Deterministic Causation*

Wolfgang Spohn

Fachgruppe Philosophie
Universität Konstanz
D-78457 Konstanz

Abstract: This paper is the most complete presentation of my views on deterministic causation. It develops the deterministic theory in perfect parallel to my theory of probabilistic causation and thus unites the two aspects. It also argues that the theory presented is superior to all regularity and all counterfactual theories of causation.

1. Introduction

In Spohn (1990), which is my most complete (though not complete) presentation of my thoughts on probabilistic causation, I claim at the end of the introduction „that each consideration, definition, and theorem" of that paper „can be routinely extended to deterministic causation with the help of the theory of ordinal conditional functions", as I called them then, or ranking functions, as they are called now. Since this remark, which I found obvious, remained unnoticed I intend in this paper to answer the commitment I incurred by this remark.[1]

The plan of the paper is very simple. Section 2 is preparatory and introduces the ontology of events or states of affairs to be used in all later sections. Section 3 briefly rehearses why regularity theories of causation won't do. This is to motivate the subjectivistic turn so courageously proposed by David Hume, though, at other times, he also clearly felt its absurdity. This turn consists, in effect, in relativizing causal relations to an observer or to her doxastic state. I take this turn in section 4. Ist main task is to introduce ranking functions which most adequately serve our purposes as formal representations of doxastic states. These functions are all-important; if this paper improves at all

* To appear in: W. Spohn, M. Ledwig, M. Esfeld (eds.), *Current Issues in Causation*, Mentis, Paderborn 2000.

[1] In Spohn (1991) and (1993) I presented only the beginnings of my account of deterministic casusation, because I pursued other interests in these papers. However, the basic theory is already contained in Spohn (1983b).

upon theories of deterministic causation, it's due to them. It will then be a short step to explicate the notion of direct causation in section 5. This section will also briefly investigate the circumstances of direct causal relationships. Section 6 extends the explication to indirect causation. In section 7, I compare my theory with the main alternative, the counterfactual approach of Lewis (1973b). Section 8, finally, sketches my proposal in Spohn (1993) to undo the subjectivistic relativization I have so heavily engaged in.

2. Variables, Propositions, Time

Let's start with assuming 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.[2] Each variable can realize in this or that way, i.e. can take one of several possibles values. *A small world w* is a function which tells how each variable realizes, i.e. assigning to each variable one of its possible values. W denotes the set of small worlds.

Variables are considered to be specific, not generic. Let me give an example: meteorologists are interested in generic variables like temperature, air pressure, humidity, wind etc., which can take various values (the latter, for instance, a direction and a velocity). But these generic variables realize at certain times and places; only then do we have specific variables. Thus, the temperature at noon of October 14, 1999, in Konstanz is a specific variable which may take any value on the Celcius scale and actually took 16^0C. For each of the generic meteorological variables there are hence as many specific variables as there are spatiotemporal locations considered by the meteorologist. Only the set of specific variables can be subject to causal investigation.

As usual, *proposition*s are sets of small worlds, i.e. subsets of W; I use $A, B, C, D, E,$ etc. to denote them. I could more clumsily call them states of affairs as well. But this is 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 certainly 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 small worlds agreeing with w within X, but differing from w outside X are also in A. For instance, a proposition about temperatures only consist of small (meteorological) worlds which realize air pressure, humidity, etc. for all locations considered in any way whatso-

[2] This may be a cause for concern which I take up in sect. 8.

ever. The set of propositions about X is denoted by $\mathbf{P}(X)$, and $\mathbf{P}(x)$ is short for $\mathbf{P}(\{x\})$. Hence, $\mathbf{P}(U)$ is the set of all propositions considered.

Propositions about single variables, i.e. in $\mathbf{P}(x)$ for some $x \in U$, are my candidates for causal relata. Since variables are understood to be specific, such propositions are just what Kim (1973) calls events and takes as causal relata. They are clearly not events in the sense of Davidson (1985). However, events in the latter sense are not suited for causal theorizing, anyway. All theories considered below require that causal relata be something propositional, subject to logical operations and relations. Although it is hard to see how any causal theory could dispense with this, Davidsonian events are not of this kind. This does not seem to be a good objection; we might simply switch to the proposition *that* a Davidsonian event ccurs. But such propositions are usually much too specific or fine-grained for causal purposes.[3] So, without looking deeper into this issue, we might certainly say that my candidates for causal relata are good and normal ones.

Since variables are specific, they have a natural temporal order. I shall greatly simplify 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 which possibly overlap one another. Without doubt, it would also be worthwhile to study other structures like the partial order of space-time points on Minkowski space, etc. But since my interest pertains to other essential features of causation, I leave out these comlications.

This gives rise to another bit of notation: $x<y$ says that the variable x realizes before the variable y, and if $A \in \mathbf{P}(x)$ and $B \in \mathbf{P}(y)$, we may also write $A<B$ to express that A precedes B. $U_{<y}$ is to denote the *past of* y, i.e. the set of variables earlier than y. We will frequently encounter the expression $U_{<y,\neq x}$ which denotes the past of y except x, i.e. the set of variables earlier than y and different from x. Correspondingly, $w_{<y,\neq x} = \{w' \mid w'$ agrees with w on $U_{<y,\neq x}\}$ denotes the past of y except x in w. Again, if $A \in \mathbf{P}(x)$ and $B \in \mathbf{P}(y)$, I often write $U_{<B,\neq A}$ and $w_{<B,\neq A}$ instead of $U_{<y,\neq x}$ and $w_{<y,\neq x}$.

Finally, I will assume that the frame U and the set W of small worlds it generates are finite; this entails in particular that the temporal order is discrete. This makes things much easier. Loosening this assumption is foremost a mathematical task. This is not my concern here, even though it is not philosophically insignificant.

3. Regularity Theories Fail

[3] Cf. Lewis (1986b).

Enough of preliminaries; let's move immediately to causation. The common starting point which abounds in the literature is something like this: A is a necessary and/or sufficient 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.[4]

The necessity of the temporal precedence of the cause over the effect is often doubted in the philosophical literature, for reasons I do not understand well. I take this necessity 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 are „the obtaining circumstances"? Let's defer the question, since it will receive a clear answer in section 5.

The biggest problem with our starting point is doubtlessly its talk of A being a necessary and/or sufficient condition for B. The most straightforward understanding of this is that A is a nomologically necessary and/or sufficient condition for B; this is one characteristic of a regularity theory of causation. The other is a particularly simple understanding of the laws alluded to in nomological implication; namely, as some further true proposition without a distinguished modal status. This is the real import of regularities; they just happen to obtain. This amounts to the following explication:

(R) Let $A \in \mathbf{P}(x)$ and $B \in \mathbf{P}(y)$ for some $x, y \in U$, and let $L \in \mathbf{P}(U)$ be the conjunction of the obtaining regularities, $S \in \mathbf{P}(U)$ the obtaining circumstances, and w some small world in $L \cap S$. A is then a (i) sufficient, (ii) necessary cause of B in w iff:

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

(2) $A < B$,

(3) (i) $A \cap L \cap S \subseteq B$, (ii) $\overline{A} \cap L \cap S \subseteq \overline{B}$,

(4) (i) not $L \cap S \subseteq B$, (ii) not $L \cap S \subseteq \overline{B}$.

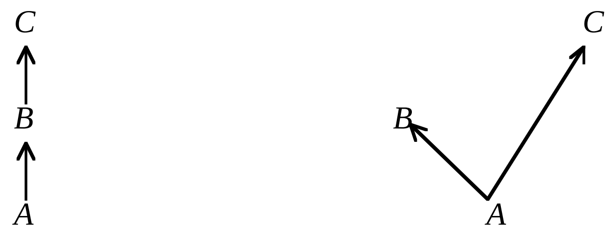
(1) says that cause and effect obtain in the world considered. (2) says that the cause precedes the effect. (3) expresses that the cause is a condition of the effect as now understood. (4) is a kind of relevance condition (usually tacitly understood), saying that the obtaining of the effect is not decided by the laws and circumstances alone.[5]

---

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

[5] This condition is required for dealing with the problem of the irrelevant law specialization by Salmon (1970), sect. 2, and famously exemplified by explaining John's non-pregnancy by his taking contraceptive pills. By the way, condition (4) (ii) is redundant, since L, S, and B all obtain w.

The basic problem of the regularity theory (R) is, as has been argued many times[6], that it cannot distinguish between causal chains and conjunctive forks:



(a) (causal chain) (b) (conjunctive fork)

But, of course, this distinction must be made by any theory of causation. It is no surprise that this fundamental flaw entails further gross defects.

Why is (R) unable to make this distinction? Ignore, for a while, the obtaining circumstances. Suppose all causal relations in (a) and (b) are necessary and sufficient.[7] This requires the very same law L to hold in (a) and in (b); namely, $(A \cap B \cap C) \cup (\overline{A} \cap \overline{B} \cap \overline{C})$. Therefore, (R) loses the distinction between (a) and (b), and wrongly classifies B in (b) as a cause of C and not as a symptom of C as it should.

The situation is a bit less obvious if we take the obtaining circumstances into account. Suppose again all causal relations to be necessary and sufficient; thus, L is as above. Clearly, we must not assume one set of circumstances for all causal relations considered; the circumstances S in (R) are, in fact, relative to the cause A and the effect B. So let S_{AB}, S_{AC}, and S_{BC}, respectively, be the circumstances relevant for judging the causal relationship between A and B, A and C, and B and C. B must not belong to the circumstances S_{AC} in both (a) and (b); otherwise, A would be no cause of C according to (R), because condition (4) would be violated. This is as it should be. The crucial question is: does A belong to the circumstances S_{BC}? It must not in case (a), if B is to come out right as a cause of C. But it must in case (b), if B is to turn out only as a symptom of C. Why, however, should A be part of the circumstances S_{BC} in the one case, but not in the other? There is no intrinsic reason; the only reason is that one knows beforehand that the two cases should be treated differently. This points to a circularity, since the regularity theory cannot define causal relations via an independently established notion of relevant circumstances. The determination of the relevant circumstances depends rather on the causal relationships.

[6] Cf., e.g., Mackie (1974), p. 81ff.

[7] This assumption can be weakened in obvious ways.

Put thus briefly, the argument is not yet conclusive; but I think all experience shows that the circularity cannot be convincingly dissolved within the confines of the regularity theory.[8] If the experience is correct, this theory indeed does not provide any means for distinguishing the two cases.

Intuitively, however, we have no difficulty at all. If *A* occurs, but *B* does not (for some surprising reason or even none), we would know in both cases what to think about *C*. In case (a), we would expect *C* not to occur as well. In case (b) the failure of *B* does not affect the causal relation from *A* to *C* so that *C* is still to occur. The deep point behind this is that the supposition that *A*, but not *B*, obtains is not only counterfactual (relative to the small world considered), but also *counternomological*, and that the regularity theory is entirely unfit for dealing with such counternomological suppositions. Therefore, no regularity theory can succeed.

The case would be totally different, if laws had some special status which, unlike mere truth, did not automatically spread to their logical consequences (so that, for instance, the relation between *B* and *C* in the case (b) of a conjunctive fork does not acquire law-like status, even though the relations between *A* and *B*, and *A* and *C*, have it). Then, however, we are back at Hume's famous question: what more is causal necessity than mere regularity?

4. Induction

Hume was peculiarly ambiguous about the question. At times, he said there is nothing more to causal necessity, thus maintaining an objective notion of causation and falling victim at the same time to the above deficiencies. Mainly, though, he held an associationist theory of causation, according to which the causal relation between two events is constituted by their being associated in our minds. Thus, causation is subjectively relativized.[9] Association, in turn, is explained as the transfer of liveliness and firmness,

---

[8] Since INUS-conditions, with conditionality understood in the sense of the regularity theory, are but a variant of sufficient causes, as defined in (R), the same criticism applies to them. Indeed, contrary to his views in (1965), Mackie (1974), p. 86, concludes that he must leave the grounds of the regularity theory. However, Graßhoff, May (2000) argue that this move is premature and that the regularity theory can be saved.

[9] I believe that the associationist theory of causation is conceptually more basic in Hume. But regularities shape our associations and explain why our associations run rather this way than that way. Thus, the associationist theory may eventually reduce to the regularity theory. It should be noticed, however, that Hume's ambiguity between causation as a philosophical relation (regularity) and as a natural relation (association) has caused many efforts at interpretation; see, e.g., Mackie (1974), ch. 1, and Beauchamp, Rosenberg, (1981), ch. 1.

the marks by which Hume characterizes belief. Thus, if *A* precedes *B* (and is contiguous to it), *A* is a cause of *B* for Hume if and only if *B* may be inductively inferred from *A* (and vice versa) – inductive inference understood in a very wide sense, embracing all kinds of plausible or defeasible inference which lead from one belief to another.

Here, I fully endorse this subjectivist turn. I am not sure whether there are strong principled reasons for doing so. My reason is rather that I am not satisfied by the alternatives, and that this turn results in a good theory of deterministic causation which can even regain objectivity – as I try to show in the sequel.

However, the equivalence of causation and induction was just arrived much too fast; if suffers from two weaknesses. First, it still does not distinguish between causes and symptoms; symptoms also allow one to inductively infer what they are symptoms for. Second, the account now envisaged needs to be underpinned by an elaborate theory of inductive inference. The latter is the task of this section; the first weakness will then be easily dispelled.

So, what might we expect a theory of inductive inference to yield? No more and no less than a dynamic account of doxastic states which specifies not only their static laws, but also their laws of change (which should be understood as laws of rationality). The form of these laws depends, of course, on how doxastic states are represented. The best elaborated representation is certainly the probabilistic, for which we have well-argued static and dynamic laws.[10] But that would lead us to a theory of probabilistic causation.

In pursuit of deterministic causation, we should hence focus rather on plain belief which holds a proposition to be true, false, or neutral, and thus admits only three grades. Prima facie, plain belief is straightforwardly represented by the set of propositions held to be true. The obvious static law for such belief sets is that they be consistent and deductively closed.[11] 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 extremal probabilities. 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 proposed in Spohn (1988) under the label „ordinal conditional functions“.[12] This conviction rests on the fact that ranking functions offer a good solution to the problem of iterated belief revision, and thus a general dynamics of plain be-

---

[10] See, e.g., Skyrms (1990), ch. 5.

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

[12] Their first appearance, though, is in Spohn (1983b), ch. 5.

lief, whereas the discussion of this problem in the belief revision literature[13] has not produced a serious rival in my view. 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 nonnegative 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 $\kappa(\emptyset) =$ Ï .

A ranking function κ is to be interpreted as a ranking of disbeliefs. 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 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 $\overline{A}$, i.e. $\kappa(\overline{A}) > 0$ 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.

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 $\kappa(\overline{A}) = 0$ 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 it is easy to represent also degrees of positive belief:

Definition 2: β is the *belief function associated with* the ranking function κ if and only if for each $A \subseteq W$ $\beta(A) = \kappa(\overline{A}) - \kappa(A)$[14]; and β is a *belief function* iff it is associated with some ranking functions.

Thus, $\beta(\overline{A}) = -\beta(A)$, 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; all 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

---

[13] See, e.g., Hansson (1998), sect. 4.

[14] Note that, due to the law of negation, at least one of the two terms is 0.

Definition 3: Let κ be a ranking function and $\emptyset \neq A \subseteq W$. Then the rank of $w \in W$ *given* or *conditional on A* is defined as $\kappa(w|A) = \begin{cases} \kappa(w) - \kappa(A) \text{ for } w \in W, \\ \infty \text{ for } w \notin A. \end{cases}$ Similarly, the rank of $B \subseteq W$ *given* or *conditional on A* is defined as $\kappa(B|A) = \min\{\kappa(w|A) | w \in B\} = \kappa(A \cap B) - \kappa(A)$. I also call the function $\kappa(.|A)$ the *A-part* of κ. If β is the belief function associated with κ, we finally set $\beta(B|A) = \kappa(\overline{B}|A) - \kappa(B|A)$.

Definition 3 is tantamount to the *law of conjunction* which states that $\kappa(A \cap B) = \kappa(A) + \kappa(B|A)$ for all propositions $A \neq \emptyset$ and B. Moreover, definition 3 entails the *law of disjunctive conditions* that $\kappa(C \mid A \cup B)$ is always between $\kappa(C \mid A)$ and $\kappa(C \mid B)$.[15]

It is obvious that a ranking function κ is uniquely determined by its A-part $\kappa(.|A)$ its $\overline{A}$-part $\kappa(.|\overline{A})$, 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.[16] Similarly, we can assume here that, if the information one receives directly concerns only the proposition A (and its negation), only the ranks of A and $\overline{A}$ are changed, whereas all the ranks conditional on A and on $\overline{A}$ remain unchanged. Therefore, the doxastic change results in a new ranking function in a fully determinate way. These remarks may suffice for indicating that ranking functions indeed allow for a fully general dynamics of plain belief.[17]

The account of conditionalization just given leads immediately to the notion of doxastic dependence and independence which will be crucial: 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 $\beta(B|A \cap C) = \beta(B|\overline{A} \cap C)$, i.e. iff $\kappa(A' \cap B'|C) = \kappa(A'|C) + \kappa(B'|C)$ for all $A' \in \{A, \overline{A}\}$, $B' \in \{B, \overline{B}\}$; unconditional independence results for $C = W$. Moreover, if $X,Y,Z \subseteq U$ are three sets of variables, X and Y

[15] This holds because $\kappa(C \mid A \cup B) = \kappa(C \cap (A \cup B)) - \kappa(A \cup B) = \min[\kappa(A \cap C), \kappa(B \cap C)] - \min[\kappa(A), \kappa(B)]$ and because $\min(n_1, n_2) - \min(m_1, m_2)$ is always between $n_1 - m_1$ and $n_2 - m_2$.

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

[17] For more details, see Spohn (1988), sect. 5, and (1999). This paper will use only the precise definition of conditional ranks.

are independent given Z relative to β (or κ) iff for all $A \in \mathbf{P}(X)$, all $B \in \mathbf{P}(Y)$, and all realizations C of Z (or atoms or logically strongest, contingent propositions in $\mathbf{P}(Z)$) 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.[18] This entails that the whole powerful theory of Bayesian nets[19], which rests on these laws, can immediately be transferred to ranking functions. Indeed, it may have become clear in the meantime that ranking functions, though their appearance is quite different, behave very much like probability measures.[20] The plan for the rest of the paper is therefore, in a way, very simple: just transfer what can be reasonably said about probabilistic causation to deterministic causation with the help of ranking functions.

Before doing so, we have to add a final observation: dependence, which negates independence, may obviously take two forms: positive relevance and negative relevance. These are of utmost epistemological importance. 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 $\overline{A}$. This is something that is 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“.[21] 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 $\beta(B|A \cap C) > \beta(B|\overline{A} \cap C)$. 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

---

[18] As I was eager to prove in Spohn (1983b), sect. 5.3, and (1988), sect. 6. For a fuller comparison see Spohn (1994).

[19] Cf. e.g., Pearl (1988), ch. 3, or Jensen (1996).

[20] 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, replacing multiplication of probabilities by addition of ranks, and addition of probabilities by taking the minimum of ranks, is a simple procedure transforming most theorems of probability theory into theorems about ranking functions. This transition has niceties, however, which are not really clarified; cf. Spohn (1994), pp. 183-185.

[21] Cf. Merin (1996).

defined embraces being a deductive reason (which amounts to set inclusion between contingent propositions). Indeed, when I earlier referred to inductive inference, this comes down to the theory of positive relevance or the relation of being a reason.[22] 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 can be distinguished:

Definition 5b: A is a

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

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. They will acquire some importance later on. To be sure, all these notions can analogously be defined with respect to negative relevance or the relation of being a counter-reason.

The theory of ranking functions developed thus far will suffice as a refined substitute for Hume's rudimentary theory of association. Let us return, thus equipped, to causation.

5. Direct Causation and its Circumstances

We started in section 3 with the idea 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. But when proceeding to the regularity theory (R), we saw that the point is rather that A is a positively relevant condition for B, and that in the old regularity framework (indeed in all frameworks conceived so far for deterministic causation), being a positively relevant

[22] Recall that inductive logic and qualitative 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 in Niiniluoto (1972). It would be most promising, I believe, to revive qualitative confirmation theory in terms of positive relevance with respect to ranking functions

condition automatically came down to being a necessary and/or sufficient condition along the lines of the clauses (3) and (4) of (R). But with the richer conceptual resources of the previous section, we should distinguish just as many kinds of causes as kinds of reasons.

What is left for clarification, hence, are the obtaining circumstances. As already indicated in my rebuttal of the regularity theory, we seem to run into trouble here. 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*[23] which are not caused by *A*.But this is obviously circular.[24] However, the circularity dissolves, if only *A*'s being a direct cause of *B* is considered; then there are no intermediate causes, i.e. no causes of *B* caused by *A*. Thus, the relevant circumstances may include all other causes of *B* in this case. 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[25] 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). 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 read probabilistically, but may be understood deterministically as well. Second, facts occurring 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. And third, 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 a hidden cause of *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 had no effect whatsoever. The truth is, however, that someone else accidentally pressed another swith for that light at the very same time. So, given these circumstances, your pressing the switch indeed caused the light *not* to go on.

---

[23] Clearly, any event may have several causes. Thus, „cause“ should here always be construed as „partial cause“, as is commonly done, and not as „total cause“.

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

[25] Presented in Spohn (1983a) and fully worked out in Spohn (1983b), ch. 3 and sect. 6.1.

Te 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. Only when the whole past of the effect B has been taken into account can the assessment not change any more. (It should be clear at this point that facts occurring after the effect have no such force.[26]) 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 is statable in this frame; this is the source of the frame-relativity of the theory developed here.

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

Definition 6: Let $A \in \mathbf{P}(x)$, $B \in \mathbf{P}(y)$ 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) $A < B$,[27]

(3) A is a reason, or, resepctively, an additional, sufficient, necessary, or weak reason for B given $w_{<B,\neq A}$ w.r.t. κ.

Here, $w_{<B,\neq A}$ denotes, as defined in section 2, the past of B except A, which collects, as argued, the obtaining circumstances.

As an illustration, look again at the case (a) of a causal chain and the case (b) of a conjunctive fork in section 3. These are easily distinguished with the help of ranking functions: suppose $\kappa(A) = \kappa(\overline{A}) = 0$, leaving us to specify only the ranks conditional on A and $\overline{A}$. One specification is this:

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

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

[26] 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.

[27] 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 which are intuitively causal intermediates. I return to this point in section 8.

A is here the only direct cause of B in w (where $A \cap B \cap C = \{w\}$), indeed a necessary and sufficient one, and B is the only direct cause of C in w, again, a necessary and sufficient one. The latter is due to the fact that A and C are independent given B as well as given $\overline{B}$; this is what probability theory refers to as the Markov property. So, we have here an example for a causal chain; indeed, the simplest one in which the ranks simply count how many times the obtaining causal relations are violated (in the sequence A, $\overline{B}$, and C, for instance, two such violations occur).

Another specification is this:

$\kappa(.\|A)$	C	$\overline{C}$
B	0	1
$\overline{B}$	1	2

$\kappa(.\|\overline{A})$	C	$\overline{C}$
B	2	1
$\overline{B}$	1	0

A is here the only direct cause of both B and C in w, indeed, a necessary and sufficient one in both cases, whereas B and C are now independent given A as well as given $\overline{A}$, or as we might also say, A screens off B from C. So, we now have the simplest example for a conjunctive fork in which the ranks again count the violations of the causal relations; the more violations, the more disbelieved.

One may wonder whether the relevant circumstances are now extremely embracive, much more embracive than intuition requires. The reason is that we have constructed „relevant“ extremely weak and hence, „relevant circumstances“ extremely strong. Indeed, for A and B in w, the whole of $w_{<B,\neq A}$ is relevant, but, as we might say, only possibly relevant on purely temporal grounds. The decisive advantage of this extremely strong construal is, however, that it is free of any circularity. This basis enables us to then search for less strict interpretations of „relevant circumstances“ which are hopefully provably equivalent to this construal.

The search is indeed successful. In Spohn (1990) sect. 4, I distinguish four wider senses of relevant circumstances that pertain to probabilistic causation, three of which are provably equivalent to the strongest sense. „Equivalent“ means here that the cause’s probability raising of the effect is exactly the same given the weaker circumstances as given the richest circumstances. These results carry over entirely to deterministic causation as explained here; but since they are a bit tedious, I refer the reader to that section in Spohn (1990).

I would like to add two observations, however. For one, it seems natural to conceive of the circumstances relevant for judging the causal relation between A and B in w simply as the set of small worlds in which that relation is as it is in w. These are what I have called the actually relevant circumstances in the widest sense in Spohn (1990), p. 130. The sad thing about this notion is that it might happen, for instance, that A is a reason for B given $w_{<A,\neq B}$ (and thus a cause of B), but not given the circumstances in the widest sense; thus, the widest sense is generally too wide.

Sufficient and necessary causation, however, is more well-behaved. We may define $S_s(A,B) = \{w \in W | A$ is a sufficient cause of B in $w\}$, and $S_n(A,B) = \{w \in W | A$ is a necessary cause of B in $w\}$. With the help of the law of disjunctive conditions (sect. 4), the following may be easily proved: first, if A is a sufficient or, respectively, necessary reason for B given $w_{<A,\neq B}$, then it is so given $S_s(A,B)$ or $S_n(A,B)$; second, it is so then given any proposition E about $U_{<B\neq A}$ with $E \subseteq S_s(A,B)$ or, respectively, $E \subseteq S_n(A,B)$, and third, $S_s(A,B)$ and $S_n(A,B)$ are the largest or weakest propositions for which the second assertion holds. Hence, if restricted to the case of sufficient or of necessary causation, the actually relevant circumstances in the widest sense could also be accepted as an equivalent definition of the obtaining circumstances, and they are indeed the weakest circumstances thus acceptable.

The other remark pertains to the construal of „relevant circumstances“ mentioned at the beginning of this section: why not conceive of the circumstances of the direct causal relation between A and B in w as the conjunction of all the other direct causes of B in w? In Spohn (1990), p. 131, I called this conjunction the circumstances in the ideal sense. This indicates that this understanding, though desirable, is often unattainable; in general, the ideal circumstances are not equivalent (in the above sense) to the circumstances taken in our strongest and basic sense. However, the equivalence holds under certain conditions: roughly that all the variables in the past of B which are individually independent of B given the rest of the actual past in B are also collectively independent of B given the rest of the actual past in B.[28] On the one hand, assuming ideal circumstances strengthens causal theory considerably[29]; on the other hand, this assumption is indeed quite strong. For our present context, the happy insight is that all these results carry over to deterministic causation, but the sad conjecture is that things don't get nicer or simpler by focussing on sufficient or necessary causation.

6. Causation

[28] For details, cf. Spohn (1990), p. 131.

[29] Cf. Spohn (1990), theorems 12, 15, and 16.

So far, we have dealt only with direct causation. How do we now extend our account to causation in general? On the one hand, we have clear structural intuitions which may quickly settle the question; on the other hand, there are further strong intuitions which cast doubt on the quick solution. Hence, we have to consider the case a bit more thoroughly.

The structural intuition is, of course, that causation is transitive. It goes without saying that direct causes are causes. And, clearly, causal relations should not further extend 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. This is what our discussion will end up with. Therefore, let us fix it now:

Definition 7: A is a *cause* of B in w relative to κ (or β) if and only if there are $A_1,..., A_n$ such that $A_1 = A$, $A_n = B$, and, for all $i = 1, ..., n$-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. Intiutively, we speak of surprising or important causes, of the crucial or most information 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.[30]

Having said this, what, then, are the real doubts about definition 7? They consist in two theoretical concerns. One has to do with positive relevance, and the other with Markovian intuitions, as one might say. Let's discuss them in turn.

The first concern is due to our most basic intuition about causation. A cause is, we said, positively relevant to its effects given the obtaining circumstances. We then managed to implement the idea for direct causation in a circle-free way. This suggests that we should try to implement this idea in full generality. However, this creates a conflict with transitivity, as causal preemption immediately demonstrates.

Consider the famous case of the traveller who attempts to cross the desert. Two enemies, however, independently seek his death. One puts poison into the traveller's water keg, and the other, not knowing of the first, later drills a hole into the keg. The traveller

---

[30] My critical reference point here is the pragmatic theory of explanation of van Fraassen (1980), ch. 5, which is insightful in its pragmatic aspects, but surprisingly uncommitting in its systematic import which hides in what van Fraassen calls the relevance relation.

does not check his equipment, and so the poor guy dies of thirst after a few days in the desert. Clearly, the drilling of the hole is positively relevant, indeed necessary and sufficient, under the obtaining circumstance for the traveller's not drinking anything for days. The lack of fluid is positively relevant, indeed necessary and sufficient, as we may imagine again, for the traveller's death. So, given transitivity, the drilling of the hole is a cause for the poor guy's death. But it is not positively relevant! If the second enemy had remained passive, the poison would have done its duty, and the story would have ended with the traveller's death with equal certainty (or with greater certainty, we might even assume, thus turning the hole drilling into a (weak) reason *against* the death). It is easy to distribute (conditional) ranks in a plausible way such that these assertions also come out true in a formal way. All this shows that the incompatibility between transitivity and positive relevance turns up not only with probabilistic causation or with such unusual constructs as additional or weak causes, but also in the most central case of necessary and sufficient causation.

At least in the example, it is intuitively clear how the conflict should be resolved: the hole in the keg is an indirect cause of the traveller's death; the second enemy has killed the traveller and can certainly not plead that the traveller would have died anyway (or that he has possibly given him a better chance). This is precisely the reason why Lewis (1973b), for whom the example presents the same problem, equates causal dependence, not with counterfactul dependence, but with the transitive closure thereof.

Before settling for the same conclusion, we should look at the second concern, however. This is raised by the prominent probabilistic idea that causal chains should be Markov chains. This idea has its force also within the deterministic setting. In a causal chain, the present member renders the past members irrelevant for the future members; once the chain has reached its present stage, how it got there does not matter for how it goes on. This Markovian intuition runs, however, against transitivity.A Markow chain from A to B and another from B to C do not necessarily combine to one large Markow chain from A to C. Moreover, the Markovian intuition apparently fits well to the positive relevance idea. It is well known[31] that if each member in a Markov chain is probabilistically positively relevant to the subsequent one, then the first is also positively relevant to the last; i.e. positive relevance spreads through Markov chains. Of course, this result carries over to ranking functions. This considerably strengthens the case against transitivity.

However, we should not generally expect causal chains to be Markov chains. Think of the following example (fighting the morbid tendency of causal theorists): at the signal

[31] See, e.g. Spohn (1990), theorem 10.

of the romantic lover (*A*) a fiddler (*B*) and a mandolin player (*C*) strike up a sweet melody in order to tenderly wake the beloved (*D*). Here we have two causal chains running from *A* to *D*, one through *B* and one through *C*. Therefore, none of them is a Markov chain; *B* does not render *A* irrelevant for *D* because of the causal chain through *C*, and vice versa. So, whenever two events are multiply connected (certainly not an uncommon situation), each of the connections cannot form a Markov chain.The Markovian intuition as we have explained it above applies only to causal chains singly connecting past and future.

There are ways to generalize the Markovian intuition to such complex situations, either by introducing the more general notion of a Markov field or, preferably, by embedding Markov chains into the obtaining circumstances.[32] However, transitivity is not regained thereby, and the spreading of positive relevance within the more general structures is lost. All these considerations also pertain to the deterministic case.

So, the situation seems to be confusing. However, in our example of the romantic lover, the structural intuition of transitivity was the overriding one. This would remain even if we looked in more detail to examples in which also the more refined Markovian notions violate transitivity. So, intuition favors transitivity, as it already has for Lewis (1973b).

However, I am not wholly satisfied by this appeal to intuition. I am more attracted by the following theoretical argument: whenever there are several plausible explications of some notion we are interested in, the theoretically most enlightening and most reasonable procedure is to look for the weakest of these explications; only this opens the way to theorize about conditions under which the stronger explications apply as well.

Concerning causation, it is the transitive closure of direct causation which obviously yields the weakest or widest permissible causal relation; causation cannot plausibly extend further (within discrete time). If, however, we would give priority to the preservation of positive relevance within causal chains or to Markovian properties of causal chains, this would result in stricter or narrower causal relations permitting only shorter causal chains. Thus, the theoretical maxim just stated also favors the above definition 7.

Satisfaction of the maxim further demands, then, investigating the conditions under which causal chains thus defined have the desired stronger properties. Section 6 of Spohn (1990) contains such an investigation in probabilistic terms. The results obtained there fully carry over to the deterministic case; but as far as I can see, they do not

---

[32] For both alternatives see Spohn (1990), p. 136.

essentially simplify. These results may or may not satisfy our expectations (I think they do), but they are open to improvement.

7. A Brief Comparison with the Counterfactual Theory

We have found reason for rejecting various versions of the regularity theory. A complete comparison with rival theories of deterministic causation is definitely beyond the scope of this paper. The counterfactual theory of Lewis (1973b, 1986a), though, is an immediate neighbor of the present one; thus, a comparison with it is fitting as well as illuminating.

A first point, which I am prone to emphasize, is that my account allows for much more explicit theorizing. This should have been sufficiently demonstrated in the previous sections, and it is supported by the observation that, to repeat, the whole theory of Bayesian nets can be stated in terms of ranking functions. It is not obvious how the theory of counterfactuals may reach comparable achievements. The reason is that little goes into the explicit theory of counterfactuals; namely, just the standard logic VC of Lewis (1973a), say, or its corresponding similarity semantics. A lot is left for intuitive explanation, as contained, for instance, in the exclusion of a back-bracking interpretation of counterfactuals and as more fully given in Lewis (1979). In particular, the nomological aspect of the theory of causation, or what goes in its place, and the reference to the obtaining circumstances are left unseparated in counterfactuals. All this hampers explicit theorizing.

This argument is certainly not decisive. The counterfactual theory may be developed in an appropriate way (I would like to see it done), and the formal virtues of theories of causation are definitely secondary compared with their adequacy. So, let us look at the latter.

We have seen that regularity theories already fail at the simple distinction between causes and symptoms. This is not a problem for Lewis' counterfactual theory or mine. The problem of preemption is a bit harder, but I find the solution of Lewis (1973b) convincing, and have essentially copied it in my account.

In his postscripts (1986a), pp. 200ff., however, Lewis distinguishes early and late preemption. Early preemption is the normal one already discussed. Late preemption, by contrast, can take three forms, the first two of which are classified as far-fetched and the third of which is taken seriously by Lewis. The forms fare differently within my account. The first case of late preemption (p. 202) is easily represented with the help of ranking functions. As soon as the preempting or the preempted cause is indirect, my theory can account for them. Only if both are supposed to be direct, it fails (like Lewis'); but this is an entirely incomprehensible supposition. The second case of late preemption involves an infinity of preempted causal chains; it does not seem impossible to deal with it in an appropriate extension of my theory. The third case, finally, obtains when the preempted causal chain, which is not completed, is cut off only by the effect itself in which the preempting causal chain results. This sounds mysterious, and it really is, I find. The problem is created by Lewis' ontology of events which tries to keep the balance between a too fragile and a too coarse-grained individuation of events. I would take alleged cases of such preemptions as a sign that the events considered are too coarse-grained; in such cases, the examples themselves dictate the standards of fragility. In any case, there is no way at all to represent the third form of late preemption within my account. Nor is there in Lewis. Therefore he introduces an extended counterfactual analysis (pp. 206f.), instead of tinkering with the individuation of events. I do not think that any strong argument is coming to the fore here.

The really discriminating case is that of (symmetric) causal overdetermination. It is clear that the counterfactual theory cannot cope with it. If C is causally overdetermined by A and B, C would have been the case even if A would not have been the case; thus, A is not a direct cause of C according to the counterfactual theory. The same holds for B. And, unlike preemption, the present case cannot be resolved by resorting to indirect causation; if the chain from A to C would have been broken, the chain from A to C would still have completed, and thus C still have occurred. In footnote 12 of (1973b), Lewis claims to lack firm naive opinions about such situations, and is thus happy to dismiss them as test cases. In his postscripts (1986a), pp. 207ff., he is more opinionated and again happy to be able to agree with Bunzl (1979) that essentially all alleged cases of causal overdetermination reduce either to ordinary joint causation or to preemption. If there should be some irreducible cases left, they might be dealt with by his extended counterfactual analysis.

Either way, causal overdetermination requires a great deal of energy on behalf of the counterfactual theory. This is in strange disharmony, I find, to the great ease with which at least prima facie cases of overdetermination can be produced; they indeed abound in everyday life. That the beloved awakes is causally overdetermined by the two musicians; each of the musicians alone would also have caused her to wake up, though not precisely in the way she did. The firing squad is an often cited more sinister example. Lewis uses neural networks for illustrating various causal schemes. My preferred examples are the huge arrangements of millions of dominos, where the first is pushed, and then one domino tips the other until all have fallen over after a few hours. There you really *see* how causal chains branch off, flow together, join forces, get preempted, etc., in an artfully designed way. Of course, there is also overdetermination; two falling dominos tip a third at the same time, and each of them would have sufficed to overthrow it.

Certainly, my intuitions are shaped by my theory as well. Still, I find it very desirable to have a simple account of this simple phenomenon and not be required to explain it (away) in some complicated or artificial fashion. And my point is that ranking functions give such an account in a straightforward way. Definition 5b in section 4 did well to allow for additional reasons. Definition 6 in section 5 similarly allowed for additional causes, and overdetermining causes are additional causes precisely in this sense.

Let us look at some numerical examples in which both A and B are directly causally relevant to C in various ways. In each of the following examples, the upper left number specifies the degree of belief $\beta(C \mid A \cap B)$ in C conditional on $A \cap B$ according to the given belief function β, etc.:

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

(a) joint necessary and sufficient causes

$\beta(C \mid .)$	B	$\overline{B}$
A	1	0
$\overline{A}$	0	-1

(b) joint sufficient, but not necessary causes

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

(c) overdetermining causes

Case (a) is the ordinary case of joint causation; two causes go together to produce an effect. If either one (or both) had been missing, the effect would not have come about.

In case (b), each of A and B is a sufficient, but not a necessary, cause of C in the presence of the other; in the absence of the other, it would have only been a necessary, but not sufficient, cause. This still seems to me to be a case of joint causation; A and B go together to produce C.

In case (c), by contrast, 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 it is still positively relevant to, i.e. a cause of C, but then it can only be an additional cause. This scheme, I find, fits naturally to 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 (a) and (b). If in a case of overdetermination the effect does not occur, for some or no reason, at least two things have gone wrong at once. This is at least doubly unbelievable, as represented in (c).

The reason why this simple account is 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 scheme (c) is available to counterfactual theories.

This shows that there is a price to pay for my simple account, and we shall see in the next and final section that it has its problems, too; after all, I did not want to deny that there is something puzzling about causal overdetermination. Still, I count the availability of this simple account as a positive advantage over counterfactual theories.

8. Objectivization

Let me finally discuss the price we paid, not only in the previous section, but all along; the price of subjectively relativizing causation to an observer or epistemic subject. I hope I have shown that we did not get little for that price; indeed, things unattainable to others. But the price is high, and some may think too high. So we should take efforts to get it back, or at least part of it; that is, to reestablish objective causation on the subjective basis presented above.

Before doing so, however, I would like to finish my comparison with counterfactual theories by stating that I also find some value in paying such an articulate price. The above explications have decidedly moved to the subjectivistic side, and it will be a well-defined question how to regain objectivity. By contrast, the stance of the counterfactual theory towards the objectivity issue is wanting, I find. The official doctrine is, of course, that the counterfactual theory offers an objective account of causation; this definitely counts in its favor. However, this objectivity gets absorbed in the notion of similarity on which Lewis' semantics for counterfactuals is based. There are some relatively clear-cut rules for that similarity as specified in Lewis (1979) (though these rules show at the same time that „similarity" may be a misleading term in this context); but they leave a large range of indeterminateness. Rightly so, Lewis would say, and I would agree. Still, I wonder whether such similarity judgments are significantly better off in the end than, say, judgments about beauty, and hence whether the semantics for counterfactuals should not rather take an expressivist form like the semantics of „beautiful". The question is difficult to decide, and I do not want to decide it here. The only point I want to make is that the whole issue is clouded behind the objectivistic veil of the counterfactual theory. It is clearer, I find, to jump right into subjectivity and, given all of it (that's the presupposition of the question), to ask how much of it is necessary.

So, how much of it *is* necessary? The objectivization of the above account has two aspects. The first is to get rid of its frame-relativity. This can be done by appealing to the universal frame consisting of all variables whatsoever. But this appeal is doubtlessly obscure. The somewhat homelier method is to assert that the causal relations obtaining within a given small frame will be maintained in all (but finitely many) 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.[33]

The main aspect of objectivization, however, pertains to the ranking functions. In my account, they played a role corresponding to that of laws or regularities in the regularity theory of causation, and we have seen that they play their role much more successfully. However, their only clear interpretation is as subjective doxastic states. Is there a way to gain a more objective view of them?

Of some of them, yes.[34] 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 $\overline{A}$ and $w_{<B,\neq A}$, then $\overline{B}$, 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, though in a purely truth-functional, non-modal form. Hence, L is simply a regularity with entirely unproblematic and objective truth conditions.

The next question is whether ranking functions can be reconstructed from their associated causal laws. To the extent in which this is feasible, we may rightfully confer the objectivity of the causal laws to the ranking functions reconstructed from them.

The reconstructibility is clearly limited; 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. We may call them *fault counting functions*: for a given 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 all the figures 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: (1) a certain

---

[33] Cf. also Spohn (2000).

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

principle of causality has to hold, and (2) each direct cause must immediately precede its direct effect.[35] These conditions certainly invite further scrutiny and evaluation. Here, I will confine myself to three remarks:

First, it would have been very 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 frames 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 (2) via the objectivizability of causal relations; it is objectivization which requires frames to be rich enough to always provide immediately preceding direct causes.

The other remark is that additional causes (and weak causes as well) 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 should only contain the material implication „if $w_{<B,\neq A}$, then B“, since this is what is believed in κ in any case. But if this is what is contained in the law, we cannot read off from the law whether A is positively or negatively relevant or irrelevant to B given $w_{<B,\neq A}$. In view of how I have accounted for causal overdetermination, this also means that such overdetermination cannot be objectivized. Thus, there is trouble with overdetermination according to my account, too. The difference is that I have both a simple account of overdetermination as well as an explanation of our strong, and often successful, tendency (if Bunzl 1979 and Lewis 1986a are right) to explain it away.

The final remark is that all this entails a certain view of causal laws. If we look only at the result of objectivization, then causal laws are mere regularities, and my account becomes indistinguishable from the regularity theory. But the objectivizing theory behind that result also gives an account of the modal force of causal laws. It does not simply postulate this modal force, as does Armstrong (1983), who thus provokes bewilderment as to how to distinguish presence from absence of the modal force. It rather gives a Humean explanation of that modal force via the briefly sketched theory of objectivization; namely, by uniquely associating with a causal law a characteristic inductive behavior which is encoded in the corresponding ranking function and which interprets certain material implications contained in the law as causal relations. All this, though, would have been out of reach without a general and adequate theory of inductive

[35] For (1) cf. Spohn (1993), p. 243, and for (2) cf. p. 246. If, being more liberal than in this paper, we admit there to be several simultaneous variables, a third condition is required; cf. p. 244.

behavior or of doxastic dynamics accounting for the notion of plain belief, as I have presented it in section 4.

Bibliography

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

Cartwright, N.: Causal Laws and Effective Strategies. In: *Noûs* 13, 1979. pp. 419-437; also in N. Cartwright, *How the Laws of Physics Lie*, Oxford 1983, pp. 21-43.

Beauchamp, T., A. Rosenberg.: *Hume and the Problem of Causation*, Oxford 1981.

Bunzl, M.: Causal Overdetermination. In: *Journal of* Philosophy 76, 1979, pp. 134-150.

Davidson, D.: Reply to Quine on Events. In: E. LePore, B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford 1985, pp. 172-176.

Graßhoff, G., May, M.: Causal Regularities, this volume, (= 2000).

Hansson, S.O.: Revision of Belief Sets and Belief Bases. In: D.M. Gabbay, P. Smets (eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 3, Belief Change*, Dordrecht 1998, pp. 17-75.

Jeffrey, R.C.: *The Logic of Decision*, Chicago 1965, ^{2}1983.

Jensen, F.V.: *An Introduction to Bayesian Networks*, London 1996.

Kim, J.: Causation, Nomic Subsumption, and the Concept of an Event. In: *Journal of Philosophy* 70, 1973, pp. 217-236.

Lewis, D.: *Counterfactuals*, Oxford 1973a.

Lewis, D.: Causation. In: *Journal of Philosophy* 70, 1973b, pp. 556-557; also in Lewis (1986c), pp. 159-172.

Lewis, D.: Counterfactual Dependence and Time's Arrow. In: *Noûs* 13, 1979, pp. 455-476, with post-scripts also in Lewis (1986c), pp. 32-66.

Lewis, D.: Postscripts to „Causation". In: Lewis (1986c), pp. 172-213, (= 1986a).

Lewis, D.: Events. In: Lewis (1986c), pp. 241-269, (= 1986b).

Lewis, D.: *Philosophical Papers, vol. II*, Oxford 1986c.

Mackie, J.L.: Causes and Conditions. In: *American Philosophical Quarterly* 2, 1965, pp. 245-264.

Mackie, J.L.: *The Cement of Universe*, Oxford 1974.

Merin, A.: *Die Relevanz der Relevanz.* Habilitationsschrift, Stuttgart 1996. Engl. translation to appear.

Niiniluoto, I.: Inductive Systematization: Definition and Critical Survey. In: *Synthese* 25, 1972, pp. 25-81.

Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, Ca., 1988.

Salmon, W.C.: Statistical Explanation. In: R.G. Colodny (ed.), *The Nature and Function of Scientific Theories*, Pittsburgh 1970, pp. 173-231.

Skyrms, B.: *The Dynamics of Rational Deliberation*, Cambridge, Mass., 1990.

Spohn, W.: Deterministic and Probabilistic Reasons and Causes. In: C.G. Hempel, H. Putnam, W.K. Essler (eds.), *Methodology, Epistemology, and Philosophy of Science. Essays in Honour of Wolfgang Stegmüller on the Occasion of his 60th Birthday*, Reidel 1983a, pp. 371-396.

Spohn, W.: *Eine Theorie der Kausalität*, Habilitationsschrift, München 1983b.

Spohn, W.: Ordinal Conditional Functions. A Dynamic Theory of Epistemic States. In: W.L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, vol. II, Dordrecht 1988, pp. 105-134.

Spohn, W.: Direct and Indirect Causes. In: *Topoi* 9, 1990, pp. 125-145.

Spohn, W.: A Reason for Explanation: Explanations Provide Stable Reasons. In: W. Spohn, B.C. van Fraassen, B. Skyrms (eds.), *Existence and Explanation*, Dordrecht 1991, pp. 165-196.

Spohn, W.: Causal Laws are Objectifications of Inductive Schemes. In: J. Dubucs (eds.), *Philosophy of Probability*, Dordrecht 1993, pp. 223-252.

Spohn, W.: On the Properties of Conditional Independence. In: P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher. Vol. 1: Probability and Probabilistic Causality*, Dordrecht 1994, pp. 173-194.

Spohn, W.: Ranking Functions, AGM Style. In: B. Hansson, S. Halldén, N.-E. Sahlin, W. Rabinowicz (eds.), *Internet Festschrift for Peter Gärdenfors*, Lund 1999, http://www.lucs.lu.se/spinning/

Spohn, W.: Bayesian Nets Are All There Is To Causal Dependence. To appear in: M.C. Galavotti et al., *Stochastic Dependence and Causality*, Stanford 2000.

Suppes, P.: *A Probabilistic Theory of Causality*, Amsterdam 1970.

van Fraassen, B.C.: *The Scientific Image*, Oxford 1980.