

1. Introduction

Everybody agrees that the distinction between direct and indirect causation is important. And it seems easy to draw, *if* an analysis of causation in general is available. The causal influence of one event on another is direct, if it is not mediated by other events in between; otherwise it is indirect. The trouble is with the proviso. Indeed, I contend that the order of analysis must be reversed because the distinction is required for a successful analysis of causation. Such an analysis perhaps proceeds best in two steps: the first analyses direct causation, and the second extends the analysis to indirect causation and thus to causation in general. Such a strategy is at least plausible. For direct causation is a very special case and so may be supposed to be more easily explicable. Then, one might say that the relation “*A* is a cause of *B*” is just the transitive closure of the relation “*A* is a direct cause of *B*”, thus completing the full analysis of causation.¹ The complete story is not so simple; but the idea will turn out to be right. In (1983a), I dealt mainly with the first step — direct causation. Here, I deal mainly with indirect causation.

Section 2 introduces the conceptual machinery required throughout the paper. Section 3 recapitulates my 1983 explication of direct causes. Section 4 considers the circumstances of direct causal relations in greater detail. Section 5 presents the main difficulties with indirect causes. Section 6, finally, proposes a strategy for dealing with these difficulties and shows that it will work.

The whole enterprise is subject to two major constraints. First, I shall discuss only causation of single events. The hope, of course, is that it will emerge from knowledge of causation in the single case what causal laws are. This procedure seems to me to be more likely to succeed than the reverse strategy endorsed by several philosophers,² though I shall not argue the point here.

Secondly, I am concerned here only with probabilistic causation because this is the context in which all the

problems dealt with here have been raised and have been discussed most extensively. It should be mentioned, however, that each consideration, definition, and theorem of the present paper can routinely be extended to deterministic causation with the help of the theory of ordinal conditional functions (OCFs) I proposed in (1988); only some marginal adjustments may be needed. This follows because deterministic conditional independence defined for OCFs obeys essentially the same laws as probabilistic conditional independence.³ It is thus possible to unify the deterministic and the probabilistic approach.⁴

2. The conceptual and formal framework

Each discussion of probabilistic causation proceeds from an explicitly given probability space: let I be a non-empty set of variables or factors. Each variable $i \in I$ is associated with a set Ω_i of at least two possible values i may take. The cross product Ω of all the Ω_i is the set of all functions ω defined on I such that, for each $i \in I$, $\omega_i \in \Omega_i$; intuitively, each ω represents a possible course of events — a possible world in philosophers’ talk, or a possible path in the mathematician’s terminology. I , each Ω_i , and hence Ω are assumed to be finite. This severe restriction has several advantages. One of these is that there is no need to worry about measurability because each subset of Ω may be assumed to represent a state of affairs or an event in the mathematicians’ sense, but not the philosophers’. Moreover, we assume a probability measure P assigning a probability to each state of affairs, i.e. to each subset of Ω . This completes the description of the underlying probability space.

This explicitness has an important philosophical consequence: namely, that everything said about causation is relative to the descriptive frame given by the set I of variables. Many discussions of examples suffer, I think, from an inadequate recognition of this relativiza-

tion. It is essential because the causal relations may indeed vary with the frame. Consider, e.g., a series of throws of a die by a machine: relative to a coarse probabilistic description which contains only variables representing the throws, no throw will be causally relevant to the next one. Relative to a finer description, however, which, for each time, allows for variables representing the mechanical state of the whole system (but which may still be probabilistic, say, because of neglect of air resistance), each throw will be causally relevant to the subsequent ones. One is, perhaps, inclined to think of causation as an absolute notion. However, from the current starting point the only way to get rid of the relativization is via the most fine-grained descriptive frame embracing *all* variables whatsoever. I am not sure whether such a move makes sense; it is at least philosophically problematic. Here, I will be content with the relativized notion of causation.

The relativization of causes is even more apparent in the distinction between direct and indirect causation. A state of affairs which is a direct cause relative to a coarse descriptive frame not mentioning the mediating links may well turn out to be an indirect cause relative to a more complete descriptive frame.

If time is continuous and if variables are associated with points and not with intervals of time, then, presumably, direct causes either do not exist or are simultaneous with their direct effects. In either case, the strategy of explicating causation via direct causation would not work because, in either case, causation would certainly not be the transitive closure of direct causation. So the strategy of analysis here demands a descriptive frame with discrete time. The idea is that the results obtained for discrete time may be generalized to continuous time in a fashion similar to the way in which the theory of stochastic processes has been extended, and the hope is that this will raise only well-known mathematical, but not new conceptual or philosophical problems. I shall not attempt here, however, any such generalization.

I assume a weak order⁵ \leq on the set I of variables which represents the order of the times at which the variables are realized; $<$ is to denote the corresponding irreflexive order relation. Since I is finite, time is bound to be discrete. By assuming the order to be weak, simultaneous variables are in general allowed; the few exceptions will be explicitly noted. However, I shall not consider simultaneous causation; I am not sure whether this would be desirable.⁶ And I plainly exclude back-

wards causation; it will be clear that this is vital to the theory to be proposed here.

An analysis of causation faces a number of well-known and unsolved problems relating to variables which have more than two possible values.⁷ One may evade these problems by considering only binary variables. But there is a hitch to this restriction. The causal theorems to be proved essentially derive from the laws of conditional probabilistic independence, and there is one such law peculiar to binary variables (see Theorem 2(f) below) which may have unforeseen and undesired consequences.⁸ Therefore, variables will be assumed to be binary only when required, and the problems with variables with more than two values will be neglected.

Finally, I shall assume that the probability measure P is strictly positive, i.e. that $P(\{\omega\}) > 0$ for all $\omega \in \Omega$; hence, the conditional probability $P(B|A)$ is defined for each $A \neq \emptyset$. Since Ω is finite, this assumption is unproblematic. The reason for it is that all probabilistic theories of causation run into serious trouble with the limiting probabilities 0 and 1.⁹

How are probabilities to be understood in the present context? Any way you like. For instance, if one takes probability objectively, preferably in a propensity interpretation, then the definitions below attempt to explicate causation as it objectively is. If, however, probabilities are understood epistemically as those of a certain subject at a certain time, then these definitions account for the causal conception of that subject at that time.

For philosophical reasons, I prefer the second understanding of probability. There are two main reasons. First, objective probability is the much more problematic notion, and it seems to be heavily intertwined with causality.¹⁰ The most promising attempt to understand it is, I think, via subjective probability.¹¹ This suggests to me that the appropriate order is to start with subjective probability, to explicate causation within the subjectivistic framework, and then to try to objectivize both.

Secondly, I have general reservations about too realistic an understanding of causation. There is a need for explaining the most pervasive and prominent epistemological role which the notion of causation plays. If one takes causation simply as a constituent of the real world, then the only explanation one can give seems to be this: "Causation is, of course, a fundamental and pervasive trait of reality; thus it is small wonder that

the notion of causation plays a fundamental and pervasive role in our picture of reality". However, the same argument would hold, say, for quarks or electromagnetic forces. Thus, this kind of explanation assimilates the epistemological role of the notion of causation to that of our notions of other important things like quarks or electromagnetic forces. This seems to me to be a distortion; according to the views of Hume, Kant, and other philosophers,¹² the notion of causation has not only an important, but a peculiar epistemological role which cannot be sufficiently explained from a realistic point of view. However, this essay is deliberately neutral with respect to these deep and crucial philosophical issues. Its focus is on the *logic* of causation, and it is intended to inform the philosophy of causation.

The following notation will be used throughout: variables, i.e. elements of I , will be denoted by i, j, k , and l , subsets of I by J, K, L, M , and N (with or without subscripts). (i, j) refers to the open interval between i and j , i.e. to $\{k \in I \mid i < k < j\}$, and $[i, j]$ to the closed interval $\{k \in I \mid i \leq k \leq j\}$; $\langle j \rangle$ denotes the past of j , i.e. $\{k \in I \mid k < j\}$, and $\langle j - K \rangle$ the past of j except K , i.e. $\langle j \rangle - K$.¹³ Instead of $\langle j - \{i_1, \dots, i_n\} \rangle$ we simply write $\langle j - i_1, \dots, i_n \rangle$.

Possible paths, i.e. elements of Ω , will be denoted by v and ω , states of affairs, i.e. subsets of Ω , by A, B, C, D , and E . We often have to refer to partial paths or, rather, to the set of their completions, which are states of affairs: for each $\omega \in \Omega$ and $J \subseteq I$ we define ${}^\omega J = \{v \in \Omega \mid v(i) = \omega(i) \text{ for all } i \in J\}$,¹⁴ and I write ${}^\omega i$ instead of ${}^\omega \{i\}$. In general, states of affairs which are concerned only with variables in some set J are called J -measurable states or simply J -states; mathematicians also call them J -cylinders. The formal definition is that A is a J -state iff, for all v and ω agreeing on J , $v \in A$ iff $\omega \in A$. Thus, A is a J -state iff $A = \cup \{{}^\omega J \mid \omega \in A\}$; and in particular each ${}^\omega J$ is a J -state.

The laws of conditional probabilistic independence lie at the bottom of the whole inquiry and therefore need at least to be stated.

DEFINITION 1. The states of affairs A and B are *independent conditional on* C , i.e. $A \perp B/C$, iff $P(A \cap B \mid C) = P(A \mid C) P(B \mid C)$. And the sets K and L of variables are *independent conditional on* the set M of variables, i.e. $K \perp L/M$, iff, for all K -states D , L -states E , and $\omega \in \Omega$, $D \perp E/{}^\omega M$. I shall often mix the two notations, i.e., more precisely: $K, A \perp L, B/M, C$ is to

mean that, for all K -states D , L -states E , and $\omega \in \Omega$, $A \cap D \perp B \cap E/C \cap {}^\omega M$.

The independence of states of affairs obeys

THEOREM 1.

- (a) If $A \perp B/C$, then $B \perp A/C$;
- (b) if $P(C) \neq 0$ and $C \subseteq A$, then $A \perp B/C$;
- (c) if A and A' are disjoint and $A \perp B/C$, then $A \cup A' \perp B/C$ iff $A' \perp B/C$;
- (d) if $A \perp C/D$, then $A \perp B \cap C/D$ iff $A \perp B/C \cap D$.

The independence of sets of variables obeys

THEOREM 2.

- (a) If $K \perp L/M$, then $L \perp K/M$;
- (b) if $K \subseteq M$, then $K \perp L/M$;
- (c) if $K' \subseteq K \cup M$, $L' \subseteq L \cup M$, $M \subseteq M' \subseteq K \cup L \cup M$, and $K \perp L/M$, then $K' \perp L'/M'$;
- (d) if $J \perp K/L \cup M$ and $J \perp L/M$, then $J \perp K \cup L/M$;
- (e) if K and L are disjoint, $J \perp K/L \cup M$, and $J \perp L/K \cup M$, then $J \perp K \cup L/M$ — provided P is strictly positive;
- (f) if i is a binary variable, $K \perp L/M$, and $K \perp L/M \cup \{i\}$, then $K \cup \{i\} \perp L/M$ or $K \perp L \cup \{i\}/M$.

For proofs see, e.g., Dawid (1979) or Spohn (1980). In particular Theorem 2(e) will be important; this is a further reason for assuming a strictly positive probability measure.¹⁵ This list of properties of conditional independence is not complete,¹⁶ but Geiger, Pearl (1988) present a number of interesting partial completeness results.

Concerning causal notation, three things must be observed. First, the causal relata are always states of affairs which are states of a single variable and thus are, so to speak, logically simple; I do not see the need to consider logically complex states of affairs as causes or effects.¹⁷ Second, whether A is a cause of B depends, of course, on the given world or path; there may well be two worlds such that A causes B only in one world, but not in the other. This path-relativity will be made explicit in the notation. Third, only facts can be causes or effects; A can cause B in ω only if A and B obtain in ω , i.e. if $\omega \in A \cap B$.

$A \xrightarrow{\omega} B$ is to mean that A is a direct cause of B in ω ; and $A \xrightarrow{\omega} B$ is to mean that A is a (direct or indirect) cause of B in ω . This notation, and all the notation to follow, always carries the presupposition

that $\omega \in A \cap B$ and that there are variables i and j such that A is an i -state, B is a j -state, and $i < j$. A similar notation for counter-causation, causal relevance and irrelevance, etc. will be introduced later on.

3. Direct causes

A is a cause of B iff A precedes B and raises the epistemic or metaphysical rank of B under the obtaining circumstances. This is the basic conception of causation which has found the widest agreement. In the deterministic case, it covers regularity theories and counterfactual approaches as well as analyses in terms of necessary and/or sufficient conditions which all differ on the relevant meaning of "raises the epistemic or metaphysical rank". My proposal is still different, namely to explicate this phrase in terms of OCFs. In the probabilistic case, however, there is only one interpretation of this phrase, that is, that A raises the probability of B .

The phrase "the obtaining circumstances" is also unclear; it is, in a sense, the subject of the whole paper. For direct causation, however, there is a particularly simple definition. This indeed is the main reason for splitting the account of causation into two steps. As I have argued in (1980), pp. 79ff., (1983a), pp. 384ff., and (1983b), pp. 80ff., each fact preceding the direct effect B and differing from the direct cause A is to count among the obtaining circumstances of the direct causal relation between A and B ; whenever judgment about that relation is based on less, it may be just the neglected facts which would change the judgment. This means that the obtaining circumstances consist of the whole past of B with the exception of A . I shall turn this into a formal definition and briefly compare it with other proposals.

DEFINITION 2. Let A be an i -state, B a j -state, $i < j$, and $\omega \in A \cap B$. Then, A is a direct cause of B in ω , i.e. $A \xrightarrow{\omega} B$, iff $P(B|A \cap \omega\{<j-i\}) > P(B|\bar{A} \cap \omega\{<j-i\})$.¹⁸ A is a direct counter-cause of B in ω , i.e. $A \xrightarrow{\omega} \leftarrow B$, iff $P(B|A \cap \omega\{<j-i\}) < P(B|\bar{A} \cap \omega\{<j-i\})$. A is directly causally relevant to B in ω , i.e. $A \xrightarrow{\omega} \diamond B$, iff $A \xrightarrow{\omega} B$ or $A \xrightarrow{\omega} \leftarrow B$. Finally, A is directly causally irrelevant to B in ω , i.e. $A \xrightarrow{\omega} \dashv B$, iff not $A \xrightarrow{\omega} \diamond B$.

In a way, Definition 2 proposes a radical solution to Simpson's troublesome paradox. If one conditionalizes

on the whole past of the effect, there is no further subdivision of that past which could change the conditional probabilities. Of course, this is true only relative to a fixed descriptive frame; but this only emphasizes the importance of relativization.

Suppes (1970), pp. 41f., moves from his own definition of *prima facie* causes toward Definition 2 by acknowledging the legitimacy and usefulness of relativizing his definitions to some background information. However, he does not expand on this suggestion. One may think that one need not mention the background as long as it is constant. But according to Definition 2, different direct causal relations refer to different backgrounds, to different obtaining circumstances. So it is mandatory to make the reference explicit.

When discussing Simpson's paradox, Suppes (1984) doubts that the problem posed by it is solvable absolutely. He says, for example, that "there is no end to the analysis of data in a practical sense" (p. 56). I agree. But surely there is a natural end to the analysis of data within a fixed descriptive frame, a point with which Suppes, in turn, seems to agree (p. 57). This is captured by Definition 2.

Good's theory (1961) differs from Definition 2 in several ways, but the crucial point is that in defining the tendency of A to cause B Good considers other conditional probabilities. He conditionalizes on the whole past of the cause and on all true laws of nature,¹⁹ whereas I conditionalize on the whole past of the direct effect. I have not found a clear argument for the appeal to the true laws.²⁰ Indeed, as far as I am concerned, it spoils much of the philosophical interest of the whole enterprise; for, the hope is that a better grasp of laws of nature will emerge from the analysis of singular causation.

The main question is whether to conditionalize on the past of the cause or on the past of the effect. Definition 2 would obviously be inadequate as a general account of causation, and one might therefore favor Good's account. My theoretical reasons for not doing so will emerge later on. The basic objection, however, is provided by a simple example:

Take a two-person game; each of the players makes a choice, and the outcome is determined accordingly. So the outcome is caused by both of the choices, but these choices, let us suppose, are causally independent of each other. Indeed, we may assume that their temporal order is totally irrelevant to the whole set-up.²¹ What is the causal efficacy of the choice of the first player to the outcome? On Good's account, it varies

with the temporal order: if the second player makes his choice first, that choice must be conditionalized on, otherwise not. This seems unacceptable because the causal set-up is not really changed by changing the temporal order. On the other hand, if the choices are taken as direct causes of the outcome, Definition 2 judges the causal efficacy of one choice by considering the probabilities conditional on the other *irrespective* of their temporal order.

One might point out that there is no difference between Good's conditionalization proposal and mine if direct causes immediately precede their direct effects, i.e. if there are no temporally intermediate variables in the given descriptive frame. And one might think that it would indeed be reasonable to assume that direct causes immediately precede their direct effects.²² At the present stage, however, this assumption is quite unreasonable. It is a strong assumption which implies that, in case I is linearly ordered by \leq , the given probability space is so well-behaved as to form a Markov chain. But attention should not be confined just to Markov processes; there are many examples of causal processes which can at present not be modelled as Markovian. Indeed, from a theoretical point of view, it would be disastrous to start by assuming well-behaved causal processes. What is needed is a *general* account of causation in terms of which the virtues of the various forms of well-behaved causal processes can be characterized. Hence, the strong assumption should be investigated at a later stage.

Cartwright (1979) is interested in causal laws rather than in singular causation. Nevertheless, it is instructive to compare her views with Definition 2. She argues that all the variables influencing B but not influenced by A constitute the obtaining circumstances of the causal relation between A and B and that conditionalization with respect to them tells us whether A is a cause of B . Though I conditionalize on much more, the conflict is less important than it seems. Cartwright rightly insists that one must not conditionalize with respect to variables mediating between cause and effect; indeed, if their values are given, the cause cannot be expected to be positively relevant to the effect. But in the special case of direct causation there are no mediating variables; and the difference is then reduced to the fact that I conditionalize also with respect to all variables which precede, but do not influence the effect, whereas she does not.

I think that the more extensive conditionalization proposal is harmless, but she does not. She says on p.

432 that "partitioning on an irrelevancy can make a genuine cause look irrelevant, or make an irrelevant factor look like a cause" and goes on to illustrate this alleged possibility. Eells and Sober (1983), who also conditionalize on irrelevant factors, argue on p. 42 that this illustration does not support Cartwright's restricted form of conditionalization; and I agree with them.

Maintaining Definition 2 has an important consequence. The upshot of Cartwright's paper is that there is no non-circular characterization of causation in probabilistic terms. But if I am right, this is certainly not true of direct causation. Hence, there is hope that Cartwright's skeptical view is not true of causation in general.

4. The circumstances of direct causes

The foregoing defense notwithstanding, it must be admitted that Definition 2 does not embody the *only* possible explication of obtaining circumstances. There are five further explications; and it is important to clarify them and to see the extent to which they are equivalent.

Definition 2 was based on the observation that each fact preceding the direct effect B and differing from the direct cause A is relevant as a circumstance. Here, "relevant" was used in the widest possible sense, namely as "possibly relevant solely on the basis of temporal relations", which is fixed in

DEFINITION 3a. Let ω , A , B , i , and j be as in Definition 2. Then *the temporally possibly relevant circumstances of* (the direct causal relation between) A and B in ω are defined as $C_{\omega}^{++}(A, B) = \omega \{ < j - i \}$.

This widest sense of "relevant" yields, as is to be expected, the narrowest circumstances. But there is a stricter sense of "possibly relevant". Whether a variable is relevant to the relation between A and B may also depend on the probabilities involved. To specify this idea, we will need

DEFINITION 4. $R_{\omega}(B)$ is to denote the set of all variables directly causally relevant to B in ω , i.e. $R_{\omega}(B) = \{ k \in \{ < j \} \mid \text{not } B \perp k / \omega \{ < j - k \} \}$. And $R(B)$ is to denote the set of all variables directly causally relevant to B in some world, i.e. $R(B) = \bigcup_{\omega \in \Omega} R_{\omega}(B) = \{ k \in \{ < j \} \mid \text{not } B \perp k / \{ < j - k \} \}$.

These sets will play an important role. A first crucial observation is

THEOREM 3. $R(B)$ is the smallest subset R of $\{<j\}$ such that $B \perp \{<j - R\}/R$.

By Definition 4, we have $k \in \{<j - R(B)\}$ iff $B \perp k/\{<j - k\}$; and from this Theorem 3 follows with the help of Theorem 2(e). Thus, $R(B)$ is the minimal set of variables preceding B which screens off all the other preceding variables from B ; i.e. given their values, B is probabilistically independent of all the rest of the possible past of B . This yields another sense of “relevant”, namely, “possibly relevant on the basis of temporal relations and probabilities alone”:

DEFINITION 3b. *The probabilistically possibly relevant circumstances of (the direct causal relation between) A and B in ω* are defined as $C_\omega^+(A, B) = {}^\omega(R(B) - \{i\})$.

What one usually has in mind, however, is not possible, but actual relevance; intuitively, it should suffice to consider only the actually relevant circumstances. Here is a first attempt of explication: Definition 2 can be interpreted metalinguistically as giving the truth conditions of the sentence “ A is a direct cause of B ”, i.e. as specifying when this sentence is true in a world ω . Viewed in this way, it seems plausible to say that the actually relevant circumstances of A 's being a direct cause of B just consist in the fact that A is a direct cause of B , i.e. in the set of all the worlds which relate A and B in this way; likewise for “direct counter-cause” and “direct causal irrelevance”. To render this idea precise we need the signum function for reals defined as $\text{sgn}(0) = 0$ and $\text{sgn}(x) = x/|x|$ for $x \neq 0$.

DEFINITION 3c. *The actually relevant circumstances of (the direct causal relation between) A and B in ω in the widest sense²³* are defined as $C_\omega''(A, B) = \{v \mid \text{sgn}[P(B|A \cap v\{<j - i\}) - P(B|\bar{A} \cap v\{<j - i\})] = \text{sgn}[P(B|A \cap \omega\{<j - i\}) - P(B|\bar{A} \cap \omega\{<j - i\})]\}$.

The deterministic analogue of this definition is not uninteresting, but the probabilistic concept is quite useless because it is not generally true that $\text{sgn}[P(B|A \cap C_\omega''(A, B)) - P(B|\bar{A} \cap C_\omega''(A, B))] = \text{sgn}[P(B|A \cap \omega\{<j - i\}) - P(B|\bar{A} \cap \omega\{<j - i\})]$; that is, if one conditionalizes on the circumstances in this widest sense, one may even get different causal conclusions. So the widest sense is too wide.

Here is a modification: The inadequate proposal holds that the actually relevant circumstances of A 's being a direct cause of B just consist in the fact that A is a direct cause of B . Now it seems that they rather consist in the fact that A is a direct cause of B in the way it actually is — where this additional clause refers to the specific numerical change of the probability of B which is actually due to A . The idea is captured in

DEFINITION 3d. *The actually relevant circumstances of (the direct causal relation between) A and B in ω in the wide sense* are defined as $C'_\omega(A, B) = \{v \mid \text{for each } A' \in \{A, \bar{A}\} P(B|A' \cap v\{<j - i\}) = P(B|A' \cap \omega\{<j - i\})\}$.

As can be easily shown, for each $\{<j - i\}$ -measurable $D \subseteq C'_\omega(A, B)$, $P(B|A' \cap D) = P(B|A' \cap \omega\{<j - i\})$ and hence $B \perp \{<j - i\}/A' \cap D$ for $A' \in \{A, \bar{A}\}$; in fact, $C'_\omega(A, B)$ is the largest $\{<j - i\}$ -measurable set for which this is true. Thus, $C'_\omega(A, B)$ represents the widest circumstances such that conditionalization on them agrees with conditionalization on any more narrowly taken circumstances of necessity and not by accident because of lucky averaging.²⁴ This strongly indicates that we have hit upon a reasonable explication.

So let me study $C'_\omega(A, B)$ a bit more closely. One valuable piece of information concerns which cylinders are subsets of $C'_\omega(A, B)$. It is given by

THEOREM 4. Let ω, A, B, i , and j be as in Definition 2. For each $v \in C'_\omega(A, B)$ and $K \subseteq \{<j - i\}$ we then have $v\{<j - K \cup \{i\}\} \subseteq C'_\omega(A, B)$ iff $B \perp K/A' \cap v\{<j - K \cup \{i\}\}$ for each $A' \in \{A, \bar{A}\}$.

For proof it is sufficient to consider Definitions 1 and 3d.

The theorem points to a useful distinction in $C'_\omega(A, B)$. Each $v \in C'_\omega(A, B)$ differs from ω on some variables. The only interesting differences are in $\{<j - i\}$, because outside $\{<j - i\}$ the members of $C'_\omega(A, B)$ may vary arbitrarily, anyway. Thus, let $K = \{k \in \{<j - i\} \mid v(k) \neq \omega(k)\}$. Now the distinction is this: one case is that v is in $C'_\omega(A, B)$ because all variations of ω on K are in $C'_\omega(A, B)$, i.e. because $\omega\{<j - K \cup \{i\}\} \subseteq C'_\omega(A, B)$ or, equivalently, $B \perp K/A' \cap \omega\{<j - K \cup \{i\}\}$ for $A' \in \{A, \bar{A}\}$. The other case is that these conditional independencies do not hold. In this case, v is, in a sense, only accidentally in $C'_\omega(A, B)$, i.e. not because the

variables in K do not matter to B given $\omega\{<j - K \cup \{i\}\}$ and A or \bar{A} . Rather, the variables in K do matter; it is only that in some particular realizations of K the relevant conditional probabilities come out the same as for ω and that v represents one such realization of K .

This suggests that the actually relevant circumstances of A and B in ω should be conceived a bit more narrowly, namely as comprising only all the arbitrary variations of ω in $C'_\omega(A, B)$.

DEFINITION 3e. *The actually relevant circumstances of (the direct causal relation between) A and B in ω in the narrow sense* are defined as $C_\omega(A, B) = \cup \{\omega\{<j - K \cup \{i\}\} \mid K \subseteq \{<j - i\}$ and $B \perp K/A' \cap \omega\{<j - K \cup \{i\}\}$ for each $A' \in \{A, \bar{A}\}$.

It will soon become clear why this is the preferred sense of the obtaining circumstances of a direct causal relation.

The five concepts of “obtaining circumstances” introduced so far are related in the following way:

THEOREM 5. $C_{\omega}^{++}(A, B) \subseteq C_{\omega}^{+}(A, B) \subseteq C_{\omega}(A, B) \subseteq C'_{\omega}(A, B) \subseteq C''_{\omega}(A, B)$; and if D and D' are any of these circumstances except $C''_{\omega}(A, B)$, then $P(B \mid A' \cap D) = P(B \mid A' \cap D')$ for each $A' \in \{A, \bar{A}\}$.

One may object that the most obvious suggestion has been ignored. Isn't it very natural to think that the actual circumstances of the direct causal relation between A and B are just all of the other actual direct causes and counter-causes of B ? Indeed. This is precisely the proposal of Cartwright (1979) restricted to direct causes; Mellor (1988), p. 234, explicitly endorses it, too; and it seems to be a natural “actualization” of Definition 3b where the circumstances of A and B in ω that are possibly relevant in the probabilistic sense are defined as the conjunction of all the facts in ω which are possibly directly causally relevant to B . This suggestion is fixed in

DEFINITION 3f. *The ideal circumstances of (the direct causal relation between) A and B in ω* are defined as $C^*_{\omega}(A, B) = \cap \{D \mid D \text{ is a } k\text{-state for some } k \neq i \text{ and } D \overset{\omega}{\dashv} B\} = \omega(R_{\omega}(B) - \{i\})$.²⁵

For the moment “ideal” means something bad. The basic trouble is that we cannot prove that $C^*_{\omega}(A, B) \subseteq C'_{\omega}(A, B)$. This means that the relevant probabilities

conditional on the ideal circumstances may well differ from those conditional on the circumstances in the senses accepted so far. How can this happen? This is made clearer by a more positive result:

THEOREM 6. $C_{\omega}(A, B) \subseteq C^*_{\omega}(A, B)$, and the identity holds iff for $K = \{k \in \{<j - i\} \mid B \perp k/\omega\{<j - k\}\} = \{<j - R_{\omega}(B) \cup \{i\}\}$ we have $B \perp K/A' \cap \omega\{<j - K \cup \{i\}\}$ for each $A' \in \{A, \bar{A}\}$.

Again, the proof essentially requires writing out the appropriate definitions. The theorem says that the identity holds if and only if the variables which are individually independent of B given the rest of the actual past of B are also collectively independent of B given A and the rest of the actual past of B as well as given \bar{A} and the rest of the actual past of B . Both aspects of this condition are easily violated, but it will suffice to exemplify this for the aspect relating to A and \bar{A} (and not for the one about collective independence):

Suppose A precedes D , D precedes B , $A \cap D \cap B = \{\omega\}$, $P(B \mid A \cap D) = 0.9$, $P(B \mid A \cap \bar{D}) = 0.9$, $P(B \mid \bar{A} \cap D) = 0.1$, and $P(B \mid \bar{A} \cap \bar{D}) = 0.5$. Here is a dream: there is hardly anything more delicious than red orange juice, but it is not offered in the deli-shops. So I thought that this was a way to become rich (B) and started a red orange juice enterprise. But what should I charge? Either \$2.99 (D) or \$1.99 (\bar{D}) per half a gallon; the prices in between are taboo, and higher or lower prices would be disastrous. In my dream I was lucky; nobody had the same idea (A). But then it is quite plausible to assume that it does not matter how I fix the price. If I fix the price to be high, I sell less with a larger profit per unit; otherwise, I sell more with a smaller profit per unit. My prospects of B are equally favorable. Thus, according to the numbers and Definition 2, D is directly causally irrelevant to B in ω . If there were competitors (\bar{A}), however, the price would of course make a big difference. Now look at the relation between A and B . A is a direct cause of B in ω , and also in $A \cap \bar{D} \cap B$; the fact that I have a monopoly is in any case advantageous to B . What are the circumstances of $A \overset{\omega}{\rightarrow} B$? The crucial comparison is that $C_{\omega}(A, B) = D$, but $C^*_{\omega}(A, B) = \Omega$. Thus, we face here the strange fact that D is directly causally irrelevant to B , but relevant to $A \overset{\omega}{\rightarrow} B$.

This possibility is, I think, responsible for quite some perplexity found in the literature. One may explain it away by resorting to a finer causal analysis in which D

turns out to be indirectly causally relevant to B ; but it is an open question whether this strategy always works. One may take it as constituting an objection against Definition 2; but this does not invalidate the other reasons for our explication. Maybe there are other ways to deal with the problem, but I think the possibility must be admitted that the two causal roles of D fall apart, i.e. that D 's being relevant to the direct causal relation of other facts to B does not coincide with D 's itself being directly causally relevant to B . However, if such behaviour is considered an anomaly, I propose to state an assumption excluding it. Then one can study how causal structures behave in general and how much more nicely they behave when this assumption is satisfied. Indeed, this assumption will play an important role later on.

What is the assumption? It was already stated in Theorem 6; it is the identity of $C_\omega(A, B)$ and $C_\omega^*(A, B)$. This explains why I have called $C_\omega^*(A, B)$ the ideal circumstances of A and B in ω ; it specifies how the circumstances ideally are, but need not be. Finally this is the deeper reason why $C_\omega(A, B)$ is the preferred explication of the actually relevant circumstances; among all the otherwise equally acceptable explications this is the only one which lends itself to a statement of the assumption of ideal circumstances.

5. The difficulties with indirect causation

It is now time to tackle the explication of indirect causation and hence of causation in general which, as the literature shows, is a difficult matter. Why? The general reason is that, even within our parsimonious framework, there is a bewildering plethora of plausible conditions for causation which cannot be simultaneously satisfied. The main purpose of this section is to present and untangle these conditions. Three kinds of conditions will be dealt with extensively and two others mentioned. A secondary goal is to show that the difficulties with these conditions are largely independent of the particular definition of direct causation one adopts. Therefore, little use of Definition 2 is made in this section; the synthesis is undertaken only in the final section.

The first condition is rather a matter of faith: namely that an explication of causation be simple. This sounds quite airy because simplicity ratings often diverge. But it helps to avoid the manifest danger of lapsing into the strategy of trying to solve difficulties by piling up clauses

and provisos, each of them plausible, but all together unintelligible.

The second condition is that there must be a good overall fit between an explication and the many more or less problematic examples found in the literature. Obviously the whole story necessary to show that a given explication satisfies this condition is long, indeed too long for this essay. But I have reservations about abbreviating the story. There is some tendency to focus on this or that problematic type of example as the central touchstone of any theory of causation. But this would be too narrow an attitude; there are too many types of examples to be considered, and intuitions about examples are not fixed enough to constitute an unshakable reference point. As I said, a good *overall* fit is to be achieved, even if this standard opens a door to vagueness and subjectivity. Moreover, examples are in a sense theoretically barren. We do not understand them as long as we have no theoretical structure enabling us to integrate them and to explain why they are examples for this or against that; and staring at them probably is bad heuristics for arriving at that structure. This is why I concentrate here on three further kinds of conditions of a theoretical nature.

The third kind of condition consists in structural conditions concerning the formal structure of causal relations. The fourth kind consists in Markovian conditions: there is a strong intuition that causal chains are Markov chains; and of course an indirect cause should be connected to its indirect effect by some causal chain. The fifth kind consists in positive relevance conditions: there is also a strong intuition that a cause is in some sense positively relevant to its effect; it is, indeed, embodied in the basic conception of causation cited in the very first sentence of section 3, and Definition 2 also relies on it.

There are alternative ways of specifying each kind of condition. It will turn out that the most plausible candidates are mutually incompatible. Recognition of this fact is important to the explanation of a number of examples and confusions. Let us look at these conditions in more detail.

Structural conditions. The first structural condition for the relation $\xrightarrow{\omega}$ of being a (direct or indirect) cause in ω is trivial, but should be made explicit:

(S0) *Lower bound:* If $A \xrightarrow{\omega} B$, then $A \xrightarrow{\omega} B$.

I shall continually use (S0) without mention. The next condition sets an upper bound:

(S1) *Upper bound*: If $A \xrightarrow{\omega} B$, then A stands to B in the transitive closure of $\xrightarrow{\omega}$.

This condition is not acceptable for continuous time. But given discrete time, (S1) seems compelling; I cannot imagine how indirect causation could extend farther than what is allowed by direct causal steps. The next all-important condition is

(S2) *Transitivity*: If $A \xrightarrow{\omega} B$ and $B \xrightarrow{\omega} C$, then $A \xrightarrow{\omega} C$.

(S0), (S1), and (S2) entail that $\xrightarrow{\omega}$ is the transitive closure of $\xrightarrow{\omega}$. Thus, as mentioned in the introduction, these conditions yield a definition of causation in general. So where is the snag? It lies in the fact that all Markovian and positive relevance conditions violate transitivity. This will become fully clear below. But the gist is easily summed up:

Though transitivity looks very natural, one would expect transitivity to ensue from a general definition of causation. If it is the other way around, naturalness is tarnished. Now, if deterministic causes are defined as sufficient and/or necessary conditions, transitivity follows immediately — at least when such conditions are explained in terms of logical or nomological entailment. This is certainly the strongest source of the intuition of transitivity. But even in the case of deterministic causation the issue is not clear. If such conditions are explained in terms of the subjunctive conditional, transitivity fails because the subjunctive conditional fails to be transitive.²⁶ Thus, even in this case a conflict arises. Lewis (1973b) resolves it by axiomatically accepting transitivity, at the cost of renouncing the general equation between causation and sufficient and/or necessary conditions and taking transitivity as a primitive property.

In the case of probabilistic causation, the issue is even less clear. Here, a direct causal impact has, so to speak, no necessitating force, but is only weak and imperfect.²⁷ Hence, it seems plausible that such a weak impact is not preserved over long causal chains, but fades sooner or later. For instance, given our very coarse and only probabilistic meteorological models, each day's weather may be granted to causally influence the next day's weather. But does the weather, say, at the turn of the last century still influence today's weather? It does not seem so; somewhere in between the influence has faded completely, even though it may be difficult to tell precisely when or where. If this is plausible, the intuition

of transitivity totters. Indeed, this intuition is not generally respected by theorists of probabilistic causation. For example, Suppes (1970), p. 58, dryly states that all causal relations he has defined fail to be transitive as long as the limiting probabilities 0 and 1 are not involved.²⁸

Thus, a profound uncertainty about this issue may be observed, and there is reason for looking for alternatives to transitivity. Here is a possible approach: Certainly, each indirect cause and effect should be connected by a causal chain. Everything then depends on how causal chains are characterized, and they may indeed be characterized in several, apparently non-equivalent ways:

DEFINITION 5.

- (a) $\langle A_1, \dots, A_n \rangle$ is a *weak causal chain in ω* iff $A_1 \xrightarrow{\omega} A_2 \xrightarrow{\omega} \dots \xrightarrow{\omega} A_n$.
- (b) $\langle A_1, \dots, A_n \rangle$ is a *connected causal chain in ω* iff it is a weak causal chain in ω and, for all r and s with $1 \leq r < s \leq n$, $A_r \xrightarrow{\omega} A_s$.
- (c) $\langle A_1, \dots, A_n \rangle$ is a *strict causal chain in ω* iff it is a connected causal chain in ω and, for no r and s with $r < s$, $A_r \xrightarrow{\omega} A_{s+1}$.
- (d) $\langle A_1, \dots, A_n \rangle$ is an *effective causal chain in ω* iff it is a weak causal chain in ω and, for all $r > 1$, $A_1 \xrightarrow{\omega} A_r$.
- (e) $\langle A_1, \dots, A_n \rangle$ is an *affective causal chain in ω* iff it is a weak causal chain in ω and, for all $s < n$, $A_s \xrightarrow{\omega} A_n$.

Of course, (a), (b), and (c) are the more promising definitions; the reason for introducing also (d) and (e) will be clear in due course. Note the difference between (b) and (c): if $A \xrightarrow{\omega} B \xrightarrow{\omega} C$ and also $A \xrightarrow{\omega} C$ — a situation which has in no way been excluded so far —, then $\langle A, B, C \rangle$ is a connected, but not a strict causal chain in ω . Correspondingly, there are five structural conditions for $\xrightarrow{\omega}$:

(S3) *Structural chain conditions*: Whenever $A \xrightarrow{\omega} B$, but not $A \xrightarrow{\omega} B$, then there are $C_1, \dots, C_n (n \geq 1)$ such that $\langle A, C_1, \dots, C_n, B \rangle$ is an (a) weak, (b) connected, (c) strict, (d) effective, (e) affective causal chain in ω .

It might be tempting to reverse (S3), in particular part (b), i.e. to take the fact that in a series of states starting with A and ending with B all causal relations except the

one from A to B obtain to imply that A also causes B . Formally:

(S3b') *Reversed chain condition*: if there is a connected causal chain $\langle C_1, \dots, C_n \rangle$ in ω such that $\langle A, C_1, \dots, C_n \rangle$ is an effective causal chain in ω and $\langle C_1, \dots, C_n, B \rangle$ is an affective causal chain in ω , then $A \xrightarrow{\omega} B$, i.e. $\langle A, C_1, \dots, C_n, B \rangle$ is also a connected chain in ω .

Another idea is that: by assuming transitivity the causal relation $\xrightarrow{\omega}$ is boosted to its maximal extension within the upper bound (S1). Thus, if transitivity is dropped, the effective range of a state A in ω , i.e. the set of its effects in ω , may comprise less than all states which can be reached from A via weak causal chains. How much less? It is hard to say. But in any case, it seems impossible that the effective range of A extends farther than the effective ranges of all its immediate causal successors:

(S4a) *Local effective maximum*: whenever $A \xrightarrow{\omega} B$, but not $A \xrightarrow{\omega} C$, then there is a C with $A \xrightarrow{\omega} C \xrightarrow{\omega} B$.

The same consideration holds, of course, for the affective range of A , i.e. the set of its causes:

(S4b) *Local affective maximum*: whenever $B \xrightarrow{\omega} A$, but not $C \xrightarrow{\omega} A$, then there is a C with $B \xrightarrow{\omega} C \xrightarrow{\omega} A$.

These suggestions demonstrate the ease with which further conditions may be invented. But there is no point in doing so. More interesting is the relation between the conditions stated so far. This is given completely by

THEOREM 7.

- (a) Upper bound (S1) and transitivity (S2) are equivalent to the assertion that $\xrightarrow{\omega}$ is the transitive closure of $\xrightarrow{\omega}$.
- (b) Upper bound (S1) and the reversed chain condition (S3b') are also equivalent to this assertion.
- (c) Given (S1), transitivity (S2) implies the connected chain condition (S3b); but the reverse does not hold.
- (d) The connected (S3b) and the strict (S3c) chain condition are equivalent.
- (e) The connected chain condition (S3b) implies the effective (S3d) and the affective (S3e) chain condi-

tion; but even jointly, (S3d) and (S3e) do not imply (S3b).

- (f) Local effective maximum (S4a) is equivalent to the affective chain condition (S3e).
- (g) Local affective maximum (S4b) is equivalent to the effective chain condition (S3d).
- (h) Each of (S3d) and (S3e) imply the weak chain condition (S3a); but the reverse does not hold.
- (i) Upper bound (S1) is equivalent to the weak chain condition (S3a).

PROOF:

(a) is trivial.

(b) For the direction (\Rightarrow) suppose that it has been shown for all $r < n$ that we have $C_1 \xrightarrow{\omega} C_r$ for each weak causal chain $\langle C_1, \dots, C_r \rangle$ in ω of length r . Now, let $\langle C_1, \dots, C_n \rangle$ be a weak causal chain in ω of length n . Because of the supposition the premises of (S3b') are satisfied, and so we may infer that $C_1 \xrightarrow{\omega} C_n$. Hence, the existence of a weak causal chain in ω from A to B already ensures $A \xrightarrow{\omega} B$. With upper bound (S1) this implies the desired result. The other direction (\Leftarrow) is trivial.

(c) The direction (\Rightarrow) is trivial. Concerning the reverse, imagine that $A \xrightarrow{\omega} B \xrightarrow{\omega} C$, but not $A \xrightarrow{\omega} C$. This situation satisfies (S3b), but not transitivity (S2).

(d) Each strict causal chain is connected; thus (S3c) implies (S3b). On the other hand, suppose that $\langle A_1, \dots, A_n \rangle$ is a connected causal chain in ω . Let $A_1 = B_1$; $B_2 = A_r$, where r is the maximal index for which $A_1 \xrightarrow{\omega} A_r$; $B_3 = A_s$, where s is the maximal index for which $A_r \xrightarrow{\omega} A_s$; etc. Thus, for some m $B_m = A_n$. Obviously, $\langle B_1, \dots, B_m \rangle$ is a strict causal chain in ω . This shows that each connected causal chain has a strict causal subchain with the same start and end. Hence, (S3b) also implies (S3c).

(e) The direction (\Rightarrow) is trivial. That the reverse does not hold, may be seen in the following way: Take a weak causal chain $\langle A_1, \dots, A_5 \rangle$ in ω with 5 members, and suppose that $A_r \xrightarrow{\omega} A_s$ for all $r < s$ with the exception of $r = 2$ and $s = 4$. Moreover, assume states B and C such that both, $\langle A_1, B, A_4 \rangle$ and $\langle A_2, C, A_5 \rangle$, are effective and affective causal chains in ω , but neither $A_1 \xrightarrow{\omega} C$ nor $B \xrightarrow{\omega} A_5$ holds. In this situation, there is an effective and affective causal chain from each D to each E for which $D \xrightarrow{\omega} E$, but there is no connected causal chain from A_1 to A_5 .

(f) For the direction (\Rightarrow), suppose that $A \xrightarrow{\omega} B$, but not $A \xrightarrow{\omega} C$. According to (S4a), there is a C_1 with

$A \xrightarrow{\omega} C_1 \xrightarrow{\omega} B$. Now, if $C_1 \xrightarrow{\omega} B$, we are finished. If not, we again apply (S4a) and find a C_2 with $C_1 \xrightarrow{\omega} C_2 \xrightarrow{\omega} B$. And so on. In the end, this process yields an affective causal chain in ω from A to B . The reverse direction (\Leftarrow) is trivial.

- (g) This is perfectly symmetric to (f).
- (h) and (i) are trivial. \square

The fact that some reverse inferences are not valid is perhaps a little surprising. In any case, Theorem 7 exhibits clearly the differing strengths of the various conditions.

It is not yet the time to decide which structural conditions are the most plausible ones. But one conclusion is quite obvious: if one should give up the transitivity of $\xrightarrow{\omega}$ and decide to settle for something weaker, one gets onto a slippery slope, at least from a purely structural point of view. For instance, it is an unpleasant fact that then one has to cope with various non-equivalent concepts of a causal chain. However, if transitivity is assumed, each weak causal chain is also effective, affective, and connected, and, thus, no ambiguity would arise. A further point is that, intuitively, it may be not so clear which chain condition direction to endorse. The reverse direction — stated in (S3b') in its weakest form — may also seem plausible; but it implies transitivity according to Theorem 7(b). Thus, in the light of structural considerations alone, transitivity (S2) has a clear preponderance over the alternatives. So, let us look more closely at the reasons against transitivity.

Markovian conditions. There is a strong intuition that indirect effects are screened off from their indirect causes by the mediating links, i.e. that the indirect causal efficacy of a state is completely contained in the mediating links, or, in other words, that, if the intermediate members of a causal chain are realized in some way or other, then the past of the chain is irrelevant to its future. This intuition is commonly expressed in the Markovian way; indeed, it is often said that Markov chains have no memory, that they are characterized by the absence of after-effect. The central concept is:

DEFINITION 6. $\langle i_1, \dots, i_n \rangle$ is a (*finite*) *Markov chain* iff $i_{r+1} \perp \{i_1, \dots, i_{r-1}\} / i_r$ for all $r = 2, \dots, n - 1$. Moreover, $\langle i_1, \dots, i_n \rangle$ is a *causal Markov chain* in ω iff it is a Markov chain and a weak causal chain in ω (where $\langle i_1, \dots, i_n \rangle$ is a weak causal chain in ω iff $\langle {}^\omega i_1, \dots, {}^\omega i_n \rangle$ is).

Unfortunately, there are various choices for rendering precise the Markovian intuition. Is it intuitively a necessary condition for a causal chain to be a causal Markov chain? Or a sufficient condition that it be both a Markov and a strict causal chain? This is not easy to decide. Let us look at one attempt a bit more closely:

(M1) *Markov chain condition*: suppose that $A = {}^\omega i$ and $B = {}^\omega j$. Then $A \xrightarrow{\omega} B$ iff there exist k_1, \dots, k_n ($n \geq 0$) such that $\langle i, k_1, \dots, k_n, j \rangle$ is a causal Markov chain in ω .

(M1) is a biconditional and thus bolder than (S3). But it rests on the same basic idea, namely, that indirect causation must be mediated by a causal chain; and it adds a particular explication of causal chains. Indeed, (M1) provides an explicit definition of $\xrightarrow{\omega}$ according to which it behaves thus:

THEOREM 8. The Markov chain condition (M1) implies the strict chain condition (S3c), but it does not imply transitivity (S2).

PROOF: The first part follows from the well-known fact that, if $\langle i_1, \dots, i_n \rangle$ is a Markov chain, then any subsequence $\langle j_1, \dots, j_m \rangle$ of $\langle i_1, \dots, i_n \rangle$ is also a Markov chain. The failure of transitivity is due to the fact that the serial connection of two or more Markov chains will not generally result in one large Markov chain. \square

Hence, the Markov chain condition does away with the unwelcome splitting up of the structural characterizations of causal chains. Apparently, it is a serious alternative to the assumption of transitivity. But there are, on the contrary, also clearly disconcerting features.

First, it is easy to see that according to (M1) there may be connected causal chains which are not Markovian. Suppose that $A = {}^\omega i$, $B = {}^\omega j$, $C = {}^\omega k$, and $D = {}^\omega l$, and that $A \xrightarrow{\omega} B \xrightarrow{\omega} D$, $A \xrightarrow{\omega} C \xrightarrow{\omega} D$, and $A \xrightarrow{\omega} D$. Then both, $\langle A, B, D \rangle$ and $\langle A, C, D \rangle$, are connected causal chains in ω . But in order to satisfy (M1) only one of $\langle i, j, l \rangle$ and $\langle i, k, l \rangle$ needs to be a Markov chain; there is nothing so far to guarantee that the other is so, too. If $\langle i, j, l \rangle$ is the Markov chain, should then the other, $\langle i, k, l \rangle$, be denied to be a genuine causal chain? Or should one stipulate that such a situation does not arise, i.e. that not only some, but all connected causal chains leading from one state to another are Markovian? No; a more trenchant conclusion is called for.

Consider an illustration of the very common abstract

situation just described. At a signal 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, as required, two causal chains running from A to D, one through B and the other through C. It is plausible in this case, and easily done, to distribute the probabilities in such a way that, given B alone, A is still probabilistically positively relevant to D (via C, so to speak), and also given C alone; the situation is symmetric with respect to B and C. This, however, means that, contrary to (M1), no causal chain between A and D is Markovian and thus that (M1) somehow fails to capture the Markovian intuition.

A well-known move for coping with such problems is to generalize the concept of a Markov chain to that of a Markov field.²⁹ In these terms, the case exemplifies a Markov field characterized by the conditional independence $I \perp i / \{j, k\}$ which says that D is screened off from A only jointly by B and C. In principle, I fully endorse this strategy,³⁰ but not at the present stage where it seems to me to overshoot the mark. If one adopts this strategy, the conceptual key role is taken over by the notion of a Markov field and the corresponding causal notion of a causal net, which are more complex notions and more difficult to grasp. The structural and the Markov conditions would then have to be expressed in these more complex terms. And causal chains become derivative entities definable only as certain parts of causal nets. This seems to be the wrong direction of analysis; we should build up complexities from simpler units already understood.

Indeed, a less radical move will do; a slight, though basic conceptual modification will save the old strategy. In explicating direct causes, the positive correlation between direct cause and effect was considered not in isolation, but embedded in the given past course of events. Similarly, the members of a Markov chain should be taken not in isolation, but rather as embedded in a given setting. Thus, I propose:

DEFINITION 7. $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain iff $i_1 < \dots < i_n$ and if, for all $r = 2, \dots, n - 1$, $i_{r+1} \perp \{i_1, \dots, i_{r-1}\} / i_r, \omega \{< i_{r+1} - i_1, \dots, i_r\}$; this means that the conditional independence characteristic of a Markov chain holds only given the rest of the past of i_{r+1} in ω . Moreover, $\langle i_1, \dots, i_n \rangle$ is a causal ω -Markov chain iff $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain and a weak causal chain in ω .

Let us modify (M1) correspondingly:

(M2) ω -Markov chain condition: Suppose that $A = {}^\omega i$ and $B = {}^\omega j$. Then $A \xrightarrow{\omega} B$ iff there exist k_1, \dots, k_n ($n \geq 0$) such that $\langle i, k_1, \dots, k_n, j \rangle$ is a causal ω -Markov chain.

This amendment takes care of the example of the romantic lover; there, $\langle i, j, l \rangle$ and $\langle i, k, l \rangle$ both plausibly are ω -Markov chains. Indeed, I think that (M2) reflects the Markovian intuition better than (M1). Generally, the expectation should be that a more proximate cause screens off the effect from a more remote cause only given the circumstances and not unconditionally. The structural properties have not changed, however:

THEOREM 9. The ω -Markov chain condition (M2) implies the strict chain condition (S3c), but it does not imply transitivity (S2).

PROOF: It is easily shown that, if $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain, then any sub-sequence $\langle j_1, \dots, j_m \rangle$ of $\langle i_1, \dots, i_n \rangle$ is also an ω -Markov chain. This immediately entails the first part. Again, the serial connection of two or more ω -Markov chains will in general not result in one large ω -Markov chain. Thus, transitivity need not hold. \square

I conclude that (M2) is a viable alternative to transitivity (S2). But the matter is still open, and the evidential basis should be further augmented.

Positive relevance conditions. A third important theoretical constraint is introduced by the conception that a cause is in some sense positively relevant to its effects. The account of direct causation above is based on that conception; and theoretical unity seems to be best preserved by further relying on it. This sets the task to extend section 4 and to determine the circumstances also of indirect causal relations, which is in fact so intricate that it can only be started, but not completed here.

Recall first Cartwright's circularity problem. In the case of direct causation, it could be argued that the whole past of the effect may be taken as obtaining circumstances. But, for indirect causation, obviously this will not do. Precisely because of the Markovian intuition, some causal intermediates must be excluded from the circumstances in this case; and the problem is to say which ones. However, we need not yet worry about the threat of circularity because we are now after plausible conditions only and not after definitions.

A useful perspective on the problem is gained when

we think of the relativity of the direct/indirect-distinction to the given descriptive frame. The core of the idea of positive relevance is, I think, the expectation that what is an indirect cause in the given descriptive frame should be a direct cause in some reduced descriptive frame. The advantage of putting the core idea in this way is that it avoids reference to a specific explication of direct causal relevance (though I shall employ the account of section 3 later on). The problem takes then the form of determining which reduced frame to consider. There are, *prima facie*, several options:

- (P) *Positive relevance conditions*: Let A be an i -state and B a j -state. Then $A \xrightarrow{\omega} B$ holds relative to the descriptive frame given by I iff $A \xrightarrow{\omega} B$ holds relative to the descriptive frame given by $I-J$
- (1) for some $J \subseteq (i, j)$,
 - (2) for some, or (3) for each, $J = \{k_1, \dots, k_n\} \subseteq (i, j)$ such that $\langle i, k_1, \dots, k_n, j \rangle$ is (a) a weak, (b) a connected, (c) a strict causal chain in ω , (d) a causal Markov chain in ω , (e) a causal ω -Markov chain,
 - (4) for $J = \{k \in (i, j) \mid k \text{ is a member of (a) a weak, (b) a connected, (c) a strict causal chain in } \omega, \text{ (d) a causal Markov chain in } \omega, \text{ (e) a causal } \omega\text{-Markov chain running from } i \text{ to } j\}$,
 - (5) for $J = (i, j)$.

As a sufficient condition for $A \xrightarrow{\omega} B$, (P1) is certainly too weak. But taken as a necessary condition, (P1) seems to be the inalienable minimum of the positive relevance idea. However, even this minimum need not be satisfied in the light of the theory proposed below.

(P5) is, as noted above, the version favored by Good (1961). The example of the two-person game in section 3 may also be used to cast doubt on (P5). Suppose that the choice A of the first player is negatively relevant to a certain outcome C , given the later, but independent choice B of the second player. On the current account, A is then a counter-cause of C . But it is easily imaginable that averaging over the choice of the second player makes A unconditionally positively relevant to C ; just assume that C is sufficiently unlikely given \bar{A} and \bar{B} . According to (P5), A would then be a cause of C . This seems inadequate.

(P4) is the best approximation to the position of Cartwright (1979), though which version of (P4) she would prefer is not clear. Is (P4) plausible? If there is

only one causal chain running from i to j , then (P2), (P3), and (P4) coincide. But if there is more than one chain, the three conditions may diverge almost arbitrarily; only the versions of (P3) are guaranteed to be stronger than the corresponding versions of (P2). In view of this divergence, it is hard to say which condition is preferable.

But we know some things. An important observation is that each version of (P) violates transitivity (S2), at least if Definition 2 is presupposed. One numerical example covers all versions. Suppose that $A \cap B \cap C = \{\omega\}$, $P(B|A) = 0.8$, $P(B|\bar{A}) = 0.4$, $P(C|A \cap B) = 0.6$, $P(C|A \cap \bar{B}) = 0.1$, $P(C|\bar{A} \cap B) = 0.9$, and $P(C|\bar{A} \cap \bar{B}) = 0.4$. Then $A \xrightarrow{\omega} B \xrightarrow{\omega} C$ holds according to Definition 2; but we have $P(C|A \cap B) < P(C|\bar{A} \cap B)$ and $P(C|A) = 0.5 < 0.6 = P(C|\bar{A})$. Thus, according to all versions of (P) we cannot have $A \xrightarrow{\omega} C$. Generally, structurally good behavior may at most be expected from the (d) and (e) versions of (P2) and (P3) which incorporate Markovian elements.

Of course, it is rather the harmony between Markovian and positive relevance conditions which is hoped for. This hope is based on the following well-known result:

THEOREM 10. Let $\langle i_1, \dots, i_n \rangle$ be a Markov chain of binary variables and $A_r = \omega i_r$ for $r = 1, \dots, n$. Then, if $P(A_{r+1}|A_r) - P(A_{r+1}|\bar{A}_r) = x_r$ for $r = 1, \dots, n-1$, $P(A_n|A_1) - P(A_n|\bar{A}_1) = x_1 \cdot \dots \cdot x_{n-1}$. This implies in particular that, if each A_r is positively relevant to A_{r+1} , then A_1 is positively relevant to A_n .³¹

However, Theorem 10 does not achieve the desired harmony. If additional variables are dispersed between i_1, \dots, i_n , then, according to Definition 2, the positive relevancies assumed in Theorem 10 need not indicate direct causal relations. Moreover, the theorem refers only to the Markovian chain condition (M1). If, however, the right strategy is to replace Markov chains by ω -Markov chains, then the theorem does not apply; and there is no corresponding theorem about ω -Markov chains.

So the situation is, in fact, as bad as I have indicated. All Markovian and positive relevance conditions are incompatible with the favourite structural condition of transitivity; and the preferred Markovian condition need not preserve positive relevance in any of the ways considered.

Suppes (1984), pp. 55ff., devotes a whole section to

“conflicting intuitions” concerning causality, and Salmon (1988b) appreciatively adopts this phrase though he has, in part, different things in mind. It is my experience that, given the current parsimonious framework and no further notions or distinctions bearing on causality, the intuitions and conflicts described in the present section are central to the discussion of probabilistic causation. Is there any way to resolve these conflicts? The final section suggests one such way.

6. Causation

For a long time, Suppes’ remark (1970, p. 58) that transitivity is not to be expected in the case of probabilistic causation held me in its grip. All proposed plausible explications so clearly failed to yield transitivity that it seemed crazy to cling to that structural property. The task could thus only be to reconcile the other intuitions and conditions; and Theorem 10 seemed to point the way. Eells and Sober (1983) obviously had the same idea in mind when investigating the lucky circumstances under which the causal relation defined on the basis of the positive relevance idea is transitive.

Moreover, when faced with several options, it is always a wise policy to choose the weakest explication possible. The various strengthenings can then be introduced and studied afterwards. Were one, on the contrary, to start with a stronger notion, all the weaker ones would simply drop out of theoretical consideration. But Theorem 7 has revealed transitivity to be a particularly strong structural condition; it implies all the other conditions, given the unassailable (S0) and (S1). So, again, it appeared better to ignore transitivity.

However, the consideration does not apply the right measure of strength. What counts is not structural strength, which was seen to be accompanied by weakness concerning other kinds of conditions. What counts is *conceptual* strength. And the fact is that the transitive closure of direct causation is the *weakest possible notion* of causation in general; it yields the causal relation with the widest possible extension, if the upper bound condition (S1) is presupposed; whenever A is a cause of B in any other feasible sense, A is also a cause of B in this sense.³² So, I shall settle for the minimal notion of causation, even if the price to be paid is what Lewis (1973b) had to pay, too, namely, that transitivity is a primitive property.

DEFINITION 8. Let A be an i -state, B a j -state, $i < j$, and $\omega \in A \cap B$. Then A is a *cause of B in ω* , i.e. $A \xrightarrow{\omega} B$, iff A stands to B in the transitive closure of $\xrightarrow{\omega}$. A is a *counter-cause of B in ω* , i.e. $A \xrightarrow{\omega} \leftarrow B$, iff $A \xrightarrow{\omega} \leftarrow B$ or for some D $A \xrightarrow{\omega} D \xrightarrow{\omega} \leftarrow B$. A is *causally relevant to B in ω* , i.e. $A \xrightarrow{\omega} \leftrightarrow B$, iff A stands to B in the transitive closure of $\xrightarrow{\omega} \diamond$. Finally, A is *causally irrelevant to B in ω* , i.e. $A \xrightarrow{\omega} \parallel B$, iff not $A \xrightarrow{\omega} \leftrightarrow B$.

This definition of counter-causation is, I think, the most plausible one. If counter-causation is to be allowed at all, then the counter-causal influence stops with the realization of the counter-effect and does not extend beyond. At least, I would firmly claim this for deterministic causation (where there may be counter-causation, too) and thus also for probabilistic causation, though less firmly.³³ For the unconvinced there is also the concept of causal relevance which comprises causation, counter-causation, and much more.

Definition 8 covers all kinds of weird cases. First, to repeat, causal chains need not be Markov chains. Also, a cause need not be positively relevant to its effect under admissible conditionalization. The numerical example in the foregoing section demonstrating the incompatibility of (P) and (S2) is a case in point; according to Definition 8, $A \xrightarrow{\omega} C$ holds in this case. $A \xrightarrow{\omega} \leftarrow C$ holds also; indeed, this is essential to its construction. This means that a state of affairs may at once be a cause and a counter-cause of another state of affairs, if at least one of these causal relations is indirect. Though this may appear counter-intuitive, it seems to be exactly the right thing to say in many cases — for instance in the famous thrombosis example of Hesslow (1976): the woman’s taking a contraceptive is a cause as well as a counter-cause of her thrombosis, mediated by different chains. Otte (1985), pp. 122f., has drawn this conclusion, also.

The counterpart may also happen: A is a cause of B , and if \bar{A} had obtained, \bar{A} would have been a cause of B , too. This is the case when causal preemption occurs or a back-up system is installed. Consider, e.g., the equally famous case of the desert traveller introduced by Hart, Honoré (1959), pp. 239ff. One enemy of the traveller pours poison into his water keg; later, but independently, the other drills a hole into the keg. The latter fact is a cause of the traveller’s death. But if the hole had not been drilled, the lack of the hole would also have been a cause of his death, because it would have caused the poison to stay in the keg. In other words, the set-up can

be such that a variable has only a relay function; it can turn on different chains which all lead to the same effect. Nevertheless, this relay function must be conceived as a causal function. The list of such oddities could easily be extended.

Settling for the weakest notion is only a start. The essential step consists in showing how to build stronger conceptions upon the weak base. Thus, the task is to specify conditions under which Definition 8 does justice to the Markovian and positive relevance intuitions. And it will not do to find just any sufficient conditions; that would presumably be easy. These conditions must be specified solely in causal terms; only then do we know which causal situations satisfy our causal preconceptions. To this task I now and finally turn.

The Markovian part is the easier one. First some terminology. Since Definition 8 entails that the four kinds of causal chains given by Definition 5(a), (b), (d), and (e) coincide, I shall now talk of causal chains *simpliciter*; only strict causal chains in the sense of Definition 5(c) have to be distinguished. Additionally, we need

DEFINITION 9. $\langle A_1, \dots, A_n \rangle$ is a *chain of causal relevance in ω* iff $A_1 \xrightarrow{\omega} A_2 \xrightarrow{\omega} \dots \xrightarrow{\omega} A_n$; and it is a *strict chain of causal relevance in ω* iff it is a chain of causal relevance in ω and, for no r and s with $r + 2 \leq s \leq n$, $A_r \xrightarrow{\omega} A_s$. Moreover, $\langle i_1, \dots, i_n \rangle$ is a (strict) chain of causal relevance in ω iff $\langle {}^\omega i_1, \dots, {}^\omega i_n \rangle$ is.

A first welcome result is

THEOREM 11. The following two assertions are equivalent:

- (a) for each $\omega \in \Omega$, all chains $\langle i_1, \dots, i_n \rangle$ of causal relevance in ω are strict;
- (b) for each $\omega \in \Omega$, all chains $\langle i_1, \dots, i_n \rangle$ of causal relevance in ω are ω -Markov chains.

PROOF: (a) \Rightarrow (b): Suppose (b) is false, i.e. there are i_1, \dots, i_n , and ω such that $\langle i_1, \dots, i_n \rangle$ is a chain of causal relevance in ω , but not an ω -Markov chain. Thus, there is an $s \leq n$ such that $i_{s+1} \perp \{i_1, \dots, i_{s-1}\} / i_s, {}^\omega \langle i_{s+1} - i_1, \dots, i_s \rangle$ does not hold. Several applications of Theorem 2(e) yield that there is an $r < s$ for which $i_{s+1} \perp i_r / \{i_1, \dots, i_{r-1}, i_{r+1}, \dots, i_s\}, {}^\omega \langle i_{s+1} - i_1, \dots, i_s \rangle$ is not true. And this means that there is an $v \in \Omega$ agreeing with ω outside $\{i_1, \dots, i_s\}$ such that $v i_r \xrightarrow{v} v i_{s+1}$.

(b) \Rightarrow (a): Suppose (a) is false, i.e. there are i_1, \dots, i_n , and ω such that $\langle i_1, \dots, i_n \rangle$ is a chain, but not a

strict chain of causal relevance in ω . Thus, we have ${}^\omega i_r \xrightarrow{\omega} {}^\omega i_{s+1}$ for some r and s with $r < s$. This immediately entails that $\langle i_1, \dots, i_n \rangle$ is not an ω -Markov chain. \square

This is a perspicuous theorem. It says that, if causal relevance spreads strictly in stages in all worlds so that in no world there obtain direct as well as indirect causal relations between any states, then all these chains are ω -Markovian, i.e. they have the preferred Markovian property. However, the order of the quantifiers is not the desired one. There is no reason to expect that all possible worlds are causally well-ordered in this way. Hence, what is needed is a universal equivalence rather than an equivalence of universal statements. Here, at last, section 4 comes into play; the assumption that the actual circumstances are ideal will be required in

THEOREM 12. Let i_1, \dots, i_n be binary variables and $\langle i_1, \dots, i_n \rangle$ a chain of causal relevance in ω . Assume that for all $r < n$ $C_\omega^*({}^\omega i_r, {}^\omega i_{r+1}) = C_\omega({}^\omega i_r, {}^\omega i_{r+1})$. Then $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain iff it is a strict chain of causal relevance in ω .

PROOF: (\Rightarrow): Suppose that $\langle i_1, \dots, i_n \rangle$ is not a strict chain of causal relevance in ω , i.e. there exist $r < s$ with ${}^\omega i_r \xrightarrow{\omega} {}^\omega i_{s+1}$. Again, this entails that $\langle i_1, \dots, i_n \rangle$ is not an ω -Markov chain.

(\Leftarrow): Suppose that $\langle i_1, \dots, i_n \rangle$ is a strict chain of causal relevance in ω and consider any $s < n$. Then for all $r < s$ ${}^\omega i_r \xrightarrow{\omega} {}^\omega i_{s+1}$, i.e., since the variables are binary, $i_{s+1} \perp i_r / {}^\omega \langle i_{s+1} - i_r \rangle$. Thus, if $K = \{k \in \{< i_{s+1} - i_s \} \mid i_{s+1} \perp k / {}^\omega \langle i_{s+1} - k \rangle\}$, we have $\{i_1, \dots, i_{s-1}\} \subseteq K$. With the assumption about the circumstances and Theorem 6 we also have $i_{s+1} \perp K / i_s, {}^\omega \langle i_{s+1} - K \cup \{i_s\} \rangle$. And this implies that $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain. \square

Of course, Theorem 12 in particular applies to causal chains. These theorems are mathematically trivial, but conceptually nice; and I do not see how they can be improved upon much. Their content is certainly plausible; intuitively, it is just the existence of direct bypasses to causal chains which violates the Markovian intuition.

Concerning positive relevance, suppose that $A_r = {}^\omega i_r$ ($r = 1, \dots, n$) and that $\langle A_1, \dots, A_n \rangle$ is a causal chain in ω . Somehow, A_1 should then be positively relevant to A_n . But how? If $\langle i_1, \dots, i_n \rangle$ were a Markov chain, Theorem 10 could be applied. But it has turned out that ω -Markov and not Markov chains are the ones relevant

to our enterprise. So if $\langle i_1, \dots, i_n \rangle$ is assumed to be an ω -Markov chain, the trouble is that the characteristic conditional independencies refer for each i_r to a different condition, and therefore Theorem 10 is not immediately applicable. But perhaps the different conditions can be equalized and the grounds for Theorem 10 thus prepared. This is the basic idea which will be worked out in the sequel.

A very simple example illustrates all of the essential aspects of that idea. Suppose that there are only four binary variables $i < j < k < l$, $A = {}^\omega i$, $B = {}^\omega j$, $C = {}^\omega k$, $D = {}^\omega l$, $A \xrightarrow{\omega} B \xrightarrow{\omega} D$, and $\langle i, j, l \rangle$ is an ω -Markov chain. In probabilistic terms this means that

- (1) $P(B|A) > P(B|\bar{A})$,
- (2) $P(D|A \cap B \cap C) > P(D|A \cap \bar{B} \cap C)$, and
- (3) $P(D|A \cap B' \cap C) = P(D|\bar{A} \cap B' \cap C)$ for each $B' \in \{B, \bar{B}\}$.

One would like to infer that, if B is omitted, A is positively relevant to D . Here, C may be taken as given or not. Thus, there are two alternatives for expressing this positive relevance:

- (4a) $P(D|A \cap C) > P(D|\bar{A} \cap C)$, or
- (4b) $P(D|A) > P(D|\bar{A})$.

Obviously, neither follows from (1)–(3), because (2) and (3) conditionalize on C , but (1) does not; that is the trouble with ω -Markov chains. The idea is to equalize the conditions in (1)–(3). Again, there are two alternatives. One may keep (2) and (3) and assume

$$(1a) P(B|A \cap C) > P(B|\bar{A} \cap C)$$

instead of (1); then (4a) can be inferred with the help of Theorem 10. Or one may keep (1) and assume

- (2b) $P(D|A \cap B) > P(D|A \cap \bar{B})$ and
- (3b) $P(D|A \cap B') = P(D|\bar{A} \cap B')$ for each $B' \in \{B, \bar{B}\}$

instead of (2) and (3); then (4b) can be inferred with the help of Theorem 10. However, (1a), (2b), and (3b) do not yet have causal form. The question thus is which causal assumptions allow them to be derived from (1)–(3). The answer differs for the two alternatives.

Look first at (1a). (1a) obviously follows from (1) together with

$$(5a) P(B|A' \cap C) = P(B|A')$$
 for each $A' \in \{A, \bar{A}\}$;

and since P is strictly positive, this is equivalent to

$$(6a) P(C|A' \cap B) = P(C|A' \cap \bar{B}) \text{ for each } A' \in \{A, \bar{A}\}.$$

Thus, (4a) may be derived from (1)–(3) by additionally assuming (6a). And (6a) has causal form; it says that B is directly causally irrelevant to C , whether A obtains or not.

Now consider (2b) and (3b). They obviously follow from (2) and (3), if

$$(7b) P(D|A' \cap B' \cap C) = P(D|A' \cap B' \cap \bar{C}) \text{ for each } A' \in \{A, \bar{A}\} \text{ and } B' \in \{B, \bar{B}\}.$$

And (7b) already has causal form; it says that C is directly causally irrelevant to D , whether A and B obtain or not.

So at least two simple alternative causal conditions are available which guarantee positive relevance of the indirect cause to the indirect effect in this example. When is neither condition satisfied? When and only when B is causally relevant to C in some world and C is causally relevant to D in some world. But in this case there are two paths of causal influence running from A to D (though not necessarily in one world); so it is not surprising that an account of how causal influence is transmitted through single causal chains is inapplicable to such a case.

These observations are valid in general. In the general case we deal with a causal ω -Markov chain $\langle i_1, \dots, i_n \rangle$, where again $A_r = {}^\omega i_r$ ($r = 1, \dots, n$). The strategy of equalizing conditionalization then amounts to finding some set M of variables such that for all $r = 2, \dots, n$ the positive relevance of A_{r-1} to A_r , as well as the characteristic ω -Markov independencies hold also conditional on ${}^\omega M$, thus enabling the inference of the positive relevance of A_1 to A_n conditional on ${}^\omega M$.

Which properties should M be expected to have? Of course, the basic property is that i_2, \dots, i_{n-1} do not belong to M . Two further properties are suggested by the above example. Alternative (b) makes it clear that, if a variable is directly causally relevant to i_r , then it must not be deleted from M . In other words, ${}^\omega M$ has to preserve the circumstances of A_r in some suitable sense (Theorem 14 below will refer to the probabilistically possibly relevant circumstances, Theorem 16 to the ideal ones). The condition that that much information about A_r must be retained is certainly plausible. Now, by omitting i_r from M we delete the most direct information about A_r . But it seems that we must as well delete any indirect information about A_r , which exceeds

the information provided by its circumstances; if such indirect information were retained, the averaging with respect to i_r — which is needed for calculating $P(A_n | A_1 \cap {}^\omega M)$ — may be biased in an undesirable way. This is what emerges from alternative (a) in the above example, in particular from (5a).

In sum, for each $r = 2, \dots, n$, ${}^\omega M$ must include the circumstances of A_r and must not contain any further information about A_r . The fact that ${}^\omega M$ then suits the desired equalization of conditions is the intuitive content of the theorems we are after. They are supplemented by two auxiliary theorems providing a way of expressing the exclusion of such further information in causal terms.

For this purpose let us define:

DEFINITION 10. The variables i and j are *causally connected in ω within $J \subseteq I$* iff there are $k_1, \dots, k_n \in J \cup \{i, j\}$ such that $k_1 = i, k_n = j$, and, for all $r = 1, \dots, n-1$, ${}^\omega k_r \overset{\omega}{\dashv} {}^\omega k_{r+1}$ or ${}^\omega k_{r+1} \overset{\omega}{\dashv} {}^\omega k_r$. And if $i < j$, I call i and j *causally connected in ω* iff they are causally connected in ω within (i, j) .³⁴

Causal unconnectedness in a lot of worlds implies a lot of probabilistic independence, at least if there are no simultaneous variables.³⁵ This is ascertained by the auxiliary

THEOREM 13. Suppose that I is linearly ordered by $<$ and that i and $j > i$ are causally unconnected in all $v \in \omega\{< i\}$. Then there are disjoint K and L such that $K \cup L = [i, j], i \in K, j \in L$, and $K \perp L/\omega\{< i\}$.

This will turn out to be a special case of Theorem 15 below. And it leads to a first result concerning the positive relevance of indirect causes:

THEOREM 14. Suppose that I is linearly ordered by $<$, that $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain of binary variables, $A_r = {}^\omega i_r, x_r = P(A_{r+1} | A_r \cap \omega\{< i_{r+1} - i_r\}) - P(A_{r+1} | \bar{A}_r \cap \omega\{< i_{r+1} - i_r\})$, and that M is a set of variables such that for all $r = 1, \dots, n$:

- (a) $i_r \notin M$,
- (b) $R(A_r) - \{i_{r-1}\} \subseteq M$,
- (c) each $j \in M$ with $j > i_r$ is causally unconnected with i_r in all $v \in \omega(M \cap \{< i_r\})$.

Then $\langle i_1, \dots, i_n \rangle$ is a Markov chain conditional on ${}^\omega M$, and hence $P(A_n | A_1 \cap {}^\omega M) - P(A_n | \bar{A}_1 \cap {}^\omega M) = x_1 \cdot \dots \cdot x_{n-1}$.

The proof is given together with that of Theorem 16. Condition (b) includes variables into M , (c) excludes variables from M . Thus, (b) and (c) work in opposite directions and may be difficult to satisfy. This can be improved upon, but again at the price of assuming circumstances to be ideal. That assumption is used in the auxiliary

THEOREM 15. Suppose that I is linearly ordered by $<$, that $\{< i\} \subseteq N \subseteq \{< j\}$, that, for all $v \in {}^\omega N$ and $k, l \in [i, j]$ with $k < l, C_v^*({}^\omega k, {}^\omega l) = C_v({}^\omega k, {}^\omega l)$, and that i and j are causally unconnected in all $v \in {}^\omega N$. Then there are disjoint K and L such that $K \cup L = [i, j], i \in K, j \in L$, and $K - N, \omega(K \cap N) \perp L - N, \omega(L \cap N) / \omega\{< i\}$.

PROOF: Let $K = \{i\} \cup \{k \in (i, j) | i \text{ and } k \text{ are causally connected within } (i, j) \text{ in some } v \in {}^\omega N\}$ and $L = \{j\} \cup (i, j) - K$. Since causal connectedness within in a fixed set is transitive, this definition implies that each $k \in K$ is causally unconnected with each $l \in L$ within (i, j) in all $v \in {}^\omega N$. In particular we thus have for all $k \in K, l \in L$, and $v \in {}^\omega N$:

- (1) if $k < l$, then $k \notin R_v({}^\omega l)$; and if $l < k$, then $l \notin R_v({}^\omega k)$.

Now we shall inductively work up from i to j . Suppose we have already shown for some $i^* \in (i, j) \cup \{j\}$ that

- (2) $\{< i^*\} \cap K - N, \omega(\{< i^*\} \cap K \cap N) \perp \{< i^*\} \cap L - N, \omega(\{< i^*\} \cap L \cap N) / \omega\{< i\}$.

Let's assume that $i^* \in L$ (for $i^* \in K$ the corresponding reasoning applies). (1) and the ideality of the actual circumstances then imply

- (3) $\{< i^*\} \cap K \perp i^* / \omega\{< i\} \cap v(\{< i^*\} \cap L)$ for all $v \in {}^\omega N$, i.e.
- (4) $\{< i^*\} \cap K \perp i^* / \omega\{< i\} \cap \omega(\{< i^*\} \cap L \cap N), \{< i^*\} \cap L - N$.

(2) and (4) finally yield according to Theorem 1(d)

- (5) $\{< i^*\} \cap K - N, \omega(\{< i^*\} \cap K \cap N) \perp (\{< i^*\} \cup \{i^*\}) \cap L - N, \omega(\{< i^*\} \cap L \cap N) / \omega\{< i\}$.

For $i^* = j$ this is the desired result. Note that for $N = \{< i\}$ Theorem 15 reduces to Theorem 13. In this special case, (3) follows from (1) with the help of Theorem 2(e) alone and without the ideality of circumstances. \square

This leads to the second result concerning positive relevance:

THEOREM 16. Suppose that I is linearly ordered by $<$, that $\langle i_1, \dots, i_n \rangle, A_r$, and x_r are as in Theorem 14, and that M is a set of variables such that for all $r = 1, \dots, n$:

- (a) $i_r \notin M$,
- (b) $R_\omega(A_r) - \{i_{r-1}\} \subseteq M$,
- (c) each $j \in M$ with $j > i_r$ is causally unconnected with i_r in all $v \in {}^\omega M$.

Suppose further that for all $v \in {}^\omega M$ and $k, l \in [i_1, i_n]$ with $k < l$ $C_v^*(v_k, v_l) = C_v(v_k, v_l)$. Then the same result obtains as in Theorem 14.

PROOF: All that must be shown is that for all $r = 2, \dots, n, A' \in \{A_{r-1}, \bar{A}_{r-1}\}$, and $v \in \Omega$:

$$(1) \quad P(A_r | A' \cap \omega\{< i_r - i_1, \dots, i_{r-1}\} \cap v\{i_1, \dots, i_{r-2}\}) = P(A_r | A' \cap {}^\omega M \cap v\{i_1, \dots, i_{r-2}\}) = P(A_r | A' \cap {}^\omega M)$$

(1) says that all the probabilities showing that $\langle i_1, \dots, i_n \rangle$ is a Markov chain conditional on ${}^\omega M$ coincide with the corresponding probabilities showing that $\langle i_1, \dots, i_n \rangle$ is an ω -Markov chain. A direct application of Theorem 10 then leads us to the desired result. To show (1) suppose that $M - \{< i_r\}$ consists of $j_1 < \dots < j_q$ and that we have already shown for some $p - 1 < q$ that

$$(2) \quad P(A_r | A' \cap v\{< i_r - i_{r-1}\} \cap \omega\{j_1, \dots, j_{p-1}\}) = P(A_r | A' \cap v\{< i_r - i_{r-1}\})$$

holds for all $v \in {}^\omega M$. Since i_r and j_p are causally unconnected in all $v \in {}^\omega M$, we may infer from Theorem 15 and the ideality of the actual circumstances that there is a partition $\langle K, L \rangle$ of $[i_r, j_p]$ such that $i_r \in K, j_p \in L$, and for $M' = M \cap \{< j_p\}$

$$(3) \quad K - M', v(K \cap M') \perp L - M', v(L \cap M') / v\{< i_r\}$$

for all $v \in {}^\omega M$. Alternatively, we get (3) from Theorem 13 and the stronger unconnectedness assumption (c) of Theorem 14 without the ideality of circumstances. (3) implies with Theorems 1(d) and 2(c)

$$(4) \quad i_r \perp j_p / v\{< i_r\} \cap v\{j_1, \dots, j_{p-1}\} \text{ for all } v \in {}^\omega M, \text{ i.e.}$$

$$(5) \quad i_r \perp j_p / A' \cap v\{< i_r - i_{r-1}\} \cap \omega\{j_1, \dots, j_{p-1}\} \text{ for all } v \in {}^\omega M.$$

(2) and (5) together imply that (2) holds for p , too, and thus also for q , i.e. that for all $v \in \Omega$

$$(6) \quad P(A_r | A' \cap {}^\omega M \cap v\{< i_r - M \cup \{i_{r-1}\}\}) = P(A_r | A' \cap \omega(M \cap \{< i_r\}) \cap v\{< i_r - M \cup \{i_{r-1}\}\}).$$

Next observe that

$$(7) \quad P(A_r | A' \cap \omega(M \cap \{< i_r\})) = P(A_r | A' \cap \omega(M \cap \{< i_r\}) \cap v\{< i_r - M \cup \{i_{r-1}\}\})$$

holds for all $A' \in \{A_{r-1}, \bar{A}_{r-1}\}$ and $v \in \Omega$ because of the assumption (b) about M and the ideality of circumstances — which is not needed, if, alternatively, the assumption (b) of Theorem 14 is used. (7) says that the R.H.S. of (6) does not depend on v . Thus, the L.H.S. of (6) does not depend on v as well. This finally yields

$$(8) \quad P(A_r | A' \cap {}^\omega M) = P(A_r | A' \cap \omega(M \cap \{< i_r\}) \cap v\{< i_r - M \cup \{i_{r-1}\}\})$$

for all $v \in \Omega$, and this is even somewhat stronger than the desired (1). \square

A careful analysis of the two proofs will show several steps which do not require the full strength of the premises. Thus, there certainly are weaker and maybe nicer conditions under which the consequence of these theorems still holds. But I wonder whether there are much weaker or much nicer conditions subject to the constraint that they be expressed in causal terms.

When does a set M exist as required by Theorem 16 (a)–(c)? I have not found an informative and more perspicuous sufficient condition. But there is a simple necessary condition:

THEOREM 17. A set M satisfying clauses (a)–(c) of Theorem 16 exists only if in all $v \in {}^\omega M$ $\langle i_1, \dots, i_n \rangle$ is the only chain of causal relevance leading from i_1 to i_n .

PROOF: Suppose that for some $v \in {}^\omega M$ there is another chain of causal relevance from i_1 to i_n which may be assumed to be of the form $\langle i_1, j_1, \dots, j_p, i_r, \dots, i_n \rangle$, where $j_p \neq i_{r-1}$. Hence, clause (b) demands that $j_p \in M$. But i_1 and j_p are causally connected in v . Hence, $j_p \notin M$ and thus a contradiction is entailed by clause (c). \square

Since there may be causal connections between earlier members and direct causal antecedents of later members of $\langle i_1, \dots, i_n \rangle$ which are not chains of causal relevance — a situation which again excludes the existence of an appropriate M —, Theorem 17 may not be strengthened to a bi conditional.

Theorem 17 limits the scope of Theorem 16 to cases where causal influence is transmitted through a single causal chain. More powerful theorems are therefore required for dealing with the transmission of causal

influence through more complex causal nets. Eells and Sober (1983), pp. 49ff., should be of help here in a similar way as Theorem 10 has guided Theorem 14 and 16.

But it is clear that this is only the beginning of a much fuller theory of causation. For instance, I have not returned to the very first characterization of causation in section 3 and tried to say what the circumstances of an indirect causal relation are. Theorem 14 and 16 suggest that ${}^{\omega}M$, for a minimal M satisfying (a)–(c) of Theorem 14 or 16, provides such circumstances. However, this suggestion is neither general nor worked out.

Still, I hope to have developed the program far enough to justify the impression that the right direction towards a prosperous theory has been found. In particular, theorems 11, 12, 14, and 16 explain how the three basic intuitions here discussed come to be held, even though they are not generally compatible. This explanation is as plausible as the assumption that the conditions which have been shown to guarantee agreement between the intuitions are taken for granted. And I think this assumption is not too implausible.³⁶

Notes

* I am indebted to the University of California at Irvine for giving me the opportunity to present much of the material during a visiting professorship during the Winter term 1988, to Maria Carla Galavotti for giving me another opportunity at Bagni di Lucca in October 1988, to Nancy Cartwright and Brian Skyrms for discussion and encouragement, and to Karel Lambert and Hans Rott for very carefully checking the manuscript and considerably improving style and content.

¹ This is precisely how Lewis (1973b) proceeds.

² For instance by Cartwright (1979), Giere (1980), and all those who take probability in causal contexts as a statistical property of event types or classes or the like. However, Cartwright herself attacks the reverse strategy in her (1988). See also Davis (1988) for a discussion of this point.

³ Compare Theorem 2 below with section 6 of Spohn (1988).

⁴ Cf. section 7 of Spohn (1988).

⁵ This means that \leq is transitive and complete.

⁶ In (1980) I allowed for simultaneous causation in a way which preserved continuity with the restricted case. I am not sure whether the same procedure would work here.

⁷ What is discussed with respect to more-than-two-valued variables is usually only causal relevance simpliciter and not positive or negative causal relevance. An exception is Suppes (1970), pp. 60ff., but it has not been further discussed, as far as I know.

⁸ Within the theory of OCFs there is no such peculiar law and thus no technical difference between binary and other variables.

⁹ This is clearly displayed by Otte (1981) who criticized Suppes (1970) essentially on this account. I have argued in (1980), pp. 92f., that the trouble-maker is essentially the fact that in standard probability theory there are no conditional probabilities for conditions having probability 0. The problem evaporates in the unification mentioned at the end of the introduction.

¹⁰ Cf., e.g., Salmon (1988a) who argues that propensities are best understood as probabilistic causes and that other objective probabilities are derived from propensities.

¹¹ Here, I refer to Lewis (1980) and Skyrms (1984), ch. 3; see also Spohn (1987).

¹² The most eloquent at present is Putnam who repeatedly argues against a naturalistic conception of causation, e.g. in (1983), ch. 12.

¹³ The hyphen denotes set theoretic difference.

¹⁴ I choose this notation because the restricted domain needs to be more salient than the path itself.

¹⁵ Theorem 2(f) does not hold for ordinal conditional functions. If their range is restricted to natural numbers, they satisfy the laws (a)–(e) without further qualification. Cf. Spohn (1988), sect. 6.

¹⁶ Studeny (1989) and Geiger, Pearl (1988), sect. 6, mention further properties, and there are still more.

¹⁷ This is so because causes in the intuitive sense are partial causes as opposed to total causes and because I want to account directly for causes without considering total causes. Particularly in the context of deductive-nomological explanation philosophers have been attracted by the idea that, conversely, the notion of a total cause is the central one which has to be explicated first. This strategy, I think, has been rejected for good reasons.

¹⁸ \bar{A} denotes the complement of A relative to Ω .

¹⁹ Cf. Good (1961), pp. 308f., and (1988), p. 27.

²⁰ In Good (1988), he only argues on p. 27 that his way of conditionalization yields the desired result that a falling barometric reading has no tendency at all to cause a storm. But this result may already be obtained by conditionalization with respect to the past of the spurious cause; no reference to laws of nature is required for this example.

²¹ Thus, even if one player chooses first, the other does not know. Though many examples have the same structure, the irrelevance of the temporal order seemed to me to be particularly perspicuous in this game-theoretic case.

²² Good (1961) indeed makes a similar assumption on p. 45 when he requires neighbors in causal chains to be contiguous in space and time.

²³ It is now "circumstances", not "relevance" which is taken in its widest sense.

²⁴ Equivalently we may say in Skyrms' terms (1980, part IA) that $C'_{\omega}(A, B)$ makes the probability of B given A or \bar{A} maximally resilient over the rest of the past of B .

²⁵ This identity follows from the fact that for $k \in R_{\omega}(B) \cap \{D \mid D \text{ is a } k\text{-state and } D \overset{\omega}{\diamond} B\} = {}^{\omega}k$.

²⁶ Cf. Lewis (1973a), pp. 32–34.

²⁷ One might well find this idea and thus probabilistic causation unintelligible; many have done so. But in the light of the last remark in the introduction the present discussion should be illuminating for them, too.

²⁸ Eells, Sober (1983) take up another remark of Suppes on that

page and investigate under which special circumstances transitivity of probabilistic causation is preserved.

²⁹ Cf., e.g., Lauritzen (1982).

³⁰ The theory of Markov fields is indeed utterly illuminating for the causal theorist; cf., e.g., Kiiveri, Speed, Carlin (1984) or the rich material presented in Pearl (1988), ch. 3.

³¹ For a proof cf., e.g., Good (1980). Eells, Sober (1983), pp. 49ff., prove a more general result about the propagation of positive relevance in particular Markov nets. Theorem 10 shows, by the way, that the meteorological example for the fading of probabilistic causal influence and thus for the failure of transitivity is not really convincing. If the meteorological models would explain the weather as a Markov process (I doubt that they actually do), then the theorem says that even the weather at the turn of the last century makes a probabilistic difference for today's weather, though an almost infinitesimally small one.

³² I owe this point to Karel Lambert; it became really clear to me in a long conversation with him.

³³ Good (1961), p. 311, Axiom 10, and Humphreys (1980), pp. 308f., seem to be guided by the same conception.

³⁴ This restriction of the connecting sequence to the interval between i and j does certainly not conform to the standard usage of "causally connected", if there is any. Note that despite this restriction causal connectedness is still weaker than causal relevance. A chain of causal relevance from i to j is always future-oriented, whereas a sequence causally connecting i and j may arbitrarily change its temporal direction.

³⁵ This is rather a technical restriction needed in the proofs below. However, it is not obvious how to avoid it.

³⁶ In fact, I believe that there is a deeper explanation for the conditions that actual circumstances are ideal, that causal chains are strict, and the like; these conditions are crucial to an objectivization of our causal picture. This conjecture would emerge more clearly in the unwritten deterministic counterpart of this paper; but it certainly cannot be part of this inquiry.

References

- Cartwright, N.: 1979, 'Causal Laws and Effective Strategies', *Noûs* **13**, 419–437; also in: N. Cartwright, *How the Laws of Physics Lie*, Clarendon Press, Oxford 1983, 21–43.
- Cartwright, N.: 1988, 'Regular Associations and Singular Causes', in: Skyrms, Harper (1988), 79–97.
- Davis, W. A.: 1988, 'Probabilistic Theories of Causation', in: Fetzer (1988), 133–160.
- Dawid, A. P.: 1979, 'Conditional Independence in Statistical Theory', *Journal of the Royal Statistical Society*, series B, **41**, 1–31.
- Eells, E., E. Sober: 1983, 'Probabilistic Causality and the Question of Transitivity', *Philosophy of Science* **50**, 35–57.
- Fetzer, J. H. (ed.): 1988, *Probability and Causality*, Reidel, Dordrecht 1988.
- Geiger, D., J. Pearl: 1988, 'Logical and Algorithmic Properties of Conditional Independence and Qualitative Independence', *Technical Report No. R-97-III*, Cognitive Systems Laboratory, University of California, Los Angeles, Oct. 1988.
- Giere, R.: 1980, 'Causal Systems and Statistical Hypotheses', in: L. J. Cohen, M. Hesse (eds.), *Applications of Inductive Logic*, Clarendon Press, Oxford, 251–270.
- Good, I. J.: 1961, 'A Causal Calculus', *British Journal for the Philosophy of Science* **11**, 305–318, **12**, 43–51, and **13**, 88.
- Good, I. J.: 1980, 'A Further Comment on Probabilistic Causality: Mending the Chain', *Pacific Philosophical Quarterly* **61**, 452–454.
- Good, I. J.: 1988, 'Causal Tendency: A Review', in Skyrms, Harper (1988), 23–50.
- Hart, H. L. A., T. Honoré: 1959, *Causation in the Law*, 2nd ed., Clarendon Press, Oxford 1985.
- Hesslow, G.: 1976, 'Two Notes on the Probabilistic Approach to Causality', *Philosophy of Science* **43**, 290–292.
- Humphreys, P.: 1980, 'Cutting the Causal Chain', *Pacific Philosophical Quarterly* **61**, 305–314.
- Kiiveri, H., T. P. Speed, J. B. Carlin: 1984, 'Recursive Causal Models', *Journal of the Australian Mathematical Society (Series A)* **36**, 30–52.
- Lauritzen, S. L.: 1982, *Lectures on Contingency Tables*, 2nd. ed., University Press, Aalborg, Denmark.
- Lewis, D.: 1973a, *Counterfactuals*, Blackwell, Oxford.
- Lewis, D.: 1973b, 'Causation', *Journal of Philosophy* **70**, 556–567; also in Lewis (1986), 159–172.
- Lewis, D.: 1980, 'A Subjectivist's Guide to Objective Chance', in: R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol. II, University of California Press, Berkeley, 263–293; also in Lewis (1986), 83–113.
- Lewis, D.: 1986, *Philosophical Papers*, vol. II, University Press, Oxford.
- Mellor, D. H.: 1988, 'On Raising the Chances of Effects', in: Fetzer (1988), 229–239.
- Otte, R.: 1981, 'A Critique of Suppes' Theory of Probabilistic Causality', *Synthese* **48**, 167–189.
- Otte, R.: 1985, 'Probabilistic Causality and Simpson's Paradox', *Philosophy of Science* **52**, 110–125.
- Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems*, Kaufmann, San Mateo, Ca.
- Putnam, H.: 1983, *Philosophical Papers*, vol. 3, University Press, Cambridge.
- Salmon, W. C.: 1988a, 'Dynamic Rationality: Propensity, Probability, and Credence', in: Fetzer (1988), 3–40.
- Salmon, W. C.: 1988b, 'Intuitions — Good and Not-So-Good', in: Skyrms, Harper (1988), 51–71.
- Skyrms, B.: 1980, *Causal Necessity*, Yale University Press, New Haven.
- Skyrms, B.: 1984, *Pragmatics and Empiricism*, Yale University Press, New Haven.
- Skyrms, B., W. L. Harper (eds.): 1988, *Causation, Chance, and Credence*, Kluwer, Dordrecht.
- Spohn, W.: 1980, 'Stochastic Independence, Causal Independence, and Shieldability', *Journal of Philosophical Logic* **9**, 73–99.
- Spohn, W.: 1983a, 'Deterministic and Probabilistic Reasons and Causes', *Erkenntnis* **19**, 371–396.
- Spohn, W.: 1983b, 'Probabilistic Causality: From Hume via Suppes to Granger', in: M. C. Galavotti, G. Gambetta (eds.), *Causalità e Modelli Probabilistici*, Editrice CLUEB, Bologna, 69–87.

- Spohn, W.: 1987, 'A Brief Remark on the Problem of Interpreting Probability Objectively', *Erkenntnis* **26**, 329–334.
- Spohn, W.: 1988, 'Ordinal Conditional Functions. A Dynamic Theory of Epistemic States', in: W. L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, Kluwer, Dordrecht, 105–134.
- Studeny, M.: 1989, 'Multiinformation and the Problem of Characterization of Conditional Independence Relations', *Problems of Control and Information Theory* **18**, 3–16.
- Suppes, P.: 1970, *A Probabilistic Theory of Causality*, North-Holland, Amsterdam.
- Suppes, P.: 1984, *Probabilistic Metaphysics*, Blackwell, Oxford.

*Institut für Philosophie
Universität Regensburg
D-8400 Regensburg, F.R.G.*