in a way, my further procedure is simply to transfer what can be reasonably said about probabilistic causation to deterministic causation with the help of ranking theory.

Before doing so, we have to add a third and final observation: dependence, which negates independence, may obviously take two forms: positive relevance and negative relevance. Intuitively, we would say that a proposition A is a reason for a proposition B (relative to a given doxastic state) if A strengthens the belief in B, i.e. if the belief in B given A is firmer than given $\bar{A}$. This is something deeply rooted in everyday language; we also say that A supports or confirms B, that A speaks for B, etc. All this comes formally to positive relevance. There are even more ways to express negative relevance; this is, for instance, the essential function of 'but' (cf. Merin [unpublished]). Hence, these notions deserve a formal explication:

Definition 5a *Let β be the belief function associated with the ranking function κ. Then A is a reason for B given C relative to β (or κ) iff $\beta(B \mid A \cap C) > \beta(B \mid \bar{A} \cap C)$. Again, the unconditional notion results for $C = W$.*

According to this definition, being a reason is a symmetric, but not a transitive relation. This is analogous to probabilistic positive relevance, but in sharp contrast to being a deductive reason, which is transitive and not symmetric. However, being a reason thus defined embraces being a deductive reason (which amounts to set inclusion between propositions $\neq \emptyset$, W). Indeed, when I earlier referred to inductive inference, this comes down to

[15] As I was eager to prove in (Spohn [unpublished], Section 5.3), and in (Spohn [1988], Section 6). For a fuller comparison see (Spohn [1994]).

[16] If one notes, moreover, how tight the relation between Bayesian nets and causation is assumed to be (see Spohn [1978], Section 3.3, Spirtes *et al.* [1993], or Pearl [2000]), the bearing of ranking theory on the theory of causation becomes already obvious.

[17] The deeper reason is that ranks may be roughly seen as the orders of magnitude of infinitesimal probabilities in a non-standard probability measure. Thus, by translating the sum of probabilities into the minimum of ranks, the product of probabilities into the sum of ranks, and the quotient of probabilities into the difference of ranks one transforms most theorems of probability theory into ranking theorems. This transition has niceties, though, which are not really clarified; cf. (Spohn [1994], pp. 183–5).

the theory of positive relevance or the relation of being a reason.[18] It is also worth mentioning that being a reason does not presuppose the reason to be actually given, i.e. believed. On the contrary, whether A is a reason for B relative to β is independent of the degree $\beta(A)$ of belief in A.

The value 0 has the special role of a dividing line between belief and disbelief. Therefore, different kinds of reasons must be distinguished:

Definition 5b *Given* C, A *is a*

$$\left\{\begin{array}{c} additional \\ sufficient \\ necessary \\ weak \end{array}\right\} \textit{reason for } B \textit{ w.r.t. } \beta \textit{ iff} \left\{\begin{array}{c} \beta(B \mid A \cap C) > \beta(B \mid \bar{A} \cap C) > 0 \\ \beta(B \mid A \cap C) > 0 \geq \beta(B \mid \bar{A} \cap C) \\ \beta(B \mid A \cap C) \geq 0 > \beta(B \mid \bar{A} \cap C) \\ 0 > \beta(B \mid A \cap C) > \beta(B \mid \bar{A} \cap C) \end{array}\right\}.$$

Hence, if A is a reason for B, it belongs to at least one of these kinds. There is just one way of belonging to several of these kinds; namely, by being a necessary and sufficient reason. Sufficient and necessary reasons are certainly salient. But additional and weak reasons, which do not show up in plain beliefs and are therefore usually neglected, deserve to be allowed for by Definition 5b.

This presentation of ranking theory suffices as a refined substitute for Hume's rudimentary theory of association. Thus equipped, we may return to causation.

4 Causation

We had started with the formula that A is a cause of B if, among other things, A is a necessary and/or sufficient condition for B under the obtaining circumstances, and we have seen that the point is rather that A is a positively relevant condition for B given the circumstances. In all frameworks for deterministic causation I know of and in particular in a regularity as well as in a counterfactual framework, being positively relevant automatically comes down to being a relevant necessary and/or sufficient condition. However, with the richer conceptual resources of the previous section, we may and should distinguish just as many kinds of causes as there are kinds of reasons. This point will become important.

The only thing so far left for clarification are the obtaining circumstances. The most plausible thing to say is that the circumstances relevant for judging the causal relation from A to B consists of all the other causes of B that are

[18] Recall that inductive logic and quantitative confirmation theory were considered to be one and the same project. Recall also that there has been a rigorous, although less successful, discussion of qualitative confirmation theory; cf. the survey of Niiniluoto ([1972]). If (Spohn [2005]) makes sense, it is a promising task to revive qualitative confirmation theory in terms of ranking theoretic positive relevance.

not caused by A. But this is obviously circular.[19] However, the circularity dissolves, if only A's being a *direct* cause of B is considered. In this case there are no intermediate causes, i.e. no causes of B caused by A; the relevant circumstances may hence include all the other causes of B. Moreover, it seems to do no harm when all irrelevant circumstances are added as well, i.e. all the other facts preceding, but not causing B. Thus, we have arrived at conceiving the obtaining circumstances of A's directly causing B as consisting of all the facts preceding B and differing from A.

A slightly more detailed argument (worked out in (Spohn [unpublished], ch. 3 and Section 6.1)) leads to the same result. Given that A and B are facts about single variables and that A precedes B (that's always tacitly understood), A's being a reason for B according to Definition 5 is obviously the deterministic analogue to A's being a *prima facie* cause of B in the probabilistic sense of (Suppes [1970], ch. 2). But the prima facie appearance may change in three ways. First, facts preceding the cause A may turn up which render A irrelevant and thus only a *spurious* cause for B. Think, e.g. of the case of the falling barometer prima facie causing the thunderstorm, but being screened off, of course, by the low air pressure. This case is usually interpreted probabilistically, but has a deterministic reading as well. Secondly, facts realizing between A and B may turn up which render A irrelevant and thus, at most, an *indirect* cause for B. Any deterministic causal chain exemplifies this possibility. These two points were already considered by Suppes. Thirdly, however, if A is irrelevant to B given some condition, further facts preceding the effect B may add to the condition such that A is again positively relevant, and thus apparently a *hidden* cause for B. Suppose you press a switch and, unexpectedly, the light does not go on. You conclude that the switch does not work and that your pressing it had no effect whatsoever. The truth, however, is that someone else accidentally pressed another switch for that light at the very same time. So, given these circumstances, your pressing the switch indeed caused the light *not* to go on.

The three cases entail that every new fact preceding the effect may, in principle, change the assessment of the causal relation from A to B and suggest, respectively, that A is a direct, or an indirect, cause of B, or neither. The assessment is guaranteed to settle only when the whole past of the effect B has been taken into account.[20] But what is the whole past of the effect? Within the given frame U, this can only mean the well-defined past as far as it can be described within this frame; this is the source of the *frame-relativity* of the theory developed here.

[19] This is basically the fundamental objection which Cartwright ([1979]) raised against all probabilistic explications of causation.

[20] It should be clear at this point that facts occurring after the effect have no such force. This does not preclude, of course, that, given incomplete knowledge about the past, the future may carry information about the past, and thus about the causal relation between A and B.

Both this and the previous consideration lead thus to the same explication of direct causation:

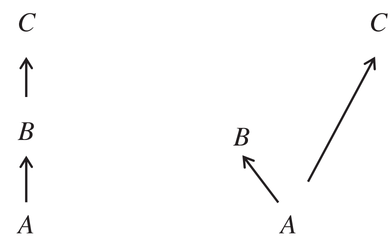
Definition 6 *Let $A \in \boldsymbol{P}(x)$, $B \in \boldsymbol{P}(y)$ for some $x, y \in U$, and $w \in W$. Then A is a direct cause or, respectively, an additional, sufficient, necessary, or weak direct cause of B in the small world w relative to the ranking function κ iff*:

(1) $w \in A \cap B$,

(2) $A < B$,[21]

(3) *A is a reason, or, respectively, an additional, sufficient, necessary, or weak reason for B given $w_{<B,\neq A}$ w.r.t. κ—where $w_{<B,\neq A} = \{w' \mid w'$ agrees with w on $\{z \in U \mid z < y$ and $z \neq y\}\}$ denotes the past of B except A as it is in w* (*which collects, as argued, the obtaining circumstances*).

As an illustration, let us look at the cases of a causal chain and of a conjunctive fork, which are hard or impossible to distinguish for a regularity account, but present no problem to counterfactual analyses.



causal chain *conjunctive fork*

They are easily distinguished also with the help of ranking functions: suppose $\kappa(A) = \kappa(\bar{A}) = 0$, leaving us to specify only the ranks conditional on A and $\bar{A}$. One specification is this:

$\kappa(. \mid A)$	C	$\bar{C}$
B	0	1
$\bar{B}$	2	1

$\kappa(. \mid \bar{A})$	C	$\bar{C}$
B	1	2
$\bar{B}$	1	0

(1) *causal chain*

where the entries in the table give the values of $\kappa(B \cap C \mid A)$ etc.

A is here a direct cause of B in w (where $A \cap B \cap C = \{w\}$), in fact a necessary and sufficient one [because $\kappa(B \mid A) = \kappa(\bar{B} \mid \bar{A}) = 0$ and $\kappa(\bar{B} \mid A) = \kappa(B \mid \bar{A}) = 1$], and indeed the only one due to temporal order. B is a direct cause of C in w, again, a necessary and sufficient one [because $\kappa(C \mid A \cap B) = \kappa(\bar{C} \mid A \cap \bar{B}) = 0$

[21] One may wonder why A is not required to immediately precede B. But clearly, this is inadmissible as long as the frame U may contain any variables whatsoever, and may thus miss variables that are intuitively causal intermediates. I return to this point in Section 6.

and $\kappa(\bar{C} \mid A \cap B) = \kappa(C \mid A \cap \bar{B}) = 1$], indeed the only one because C is independent of A given B as well as given $\bar{B}$ (i.e. the figures just stated would be the same if A were replaced by $\bar{A}$—this is what probability theory refers to as the Markov property). So, we have here an example for a causal chain, in fact the simplest one in which the ranks simply count how many times the obtaining causal relations are violated (in the sequence A, $\bar{B}$, and C, for instance, two such violations occur).

Another specification is this:

$\kappa(. \mid A)$	C	$\bar{C}$
B	0	1
$\bar{B}$	1	2

$\kappa(. \mid \bar{A})$	C	$\bar{C}$
B	2	1
$\bar{B}$	1	0

(2) *conjunctive fork*

As before, A is the only direct cause of B in w, in fact a necessary and sufficient one [because $\kappa(B \mid A)$ etc. are the same as in (1)]. But now, A is also a necessary and sufficient cause of C [because $\kappa(C \mid A \cap B) = \kappa(\bar{C} \mid \bar{A} \cap B) = 0$ and $\kappa(\bar{C} \mid A \cap B) = \kappa(C \mid \bar{A} \cap B) = 1$], and the only one because C is independent of B given A as well as given $\bar{A}$ (i.e. the figures just stated would be the same if B were replaced by $\bar{B}$). We might also say that A screens off B from C. So, we now have the simplest example for a conjunctive fork in which the ranks again just count the violations of the causal relations; the more violations, the more disbelieved.[22] One should note, though, that in both cases the causal relations could be realized through many different distributions of ranks.

One may wonder why the relevant circumstances are now extremely embracive, much more so than intuition requires. The reason is that we have construed 'relevant' extremely weakly. Thus, a lot is relevant. Indeed, all of $w_{<B,\neq A}$ is relevant for the causal relation between A and B, but, as we might say, only potentially relevant on purely temporal grounds. The crucial advantage of this construal is, however, that it is free of any circularity. On this basis we may then search for more restrictive interpretations of 'relevant' which are hopefully provably equivalent to this construal.

The search is indeed successful. We may distinguish five narrower senses of relevant circumstances (for details see Spohn [1990], Section 4, where I have investigated the issue in relation to probabilistic causation). Three of them are provably equivalent to the richest sense above. According to the fourth, the circumstances for the direct causal relation from A to B consist just of all the other direct causes of B, as suggested at the beginning of this section. This is

22 The term'conjunctive fork' has been introduced by Salmon ([1980]), in distinction to what he calls 'interactive forks'. It is still a matter of debate whether the latter can and should be explained away; cf., e.g. (Martel [2003]). My framework, in any case, cannot represent interactive forks as intended by Salmon.

equivalent to the other ones only under special conditions. The fifth sense, finally, is provably equivalent only in the case of necessary and/or sufficient causes. 'Equivalent' means here that the cause's raising of the probability or the rank of the effect is exactly the same given the narrower circumstances as given the richest circumstances. These are essentially satisfying results completing in my view the refutation of the circularity objection of Cartwright ([1979]), and they perfectly carry over to deterministic causation as explained here.

So far, we have dealt only with direct causation. How are we to extend our account to causation in general? I agree with Lewis ([1973b], [2000]) and many others that we should respect our structural intuition that causation is transitive. It goes without saying that direct causes are causes. And, clearly, causal relations should not extend further than direct causal relations—at least as long as we consider only discrete time. The three assumptions entail in fact that causation is the transitive closure of direct causation.

Definition 7 *A is a cause of B in w relative to* κ *(or* β*) if, and only if, there are* $A_1, \ldots, A_n$ $(n \geq 2)$ *such that* $A_1 = A$, $A_n = B$, *and, for all* $i = 1, \ldots, n-1$, A_i *is a direct cause of* A_{i+1} *in w relative to* κ *(or* β*).*

This allows for a lot of causes. Intuitively, though, we speak of much less. This is no cause for worry, however, as has been often observed. Intuitively, we speak of surprising or important causes, of the crucial or most informative cause, etc. But all this belongs to the pragmatics of causal talk. And if there is any hope of doing justice to the pragmatics, it is certainly only by first developing a systematic theory of causation that abstracts from pragmatic considerations, and then trying to introduce the relevant distinctions. My interest is the former, not the latter.

However, there are various real prices to pay for this definition, and this is why we find so much uncertainty in the literature about this issue, mainly, but not only, in the camp of probabilistic causation. One important price is that we thereby decide against the deeply entrenched intuition that causal chains should be something like Markov chains; the relevant conditional independences are not guaranteed by Definition 7. Another important price is that the basic idea that a cause is positively relevant to its effect under the obtaining circumstances, though useful for explicating direct causation, does not generally hold; an indirect cause may well be even negatively relevant to its effect.[23] These prices may well seem too high.[24]

[23] The mutual incompatibilities of the three intuitions are thoroughly explained in (Spohn [1990], Section 5). What I say there for the probabilistic case again applies just as well to the deterministic case.

[24] Hall ([2000]) thoroughly discusses similar conflicts and draws more complicated conclusions. See also (Lewis [2000], pp. 191ff).

The predicament should not be solved by merely pondering about which intuition is weightier or fits the examples better. A more theoretical solution is called for. In my view the following theoretical maxim is decisive: Whenever there are several plausible explications of some notion we are interested in, the theoretically most enlightening procedure is to look for the weakest of these explications; only thereby can we gain theoretical insight about the conditions under which the stronger explications apply as well.

Concerning causation, it is obviously the transitive closure of direct causation that yields the weakest or widest permissible causal relation (within discrete time). The other intuitions, by contrast, would lead to stricter causal relations permitting only shorter causal chains. Thus, the maxim just stated speaks in favour of Definition 7.

Satisfaction of the maxim further demands, then, investigating the conditions under which causal chains as specified in Definition 7 have the desired stronger properties. Section 6 of (Spohn [1990]) contains such an investigation in probabilistic terms. But again, the results obtained there fully carry over to the deterministic case. They do justice to our intuitions to an arguably sufficient extent.

5 Redundant causation

We have already seen that counterfactual analyses and my ranking theoretic account of causation do equally well in distinguishing between causal chains and forks. Let us therefore look at more discriminatory cases. Counterfactual analyses always had a hard time with the various forms of redundant causation. Hence it is interesting to see how these can be handled by the account proposed here.

A and B *redundantly* cause C iff it holds: if neither A nor B had realized, C would not have occurred; but if one of A or B had not realized, in the presence of the other C would still have occurred. The following ranking tables (in terms of belief functions), which restrict themselves to necessary and/or sufficient causes, may be instructive:

$\beta(C \mid .)$	B	$\bar{B}$
A	1	−1
$\bar{A}$	−1	−1

(3) *joint necessary and sufficient causes*

$\beta(C \mid .)$	B	$\bar{B}$
A	1	0
$\bar{A}$	0	−1

(4) *joint sufficient, but not necessary causes*

$\beta(C \mid .)$	B	$\bar{B}$
A	1	1
$\bar{A}$	1	−1

(5) *redundant causes*

One problem with redundant causation is that according to a naive counterfactual analysis neither A nor B is a cause of C. The deeper problem, though, is that redundant causation comes in various forms. In cases of symmetric overdetermination we tend to say that both A and B cause C, whereas in cases of asymmetric pre-emption we want to deny that the pre-empted cause is a cause. However, as long as we present things as in (5) there is no way to give A and B different roles.

In his ([1986], pp. 193–212), Lewis paradigmatically discusses various strategies to settle the issues. One strategy is fine-graining of events. The prince plays the mandolin and simultaneously sings a love song to wake the princess. One musical part would have sufficed for waking. Thus this is a case of overdetermination. But hearing both, the princess wakes up in a slightly different way, which is hence *jointly* caused by the simultaneous performances. Likewise, the poison in the famous desert traveller's keg would have killed him, but it is pre-empted by the hole in the keg. He rather dies of thirst, and that's a different death not producible by the poison. Lewis ([1986], pp. 197–9) explains why he does not want to fully rely on this strategy, and I agree.

He goes on to discuss the other strategy, fine-graining of causal chains.[25] There he arrives at quite complicated conclusions. Let us consider cases of *overdetermination* first. In ([1973b], footnote 12), Lewis declares such cases as useless as test cases because of lack of firm naive opinions about them, something he almost literally repeats in ([2000], p. 182). In his ([1986], pp. 207ff.) he is more optimistic and agrees with Bunzl ([1979]) that fine-graining of causal chains shows most alleged cases of overdetermination to reduce either to ordinary joint causation or to pre-emption via an intermediate Bunzl event, as he calls it. The remaining cases, if there are any, might then be resolved by his doctrine of quasi-dependence. In ([2000]) he repudiates this doctrine. However, the uncertainty has no weight, because the remaining cases are 'spoils to the victor', anyway.

Thus, overdetermination puts a lot of strain on counterfactual analyses. This is in strange disharmony with the great ease with which at least prima facie cases of overdetermination can be produced; they abound in everyday life. Moreover, I do not believe in the uncertainty of pure intuition concerning these cases; uncertain intuitions are already tinged by uncertain theory. My intuition (or my theory) is quite determined. Why not take the prima facie cases at face value? I find it desirable to have a simple account of a simple phenomenon. And there is one. We need neither fine-graining of events nor fine-graining of causal chains: ranks offer a third method for dealing with problem cases. Definition 5b allowed for additional reasons, Definition 6

[25] Salmon ([1980]) already concludes that fine-graining of events and fine-graining of causal chains or, in his terms, 'the method of more detailed specification of events' and 'the method of interpolated causal links' are the two main strategies for dealing with problematic examples.

similarly allowed for additional causes, and this is exactly what overdetermining causes are. This is displayed in the following table:

$\beta(C \mid .)$	B	$\bar{B}$
A	2	1
$\bar{A}$	1	-1

(6) *overdetermining causes*

According to this table, each of A and B would have been a necessary and sufficient cause of C in the absence of the other; in the presence of the other each is still positively relevant to, i.e. a cause of C, but then each can only be an additional cause. This scheme, I find, fits naturally all the intuitive cases of causal overdetermination: usually, if a sufficient cause occurs it is unbelievable that the effect does not occur; this applies to the cases (3) and (4) above. If in a case of overdetermination the effect does not occur, for some or no reason, at least two things appear to have gone wrong at once; and this is at least doubly unbelievable, as represented in (6).

The case of *pre-emption* appears even more complicated from the point of view of counterfactual analyses. Lewis discusses it already in his ([1973b]) where he considers normal cases, as it were, in which fine-graining of causal chains does the trick. The hole in the desert traveller's keg causes him to be thirsty, and the thirst eventually causes his death, but the thirst is no way caused by the pre-empted poison. The poison would rather have caused a heart attack leading to death. But this causal chain never went to completion, it was cut off by the hole in the keg. This solution works in counterfactual as well as in ranking terms.

In ([1986], pp. 200ff.) Lewis calls the easy case *early* pre-emption and distinguishes it from *late* pre-emption, where the causal chain from the pre-empting cause to the effect is somehow empty from the pre-empting action of the pre-empting cause onwards. Thus, in late pre-emption one does not find an event like the above traveller's thirst, and hence the easy solution does not work. According to Lewis ([1986]), late pre-emption can even take three different forms. However, we do not need to discuss them here. Lewis himself takes the first two possibilities to be too far-fetched to worry about and rejects his ([1986]) solution of the third possibility in his ([2000]).

Hall and Paul ([2003]) are not happy with Lewis' presentation of late pre-emption. For them, the mark of late pre-emption is that 'at no point in the sequence of events leading from cause to effect does there fail to exist a backup process sufficient to bring about that effect' (p. 111). And they take this to be an obvious possibility. Their example is Lewis':

> "Suzy and Billy, both throw rocks at a bottle. Suzy is quicker, and consequently it is her rock, and not Billy's, that breaks the bottle. But Billy,

> though not as fast, is just as accurate. Had Suzy not thrown, or had her rock somehow been interrupted mid-flight, Billy's rock would have broken the bottle moments later." (Hall and Paul [2003], p. 110)

Now the obvious asymmetry between Suzy and Billy is the temporal one. Lewis ([2000]) argues that it does not matter whether Suzy's and Billy's breaking the bottle are taken as two versions of the same event or as two alternative events. In any case, one must look at the fine-grained alterations, and then the case is not different from early pre-emption. Suzy's rock touches the bottle, whereas Billy's does not, and thus the causal chain from Suzy to the bottle goes to completion, whereas the one from Billy is cut.

However, Hall and Paul declare the temporal asymmetry to be inessential. They continue:

> It is perfectly easy to construct late pre-emption examples in which, had the cause not occurred—or indeed, had any of the events connecting the cause to the effect not occurred—the effect would have occurred at exactly the same time, and in exactly the same manner .... for example, suppose that the signal from C [= the pre-empting cause] exerts a slight retarding force on the signal from A [= the pre-empted cause]. Pick any point before this signal from C reaches E [= the effect], and ask what would have happened if, at that time, the signal had been absent. Answer: the signal from A would have accelerated, and we can stipulate that it would have accelerated enough to reach E at exactly the time at which the signal from C in fact reaches E. (Hall and Paul [2003], pp. 112f., brackets added by me.)

This is not perfectly easy, it is highly contrived. What is worse, the retarding effect of the chain from C to E on the chain from A to E must become smaller and smaller and converge to 0. Otherwise, there must be a time at which it is too late for the chain from A to arrive at E at the same time as the chain from C. Hence, despite the retarding force of C, the chain from A arrives at E at exactly the same time as the chain from C. And then there is no reason to take the chain from A to E as pre-empted; the case rather seems to be one of symmetric overdetermination.

One may wonder, hence, whether there are any convincing cases of late pre-emption which, by definition, have to be such that fine-graining of causal chains does not reveal an asymmetry. Yes, there are. So far, we have considered only cases of pre-emption that turned out to be cases of *cutting* (possibly after fine-graining) and thus not of late pre-emption. However, Schaffer ([2000]) has forcefully argued that there is also pre-emption by *trumping*. In Lewis' words:

> The sergeant and the major are shouting orders at the soldiers. The soldiers know that in the case of conflict, they must obey the superior officer. But as it happens, there is no conflict. Sergeant and major simultaneously shout 'Advance!'; the soldiers hear them both; the soldiers advance.

> Their advancing is redundantly caused .... But the redundancy is asymmetrical: since the soldiers obey the superior officer, they advance because the major orders them to, not because the sergeant does. The major pre-empts the sergeant in causing them to advance. The major's order *trumps* the sergeant's. (Lewis [2000], p. 183.)

Schaffer insists that his examples should be taken at face value; they are 'intuitively clear' and 'empirically and pretheoretically plausible'.[26] And Lewis concurs:

> We can speculate that this might be a case of cutting. Maybe when a soldier hears the major giving orders, this places a block somewhere in his brain, so that the signal coming from the sergeant gets stopped before it gets as far as it would have if the major had been silent and the sergeant had been obeyed. Maybe so. Or maybe not. *We do not know one way or the other. It is epistemically possible*, and hence it is possible simpliciter, that this is a case of pre-emption without cutting. (Lewis [2000], p. 183, my italics.)

Schaffer shows that four variants of the counterfactual analysis founder at trumping. In response Lewis proposes a fifth that again proceeds in terms of fine-graining events (or rather alterations). In my terminology, he proposes not to look at binary variables ('whether-on-whether dependence'), but rather at more-than-two-valued variables ('how-when-whether-on-how-when-whether dependence'). This may be successful with the major and the sergeant. However, Suppes ([1970], ch. 5) was the first to attempt to specify causal relations between multi-valued variables. This attempt was heroic, but not well received. Indeed, the probabilistic camp prefers to talk only about causal dependence between variables and to be silent on causation between events (cf. e.g. Spirtes *et al.* [1993]). Lewis ([2000]) in effect also favours talking about causal dependence between variables. Insofar I agree with the criticism of Collins ([2000], Section 4) that Lewis changes the topic. Moreover, why should there be no trumping with respect to binary variables? No reason; Lewis would have to argue that the fragility of events (in his sense) entails that there are no binary variables (in my sense).

We seem to be on the wrong track. Fine-graining of causal chains is disallowed by definition, fine-graining of variables or events helps in some cases, but not necessarily in all. However, ranks again offer a straightforward

[26] In (Spohn [unpublished], ch. 3) I have discussed a structurally similar example. It was a case in which it is unclear who trumps whom. Schaffer's main example is one with Merlin and later on Morgana casting a spell to turn the prince into a frog and a wizard's law to the effect that the *first* spell cast on a given day match the enchantment that midnight. In that case Merlin trumps Morgana. As Schaffer argues such a law is even compatible with Lewis' best-system analysis of laws. But suppose we live in a world in which each day when two wizards cast a spell they, perhaps accidentally, cast the same spell. In this case it is not clear who trumps whom. I was unsure what to think about such a case.

account of trumping. Look at the following table:

$\beta(C \mid .)$	B	$\bar{B}$
A	2	2
$\bar{A}$	1	-1

(7) *trumping*

Here, A is a cause of C w.r.t. β independently of B, though only an additional one in the presence of B, whereas B is no cause of C in the presence of A, but a sufficient cause in the absence of A. This matches well the story of the major and the sergeant. The soldiers' disobedience to the major's orders is more incredible than their disobedience to the sergeant's orders.

The reason why these simple accounts of overdetermination and trumping are available to me, but not to any counterfactual theory is obvious: ranking functions specify varying degrees of disbelief and thus also of positive belief, whereas it does not make sense at all, in counterfactual theories or elsewhere, to speak of varying degrees of positive truth; nothing can be truer than true. Hence, nothing corresponding to the schemes (6) and (7) is available to counterfactual theories. In fact, ranking functions have so many more degrees of freedom that I am confident that they are able to account for all kinds of recalcitrant examples. Still, I would like to emphasize that I am not just playing around with numbers. Ranking functions have a perfectly clear epistemological interpretation,[27] and in all formal representations of examples the ranks must be specified in a way that is at least plausible.

6 Objectivization

Ranks other than 0 and 1 as they appear in the cases (6) and (7) are not gratuitous, however; I do not want to deny that there is something puzzling about these cases. So let us finally consider the costs; this will help us to explain the puzzle without thereby depreciating the simple account given so far. The costs should have been clear all along; they consist in subjectively relativizing causation to an observer or epistemic subject. We did not get little in return, I think; indeed, we gained things unattainable to others. There are many philosophers, though, who find the price too high. Therefore, I finally want to indicate at least that there are ways to re-establish objective causation on the subjective basis presented here.

Let me first emphasize, though, that this subjective relativization is not arbitrary. It is a response (indeed, as mentioned in Section 3, Hume's response) to a deep philosophical problem Hume has raised: namely, what is the nature

[27] They can even be measured on a ratio scale by multiple contractions, as Matthias Hild has first shown; cf. (Spohn [1999]).

of nomic and causal necessity? Long is the list of great philosophers who indulged in Hume's subjectivistic turn, equally long the list of those standing firmly objectivistic, and many tried to take some middle course, the most prominent perhaps being projectivism as famously elaborated in Kant's transcendental idealism. Clearly, the issue is anything but settled. Hence, unintuitive as subjectivism certainly appears, it is not philosophically disreputable.

Lewis' response to the deep problem is his doctrine of Humean supervenience. Thereby, he hopes to be able to give an objectivist account of the problematic ilk of laws, counterfactuals, and causation (and even objective probability) instead of just postulating that ilk (as does Armstrong ([1983]) concerning lawhood and Tooley ([1987]) concerning causation). This is not the place to discuss that doctrine (cf. however, Spohn [forthcoming]). I only want to mention that it is not entirely clear how well Lewis succeeds in keeping his enterprise on the objectivistic side. His account of causation is as objective as his account of counterfactuals. The latter again turns on the objectivity of the similarity between possible worlds. There he admits at least that 'plenty of unresolved vagueness remains' (Lewis [1979], p. 472).[28] Moreover, similarity essentially refers to laws, the objectivity of which he tries to save by his 'best-system analysis of laws'. However, Lewis himself acknowledges that 'best' is quite a human category, and he consequently tries to dissolve subjectivistic implications (cf. Lewis [1994], p. 478). So, there is at least cause for concern.

On the other hand, if one starts right on the subjectivistic side as I do, one should at least attempt to accommodate objectivistic intuitions as far as possible (but it is up to the objectivist to decide whether he is satisfied by the offers). I mentioned in Section 3 how I think Hume backed up his associationist definition by a regularity account of causation. If association is replaced by ranks, a more complicated story must be told. Indeed, the objectivization of the account of causation given so far has two aspects that I can only indicate here.

The first is to eliminate the frame-relativity of the account. This may be done by appealing to the universal frame consisting of all variables whatsoever, though this appeal is doubtlessly obscure. The somewhat homelier method is to relate small worlds not to indescribably grand, but just to larger worlds, i.e. to conjecture that the causal relations obtaining relative to a small frame are maintained in the extensions of that frame. It should be a fruitful task, then, to investigate under which conditions the relations within a coarse frame are indicative of those in the refined frame.[29]

[28] At this point I was more attracted then by the 'epistemic approach to conditionals' which Peter Gärdenfors has developed since 1978 (see his [1988]) and from which ranking functions descend.

[29] Spirtes *et al*. ([1993], chapters 6, 7, and 10) have a lot to say about the probabilistic side of this issue.

The second and main aspect of objectivization, however, pertains to the ranking functions. In my account, they played a role corresponding to that of regularities in the regularity theory of causation or to that of the similarity ordering of worlds in Lewis' counterfactual analysis, and they played it more successfully. However, the only interpretation I have offered for them is as subjective doxastic states. So, what we are seeking is a way of viewing them more objectively. Is there such a way?

For some of them, yes.[30] The basic idea is this: We may assume that the propositions generated by the given frame have unproblematic objective truth conditions. Ranking functions, however, usually don't have them. A ranking function may be said to be true or false according to whether the beliefs embodied in it are true or false. But this refers only to ranks being 0 or larger than 0, it does not confer objectivity to varying distributions of ranks larger than 0. Generally, though, we might say that ranking functions are objectivizable to the extent we succeed in uniquely associating them with unquestionably objective propositions.

There is such an association answering our present needs. First observe that a *causal law* L may be associated with each ranking function κ: define L as a big conjunction of material implications, of all implications of the form 'if A and $w_{<B,\neq A}$, then B', whenever A is a sufficient direct cause of B in w relative to κ, and all implications of the form 'if $\bar{A}$ and $w_{<B,\neq A}$, then $\bar{B}$', whenever A is a necessary direct cause of B in w relative to κ. So, L is the conjunction of all causal conditionals obtaining according to κ, reduced to material implications. Hence, L is simply a true or false proposition generated by the given frame.

The crucial question is whether a ranking function can be reconstructed from its associated causal law; our objectivization strategy works to the extent to which this is feasible. The reconstructibility is limited, of course; there are always many ranking functions with which the same causal law is associated. But there is a narrow class of ranking functions which uniquely correspond to their causal laws and may thus be assigned the same truth values as their associated laws. We may call them *fault counting functions*: for a given law L simply define κ_L such that for each $w \in W$, $\kappa_L(w)$ is the number of times the law L is violated in w. In (1)–(5) above, I have used such fault counting functions.

However, even then the unique reconstructibility, and thus the objectivization of ranking functions through causal laws, works only under two conditions: (i) a certain principle of causality is required to hold, and (ii) each direct cause must immediately precede its direct effect (cf. Spohn [1993],

[30] In the following, I give a very rough sketch of what is worked out in formal detail in (Spohn [1993]).

pp. 243–246). These conditions certainly invite further scrutiny and evaluation. Here, I shall confine myself to three concluding remarks:

First, it would have been natural to wonder why Definition 6, explicating direct causation, does *not* require the direct cause to immediately precede the direct effect. Generally, this requirement would have been clearly unreasonable. As long as we do not put any constraints on the frame to be chosen, the frame considered may well omit all intermediate members of a causal chain, and thus represent the causal relation between two temporally quite distant events as a direct one. Hence, it is interesting to see that the temporal immediacy returns in condition (ii) via the objectivizability of causal relations; it is objectivization which requires frames to be rich enough to always provide immediately preceding direct causes.

The second remark is that additional causes (and weak causes) cannot be objectivized according to this theory. The reason is that, if A is an additional cause of B in w relative to κ, the corresponding causal law contains only the material implication 'if $w_{<B,\neq A}$, then B'; this is what is believed in κ. Then, however, we cannot read off from the law whether or not A is positively relevant to B given $w_{<B,\neq A}$. This entails that it is impossible to objectivize my treatment of overdetermination and trumping in the schemes (6) and (7) above which crucially relied on additional causes.[31] Objectively, these cases *must* be explained away, as Bunzl and Lewis have succeeded to a large extent in the case of overdetermination. Lewis refuses to take the same route in the case of trumping, with the surprising hint at epistemic possibilities, which I have italicized in the last of the longer quotations above in Section 5. I suspect here a confusion of epistemic and metaphysical possibilities. Scheme (7) well represents the epistemic possibilities. Objectively, though, one has to inquire how trumping works; and then a more detailed story, about the brains of the soldiers or whatever, has to be told. Thus, there is trouble with overdetermination and trumping also according to my account. The point is, however, that I have both a straightforward account of these cases as well as an explanation of our urge to explain them away.

The final remark is that all this entails a certain view of causal laws. The objectivization just sketched yields two things. On the one hand, it delivers the objectively true or false proposition L. Thus, causal laws reduce to mere regularities, as the regularity theorist always pleaded. On the other hand, it produces the objectivizable ranking functions uniquely corresponding to these propositions. This accounts for the modal (inductive, explanatory, or counterfactual) force of causal laws. It does not do so by simply postulating

[31] (7) clearly does not represent a fault counting function. One may be tempted to think that (6) does; there seem to be two faults when an overdetermined effect does not occur. Objectively, though, there is only one fault in this case, only one event not occurring as expected.

this modal force, as proposed by Armstrong ([1983]) thus provoking bewilderment as to how to distinguish presence from absence of the modal force. It rather gives a Humean explanation of that modal force via ranking theory and the appertaining theory of objectivization by uniquely associating with a causal law a characteristic inductive behaviour encoded in the corresponding ranking function.[32] All this would have been entirely out of reach, however, without a general theory of inductive behaviour or of doxastic dynamics applying to plain belief, as I have presented it in Section 3.

Acknowledgements

I am indebted to an anonymous referee whose extensive and careful comments led to numerous improvements of the paper.

Fachbereich Philosophie
Universität Konstanz
78457 Konstanz
Germany
wolfgang.spohn@uni-konstanz.de

References

Armstrong, D. M. [1983]: *What is a Law of Nature?*, Cambridge: Cambridge University Press.

Bunzl, M. [1979]: 'Causal Overdetermination', *Journal of Philosophy*, **76**, pp. 134–50.

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

Collins, J. [2000]: 'Pre-emptive Prevention', *Journal of Philosophy*, **97**, pp. 223–34.

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

Gärdenfors, P. [1988]: *Knowledge in Flux*, Cambridge, MA: MIT Press.

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

Hall, N. and Paul, L. A. [2003]: 'Causation and Pre-emption', in P. Clark and K. Hawley (*eds*), *Philosophy of Science Today*, Oxford: Clarendon Press, pp. 100–30.

Hansson, S.-O. [1998]: 'Revision of Belief Sets and Belief Bases', in D. M. Gabbay and P. Smets (*eds*), *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, Volume **3**, *Belief Change*, Dordrecht: Kluwer, pp. 17–75.

Jeffrey, R. C. [1965]: *The Logic of Decision*, Chicago: Chicago University Press.

Jensen, F. V. [1996]: *An Introduction to Bayesian Networks*, London: UCL Press.

[32] I have elaborated on this view of laws in (Spohn [2002]).

Kim, J. [1973]: 'Causation, Nomic Subsumption, and the Concept of an Event', *Journal of Philosophy*, **70**, pp. 217–36.

Lewis, D. [1973a]: *Counterfactuals*, Oxford: Blackwell.

Lewis, D. [1973b]: 'Causation', *Journal of Philosophy*, **70**, pp. 556–57.

Lewis, D. [1979]: 'Counterfactual Dependence and Time's Arrow', *Noûs*, **13**, pp. 455–76.

Lewis, D. [1986]: 'Postscripts to "Causation"', in D. Lewis (*ed.*), *Philosophical Papers*, Volume II, Oxford: Oxford University Press, pp. 172–213.

Lewis, D. [1994]: 'Humean Supervenience Debugged', *Mind* **103**, pp. 473–90.

Lewis, D. [2000]: 'Causation as Influence', *Journal of Philosophy*, **97**, pp. 182–97.

Mackie, J. L. [1965]: 'Causes and Conditions', *American Philosophical Quarterly*, **2**, pp. 245–64.

Mackie, J. L. [1974]: *The Cement of Universe*, Oxford: Clarendon Press.

Martel, I. [2003]: 'Indeterminism and the Causal Markov Condition', Working Paper Series *Philosophy and Probability* No. **4**, Philosophy, Probability, and Modeling Research Group, University of Konstanz.

Merin, A. [unpublished]: *Die Relevanz der Relevanz*, unpublished Habilitationsschrift, Stuttgart, engl. translation forthcoming.

Niiniluoto, I. [1972]: 'Inductive Systematization: Definition and Critical Survey', *Synthese*, **25**, pp. 25–81.

Pearl, J. [1988]: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann.

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

Rott, H. [2003]: 'Coherence and Conservatism in the Dynamics of Belief. Part II: Iterated Belief Change Without Dispositional Coherence', *Journal of Logic and Computation*, **13**, pp. 111–45.

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

Savage, L. J. [1954]: *The Foundations of Statistics*, New York: Dover.

Schaffer, J. [2000]: 'Trumping Pre-emption', *Journal of Philosophy*, **97**, pp. 165–81.

Shafer, G. [1996]: *The Art of Causal Conjecture*, Cambridge, MA: MIT Press.

Skyrms, B. [1990]: *The Dynamics of Rational Deliberation*, Cambridge, MA: Harvard University Press.

Spirtes, P., Glymour, C. and Scheines, R. [1993]: *Causation, Prediction, and Search*, Berlin: Springer.

Spohn, W. [1978]: *Grundlagen der Entscheidungstheorie*, Kronberg/Ts.: Scriptor (out of print, pdf-version at: http://www.uni-konstanz.de/FuF/Philo/Philosophie/Mitarbeiter/spohn.shtml).

Spohn, W. [unpublished]: *Eine Theorie der Kausalität*, unpublished Habilitationsschrift, München.

Spohn, W. [1986]: 'The Representation of Popper Measures', *Topoi*, **5**, pp. 69–74.

Spohn, W. [1988]: 'Ordinal Conditional Functions. A Dynamic Theory of Epistemic States', in: W. L. Harper and B. Skyrms (*eds*), *Causation in Decision, Belief Change, and Statistics*, Volume II, Dordrecht: Kluwer, pp. 105–34.

Spohn, W. [1990]: 'Direct and Indirect Causes', *Topoi*, **9**, pp. 125–45.

Spohn, W. [1993]: 'Causal Laws are Objectifications of Inductive Schemes', in J. Dubucs (*ed.*), *Philosophy of Probability*, Dordrecht: Kluwer, pp. 223–52.

Spohn, W. [1994]: 'On the Properties of Conditional Independence', in: P. Humphreys (*ed.*), *Patrick Suppes: Scientific Philosopher*, Volume 1 of *Probability and Probabilistic Causality*, Dordrecht: Kluwer, pp. 173–94.

Spohn, W. [1999]: 'Ranking Functions, AGM Style', in B. Hansson *et al.* (*eds*), *Internet Festschrift for Peter Gärdenfors*, Lund, <www.lucs.lu.se/spinning/>.

Spohn, W. [2000]: 'Wo stehen wir heute mit dem Problem der Induktion?', in R. Enskat (*ed.*), *Erfahrung und Urteilskraft*, Würzburg: Königshausen & Naumann, pp. 151–64.

Spohn, W. [2002]: 'Laws, Ceteris Paribus Conditions, and the Dynamics of Belief', *Erkenntnis*, **57**, pp. 373–94.

Spohn, W. [2005]: 'Enumerative Induction and Lawlikeness', *Philosophy of Science*, **72**, pp. 164–87.

Spohn, W. [forthcoming]: 'Chance and Necessity: From Humean Supervenience to Humean Projection', in E. Eells and J. Fetzer (*eds*), *The Place of Probability in Science*, Chicago: Open Court.

Suppes, P. [1970]: *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.

Tooley, M. [1987]: *Causation*, Oxford: Oxford University Press.

Vendler, Z. [1967]: 'Causal Relations', *Journal of Philosophy*, **64**, pp. 704–13.