FRANZ VON KUTSCHERA

CAUSATION

ABSTRACT. As cause we often specify an event the occurrence of which first guaranteed that of the effect. This notion is explicated in a framework of branching worlds in Sections I to V. VI and VII point out its close relations to the concept of an agent's bringing about an event. The topic of the last two sections is the distinction between causes and necessary circumstances. For this purpose conditionals are used, interpreted with respect to branching worlds without a similarity relation between them.

No simple definition of causation can cover the complex use of causal terms in everyday discourse. The best we can do is try to disentangle the different strands in this usage and define different types of causation. When I propose yet another definition, radically departing in some points from most of the current ones, it should be kept in mind, therefore, that there is no claim attached that it captures the 'true' concept of causation. My aim is just to explicate one of the notions underlying our usage: A cause is an event, which was not sure to have occurred and whose occurrence first guaranteed that of the effect. This description can also be understood epistemically, but I take it in a purely factual sense here. The effect, therefore, is conceived of as a necessary consequence of an event which in turn didn't occur necessarily. The concept of necessity employed is not a logical or nomological one. Necessary is rather what is the case, no matter what turn the future history of the world will take. The frame of my analyses are worlds with branching histories. The radical difference from most of the current concepts of causation can be seen from the consequence that in deterministic, unbranching worlds there are no causes of this type, that effects are not causes and therefore no causal chains occur. It will be shown that this concept of causing is closely related to that of bringing about by an agent. Hence it can be termed "actionistic" – although only in a very wide sense of this word. Its closest relative is the type of causation that G. H. von Wright has

Journal of Philosophical Logic **22**: 563–588, 1993.

analysed in (1974). Finally I show how to define truth conditions for conditionals in this framework and how to use them to distinguish causes from necessary circumstances.

I. TREE-UNIVERSES

In what follows I shall not specify a language in which causal sentences can be expressed, but only talk about models for such a language. The analyses, therefore, are confined to the objective level. I shall also use only very simple models with a discrete time order. Many applications demand generalisations, but here I just want to present the basic ideas as simply as possible and free from complications. The model structure employed has – for a richer time order – already been defined in (1986), but as this paper is in German, I shall give a short sketch of it here again.

Worlds are conceived of as functions from timepoints into world-states (WS, for short) so that $w(t)$ is the state of the world w at time t. For simplicity's sake it is assumed that all worlds have a beginning in time and last equally long. The set T of times, therefore, is the set of natural numbers or an initial segment thereof. The worlds of a universe are the branches of one or more trees. Hence every WS $w(t)$ with the exception of the initial states has exactly one immediate predecessor $w(t - 1)$, different from it, and two worlds w and w' that are in the same state in t, share their history before t. A WS, then, cannot be the product of different developments, of different world histories. Tree-universes are chosen because we are interested in a concept of possibility according to which only what is real or realizable is possible. What is past or present cannot be changed anymore, so it is possible only if real. For an epistemic concept of possibility tree structures would not be appropriate, as v. Wright remarked, for as far as we know the present WS can very well be the product of different developments.

We can also define tree-universes starting with a set I of WS and a relation r of immediate succession on I, instead of sets of times and worlds.

D1: A *tree-universe* is a pair $\mathfrak{U} = \langle I, r \rangle$ such that

(1) I is a set of WS.

(2) r is a binary relation on I such that for all $i, j \in I$ we have

(a) $I_0 := \{j: \neg \bigvee i(irj)\} \neq \Lambda$,

(b) $\bigwedge i \bigvee j(j \in I_0 \wedge jr^{\geqslant 0} i)$,

(c) $irj \wedge krj \supset i = k$,

(d) for all natural numbers m and n with $m < n$:

$$\bigvee ik(i \in I_0 \wedge ir^n k) \supset \bigwedge ik(i \in I_0 \wedge ir^m k \supset \bigvee j(krj)).$$

In (a) I_0 is the set of initial states and Λ the empty set. In (b) $r^{\geqslant 0}$ is the ancestral of r, i.e. $ir^{\geqslant 0}j$ holds iff there is a $n \geqslant 0$ such that $ir^n j$, where these relative powers are defined by $ir^0 j := i = j$ and $ir^{n+1}j := \bigvee k(ir^n k \wedge krj)$. (b), therefore, implies that every WS in I is an initial state or the (immediate or mediate) successor of an initial state. According to (c) every WS has only one predecessor, and in view of D3 (d) says that all worlds are equally long. From (a) to (c) it follows that for every WS i there is exactly one number n for which $\bigvee j(j \in I_0 \wedge jr^n i)$ holds. This number will be designated by $z(i)$. We can then define the sets T of time points and W of worlds as follows:

D2: $T := \{n: \bigvee i(z(i) = n)\}$.
Times will be designated by $t, t', \ldots$.

D3: $W := \{w \in I^T: \bigwedge t(w(t)rw(t+1) \wedge w(0) \in I_0)\}$.
Worlds, then, are functions from T into I, such that immediately succeeding WS are related by r, and $w(0)$ is an initial state. We then have $z(w(t)) = t$. Worlds could also be defined as maximal sets of WS on which $r^{\geqslant 0}$ is a linear ordering.

II. EVENTS

If a language is interpreted on tree-universes sentences will be assigned truth values that are dependent on worlds and times, not just on WS. The truth value of a sentence about the future, e.g., does not just depend on the present state of the world, but on its further development. Propositions as intensions of sentences, then, are subsets of $W \times T$, i.e. sets of pairs of worlds and times. Instead of "propositions" we will talk of *states of affairs* (SAs, for short). In connexion with causal relations we are interested mainly in *events*. In every world, in which it occurs, an event has a well-defined beginning and a well-defined ending. Every event occurs at most once in every world. Let $\tau, \tau', \ldots$ be time intervals $[t_1, t_2]$ with $t_1 < t_2$. Then an event may be defined as a set E of segments w_τ of worlds. w_τ is the set of all WSs $w(t)$ with $t \in \tau$. If τ_1 is the first and τ_2 the last point of τ and we have $w_\tau \in E$ and $w'(\tau_2) = w(\tau_2)$ then w'_τ is in E, too. We assume that the same event does not start in two worlds at different times as long as they coincide. For if they share a common history until the event has started in both of them, there is no reason for saying that it starts later in one than in the other. We do not assume, however, that if $w_\tau \in E$ and there is a t with $\tau_1 \leqslant t < \tau_2$ and $w'(t) = w(t)$, then there is an interval τ' with $\tau'_1 = \tau_1$ such that $w'_{\tau'}$ is in E, too. If Joe climbs a mountain in w_τ and w' is exactly as w up to t, it does not follow, that Joe climbs the mountain in w', too, for in w' he may get tired and return before he gets to the top. We arrive then at the definition:

D4: An *event* is a set E of segments of worlds w_τ such that
(a) $w_\tau, w_{\tau'} \in E \supset \tau = \tau'$,
(b) $w_\tau, w'_{\tau'} \in E \wedge w_\tau \cap w'_{\tau'} \neq \Lambda \supset \tau_1 = \tau'_1$.

There is a unique correspondence between eventlike SAs and events. If we set, for $X \subset W \times T$,

D5: $L(X, w, \tau) := \bigwedge t(\langle w, t\rangle \in X \equiv t \in \tau)$ – X holds in w exactly in τ, we can say

D6: A SA X is *eventlike* iff $\bigwedge w(\bigvee t(\langle w, t\rangle \in X) \supset \bigvee \tau L(X, w, \tau)) \wedge \bigwedge ww'\tau(L(X, w, \tau) \wedge w_\tau = w'_\tau \supset L(X, w', \tau)) \wedge \bigwedge ww'\tau\tau'(L(X, w, \tau) \wedge L(X, w', \tau') \wedge w_\tau \cap w'_{\tau'} \neq \Lambda \supset \tau_1 = \tau'_1)$.

If X is eventlike $\{w_\tau: L(X, w, \tau)\}$ is the corresponding event, and the eventlike SA $\{\langle w, t\rangle: \bigvee\tau(w_\tau \in E \wedge t \in \tau)\}$ corresponds to the event E.

We can specify events "abstractly" by sentences and speak, for instance, of the event, that peak K2 is climbed for the first time. We can also designate them "concretely" by (perfect) nominalizations like

"The first climbing of K2". The former event occurs in all worlds in which K2 is climbed for the first time – no matter by whom and under what circumstances. The latter event, however, is the first climbing as it actually occurred. Hence it occurs in another world only if it comes to pass there in exactly the same way as in our world. "Exactly", of course, is rather strong. Mostly the context specifies – more or less clearly – certain aspects under which the event is considered. The first climbing of K2, under the aspects of type A, then is the event that occurs in a world iff K2 there is climbed for the first time in a way that under all A-aspects is equal to the way it is actually climbed in our world. "Concrete" events, however, are also sets of world segments corresponding to eventlike SAs. Events are discussed in more detail in Kutschera (1993).

III. NECESSITY

If only eternal sentences are taken into account, as in most logics, the necessity of a proposition depends just on worlds. In our frame the notions of analytical and nomological necessity would have to be defined by

$$\Box(X) := \bigwedge wt(\langle w, t\rangle \in X) \text{ – } X \text{ is } \textit{analytically necessary},$$

$$L(w, X) := \bigwedge w'(wRw' \supset \bigwedge t\,(\langle w', t\rangle \in X)) \text{ – In } w\ X \text{ is } \textit{nomologically necessary}.$$

Here wRw' says that in w' the same natural laws hold as in w. We can, however, also introduce a timedependent necessity. First we define

D7: (a) $X_t := \{w: \langle w, t\rangle \in X\}$ – The SA that X holds in t,[1]
(b) $W^i := \{w: w(z(i)) = i\}$ – The set of all the worlds passing through WS i.

We can express a timedependent necessity then by

D8: $N(w, t, X) := W^{w(t)} \subset X_t$.

The SA X, therefore, is necessary from the standpoint of WS $w(t)$ iff X holds in t in all worlds coinciding with w in t. We then have

$\Box(X) \equiv \bigwedge w'tN(w', t, X)$ and $L(w, X) \equiv \bigwedge w'(wRw' \supset \bigwedge tN(w', t, X))$. For events E we get: $N(w, t, E) \equiv W^{w(t)} \subset \{w'\colon V\tau(w'_\tau \in E \wedge t \in \tau)\}$.

A world is *deterministic* if it does not contain a branching point. Every event occurring in such a world occurs necessarily from the beginning and truth coincides with necessity. A universe is deterministic if all its worlds are deterministic.

IV. CAUSES

Causal relations are relations between events. That E is a cause of E' is true in a world always or never, i.e. the relation does not depend on times. We can then write it in the form $K(w, E, E')$ – in world w E causes E'. Our basic idea for determining this relation is: E is a cause of E' if the occurrence of E is something that first guarantees the occurrence of E'. This idea must now be made precise in our frame of tree-universes. If we use the abbreviations:

D9: (a) $E^0 := \{w\colon V\tau(w_\tau \in E)\}$ – The SA that E occurs.
(b) $D(E, w, t) := W^{w(t)} \subset E^0$ – In w, t E is *determined.*
(c) $DB(E, w) := V\tau(w_\tau \in E \wedge D(E, w, \tau_1))$ – in w E is *determined from its beginning,*

our explication takes the form:

D10: $K(w, E, E') := V\tau(w_\tau \in E \wedge \bigwedge w'\tau'(w' \in W^{w(\tau_1)} \wedge w'_{\tau'} \in E \supset DB(E', w') \wedge \neg D(E', w', \tau'_1)))$.

It is based on the following considerations:
(1) A cause in w is an event occurring in w. Hence there has to be an interval τ with $w_\tau \in E$.
(2) That the occurrence of E' is not guaranteed until E occurs means firstly that it is certain that E' will occur if E occurs. "Certain", here, can have an epistemic as well as a factual sense. For a person the occurrence of an event is doxastically certain if he is convinced that it will occur. Factual or objective certainty, on the other hand, is an alethic modality. Since we want to give a purely objective sense to causal statements we take certainty as timedependent necessity, referring it to the beginning of E in w, i.e. to τ_1. This means that we take

the circumstances obtaining in w and τ_1 into account. That gives us

$$(*) \qquad \bigwedge w'(w' \in W^{w(\tau_1)} \wedge V\tau'(w'_{\tau'} \in E) \supset V\tau''(w'_{\tau''} \in E')),$$
$$\text{or } W^{w(\tau_1)} \cap E^0 \subset E'^0.$$

Since $V\tau''(w'_{\tau''} \in E')$ follows from $DB(E', w')$, $(*)$ is implied by the second conjunct of **D10**. It is a version of the traditional idea of a necessary connection between cause and effect. Since Hume has shown that no logical necessity is involved, today a nomological one is assumed, if this idea is still accepted. Our necessity is neither a logical nor a nomological one, both being independent of time. $(*)$ just says: In view of the circumstances obtaining in w and τ_1, E' must occur if E does, no matter how the world goes on. If we say that E causes E' we mostly do not want to imply that events of the type of E will always be followed by events of the type of E', much less still that this regularity holds as a matter of natural law. Eve's slipping on a banana skin was the cause of her breaking a leg, but nobody assumes that everyone's slip on a banana skin results in a broken leg. Rather it were the special circumstances that, in Eve's case, made her slip result in a broken leg.

This example might suggest that for E to be a cause of E' in w it is only necessary that an occurrence of E under the circumstances obtaining in w at τ_1, the beginning of E in w, results in E'. The explosion of a bomb is the cause of John's death even if the bomb, if it had exploded later on, when John would have been somewhere else, would not have killed him. Therefore it might be argued that only worlds have to be considered, in which E also begins at τ_1. We have to distinguish, however, between the concrete event 'The explosion of the bomb' and the abstract event, that the bomb exploded. The first one occurs at a specific time and with John close to it. The second, on the other hand, can occur at a different time in a different world and with John far away. So the first may be the cause of John's death while the second is not. In our general definition, then, we do not want to refer only to occurrences of E that start in τ_1.

(3) That the occurrence of E' was not guaranteed until E occurred means, secondly, that in w and τ_1 it is not yet certain, i.e. necessary, that E' occur. This is $\neg D(E', w, \tau_1)$, and in view of $(*)$ it implies

$\neg D(E, w, \tau_1)$, or $\neg DB(E, w)$. The intuitive idea from which we started was that causes are events that do not have to happen, events which are not determined before they begin. As events occur in the transition from WS to other WS we can also say, that causes are not yet determined at the moment of their beginning.

D10, however, demands not only $\neg D(E', w, \tau_1)$ but $\neg D(E', w', \tau_1')$ for all worlds w' from $W^{w(\tau_1)}$ and intervals τ' with $w'_{\tau'} \in E$. This is because we want to have

(**) $K(w, E, E') \wedge w_\tau \in E \supset \bigwedge w'(w' \in W^{w(\tau_1)} \cap E^0 \supset K(w', E, E'))$ – If E causes E' in w, then it is necessary from the start of E in w that all possible occurrences of E cause E'.

This is not a direct consequence of our general idea of causes and their effects, but it seems intuitively appropriate to say that it is fixed from the beginning of E in w that E cause E' whatever turn the history will take.

(4) We interpret the idea that the effect is guaranteed to occur, given the cause, in the strong sense that E' is determined from its very beginning. That would be a consequence of (*) if we had postulated that the effect begins not before the end of the cause, but this is not implied by **D10**. A chance event is a typical example of an event that is not determined from its beginning, and it should not be an effect. But so far it would be in accordance with our stipulations to say: The event that a toss of a coin results in "heads", is the effect of the event, that Joe takes the coin in hand and tosses "heads" with it. To exclude cases like that we demand that effects be determined from their very beginning. That gives us $DB(E', w)$. Even this is a very strong assumption. Firstly, together with $\neg DB(E, w)$, it implies that effects are never causes. Secondly, heating a gas does not cause it – in the sense of **D10** – to expand. To capture cases like that we have to introduce piecemeal causation in the sense of David Lewis: E causes E' piecemeal if there are temporal parts $E_1, \ldots, E_n$ of E and $E_1', \ldots, E_n'$ of E' such that E_i causes E_i' in the sense of **D10** $(1 \leqslant i \leqslant n)$. Piecemeal causation has to be used anyway for an analysis of cases of self-causation[2] with an asymmetrical causal

relation, but we have to refer to it even in situations where other theories get along without it.

If our concept of causation is to satisfy condition (**), we again have to strengthen the postulate $DB(E', w)$ to $DB(E', w')$ as in **D10**. From $DB(E', w')$ and $\neg D(E', w', \tau_1')$ we obtain $\tau_1' \leqslant \tau_1''$, so that E' starts later than E in all the worlds from $W^{w(\tau_1)}$. This is not very problematic. Examples of alleged effects that precede their causes are not very convincing,[3] and physics, at least, assumes that effects are propagated with finite velocities. It is not a consequence of **D10**, however, that the effect does not start before the cause is completed. Since we have to use piecemeal causation anyway even this would not be unacceptable, and we might replace condition $DB(E', w')$ by the stronger $\tau_2' \leqslant \tau_1''$. Then we could not say that the egg I have for breakfast is hard because it was boiled for 15 minutes, since it was already hard after 10 minutes, but we might also say that only the 10-minutes cooking was the cause of the egg's being hard. Since our definition yields a rather exclusive concept of causation as it is, we do not want to draw its boundaries still closer, however. Neither do we demand that the effect is not determined earlier than the cause – that it is not determined later is a consequence of (*). Cases of causal overdetermination are to be admitted, but in them another cause, and thereby the common effect, may be determined before E.

All this shows that there is a certain latitude in explicating the basic notion of causation, from which we started. We might, furthermore, refer the necessary connection of cause and effect not to the beginning of the cause, but to its point of determination, i.e. the last time when it was still possible that the cause would not occur. And we could add to **D10** the condition that the occurrence of E' is not a logical consequence of the occurrence of E. Even if it is not implied by $E^0 \subset E'^0$ that E, if it occurs, is a cause of E', at least the event that E' is going to occur should not be a cause of E'. Let E' be an event that does not start in any world at $t = 0$, that is not determined in w, 0, but is determined from its beginning in all the worlds from $W^{w(0)}$ where it occurs. Then $F^*(E') := \{w_\tau : \tau_1 = 0 \wedge V\tau'(w_{\tau'} \in E' \wedge \tau_2 = \tau_1' - 1)\}$ is the event that E' will begin, and it is a cause of E' according to **D10**. But as this is easily remedied and we are only after a simple model for our intuitive notion of causation that yields a first basis for discussing it, we shall leave such modifications aside, here.

Our relation of causing is asymmetrical, as any respectable causal relation should be; effects are determined from their beginning while causes are not. Therefore the relation is transitive only in trivial sense: effects are never causes.

If E_{w} is the concrete event E, as it is realized in w (under certain aspects), E_{w} is a proper subset of E. Now $K(w, E, E')$ implies $K(w, E_{w}, E')$, but neither does the converse hold nor the implication from $K(w, E, E')$ to $K(w, E, E'_{w})$ or vice versa. On the object level this raises no difficulties as the sets E and E_{w} are patently different. On the level of normal language, however, one has to be careful since there the difference between sentences "The fact that A causes the fact that B" and those of the form "Event e causes event e'", where "e" and "e'" are names of events, is not always clearly marked. A name for an event signifies this event as it happens in the world from which the causal relation is considered. If A and B express eventlike SAs corresponding to the events E and E' and if "e" and "e'" are normalization of A and B, then the two sentences are related as $K(w, E, E')$ to $K(w, E_{w}, E'_{w})$. In the example we used above: The fact that the bomb exploded, is not the cause of the fact that John died afterwards, even though the explosion of the bomb was the cause of his death. If it was necessary from the start that the bomb should explode sooner or later, the fact that it exploded, is not even a possible cause, but the explosion of the bomb is, if it was not certain from the start that it should explode exactly when it actually did.

V. DISCUSSION

In this section some objections against our concept of causation will be considered. The main objection, of course, is that for many people the succession of events in deterministic worlds is *the* paradigm for causal connections, while in such worlds there are no causes at all in our sense. Now the real world, as far as we know, is indeterministic. Aside from physical theories our normal conceptions of human action and responsibility preclude a deterministic view, too. Whether true or not, for their reconstruction at least we have to assume indeterministic worlds. So there is no lack of applications for our notion. On the other hand we cannot claim that it covers all the intuitions that govern our talk of causes in everyday life and in science. We already remarked that doxastic aspects may play a role. As cause often an event is pointed out, the occurrence of which was unknown or unusual, and which makes it more plausible that the effect came about. This idea is taken up in the probability theory of causation, if – following Hume – one thinks of subjective probabilities. The central idea of regularity theories of causation to take effects as nomological consequences of their causes is also justified intuitively. We say, for instance, that it was the fact that the temperature sank below the freezing point that caused the water pipe's bursting. This would not be a cause in our sense since the sinking of the temperature was the effect of preceding meteorological events. Objections against theories of causation often wrongly presuppose that there is just one single type of causation, and the theories, in turn, invite such objections by sharing this presupposition and claiming this one and only type to be that which they define. But even if it be granted that our definition is only concerned with one type of causation, that does not save it from all objections, of course. The notion might, after all, be incoherent or yield contra-intuitive results even in the restricted range of its applicability.

(1) *No causal chains*

Since effects are not causes there are no causal chains. But then it seems problematic to admit long intervals between the end of the cause and the beginning of the effect, as we have done in **D10**. Causation at a temporal distance is only conceivable, one might say, as mediate causation through causal chains in which every cause is contiguous with its direct effect. Without contiguity the cause alone cannot explain the effect. First, too much may happen in-between. Second, there must be a reason why the effect begins exactly at the time it does and neither earlier nor later. Since the cause is already completed it cannot explain this. Now, the first doubt is eliminated by postulating a necessary connexion between cause and effect; as long as something can still happen that would preempt the efficacy of the cause there is no necessary connection. The second doubt is more serious. If E is the cause of E', E is not thereby also the cause of E's occurring at a certain time – that would be a more specific event than E'. There may be reasons, however, that are not causes, for a later beginning of the effect. According to some natural law the effect may take a certain time – calculable from the law – to materialize after the cause is completed. I bump against a flower pot on my window sill. Some time after the bump, depending on the height, it will crash on the ground. The assumption of a mediating chain of contiguous causes and effects seems no less artificial than Aristotle's assumption of the air pushing the stone I have thrown to account for its moving on after it has left my hand.

(2) *Necessary causes*

We cannot read $K(w, E, E')$ as "In w E is *the* cause of E'". Both $K(w, E_1, E')$ and $K(w, E_2, E')$ may be true without E_1 and E_2 being identical. The two causes may not only be logically but factually independent. There would be a logical dependence for $E_1 \subset E_2$ or $E_2 \subset E_1$, and a factual dependence if in the beginning of E_1 it would be necessary for E_2 to happen if E_1 occurs, or vice versa. There must be an overlap between E_1 and E_2 in w, however, since E' is determined after one of the causes is completed. Causes, according to **D10**, then, are sufficient, but not necessary conditions for the effect. It is possible (in the beginning of a cause) that the effect occurs without the cause.

In the literature "necessary causes" play an important role, as events without which the effect would not have taken place. Paresis, to use a well-worn example, only occurs after an infection with syphillis, but even then only in about 5% of the cases. Nevertheless, we say that paresis is caused by an infection with syphillis. Since here the infection is not sufficient but only necessary for paresis to occur, it is no cause in our sense. But we certainly do not call every necessary condition of an event its "cause" either, otherwise we should have to say that a cause of Joe's motor accident was that he left his house in the morning. The difference between the two examples seems to be that in the first case nothing extraordinary has to happen between the infection and the development of paresis. The unusual thing is the infection, not the consequence, given the infection. In the second case, however, Joe's leaving his house in the morning is not the remarkable thing but, let us say, someone overlooking a red light. The same thing holds in another standard example: I give Mary a box of poisoned chocolates. She eats one and dies. Here, too, my action is not sufficient for her death – after all she might have not have eaten from the chocolates – it is just necessary for it. Still we say that my action was the cause of her death since what she did was only to be expected. Causes, it seems, are always sufficient for the effect, but often they are sufficient in a weak epistemic sense only; they make the occurrence of the effect more probable.[4] "Normal" and "usual" are typical doxastic concepts, for what we take to be normal under given circumstances is what we have come to expect. Since we are interested in a purely objective concept of causation such "necessary causes" are beyond our scope.

(3) *Causal preemption*

John jumps out of a window on the 12th floor of a building (E), which is surrounded by concrete so that it is certain that he will be dead after hitting the ground (E'). But while he flies past the 8th floor he is shot from a window (E''), so that he is in fact already dead before he reaches the ground. Here we should say that the cause of John's death was not E but E''. But since we have assumed that with E the occurrence of E' is already guaranteed, we should have to say that E, and not E'', was the cause of E'. Now, evidently, it makes a difference whether we speak of the concrete event of John's death or of E', i.e. the fact that John is dead after hitting the ground. John's death occurs earlier and is a death by shooting, and its cause, therefore, is not E but E''. David Lewis has objected against too concrete ("fragile") descriptions of events as a way out of such problems,[5] but where we have to look for a cause doubtlessly depends on the effect, and in our case there are clearly two different effects. E is the cause of E', and the shot from the 8th floor window made no difference to this effect. But E'' is the cause of John's death, and this, as it actually occurred, was not guaranteed by E. The man who shot John, therefore, will be indicted for murder, but he cannot be held responsible for the consequences of E' for they would have come about anyhow.

(4) *The principle of causality*

Not every event has a cause – none that is not determined from its very beginning has. Now according to **D10** the principle of causality is not only violated in indeterministic worlds – that wouldn't be so unusual – but also in deterministic ones. In them there is no cause at all. This, however, is not a real objection since our concept of causation was designed for indeterministic worlds. The failure of the causal principle, furthermore, does not imply that there are events that cannot be explained – there are, after all, reasons that are not causes.

The problem of a common cause makes trouble for regularity theories. It does not arise for our conception, for the effect of a cause can never be the cause of something, and therefore not the cause of another, later effect of its own cause. There remains the distinction between causes and necessary circumstances. For regularity-theories this is also a very real problem.[6] A short-circuit in a defective cable is the cause of a fire. But it could only have this effect as there was enough oxygen. Why is the short-circuit but not the presence of the oxygen the cause of the fire? The oxygen is a necessary circumstance in the sense that if there had been no oxygen the short-circuit would not have caused the fire. Can we then in our framework state truth conditions for such counterfactuals? This is important for completing our theory of causation, since every useful theory must be able to distinguish causes from necessary circumstances.

VI. ALTERNATIVES AND STRATEGIES

First, however, we shall point out connexions between our concept of causing and that of an agent bringing something about. In this section we shall supply the conceptual instruments for defining actions. Its content is a shortened version of (1986).[7]

We now conceive of the transition from one WS i to a succeeding WS j as due to the action of agents. In i every agent s has a set $A(s, i)$ of alternatives open to him, and the choice of one of his alternatives by each agent determines the actual successor state of i. Let

D11: $W_i := \{j\colon irj\}$ – the set of immediate successors of i.
Then $A(s, i)$ is a partition of W_i, i.e. a set of disjunct subsets of W_i, whose union is W_i. $A(s, i) = \{W_i\}$ is admitted. In this case the agent s has no real alternatives. This makes it possible for us to assume the same set of agents for every WS, even if only a few of them have a real choice in each of them. For the sake of simplicity we will assume the set of agents $S = \{s_1, \ldots, s_n\}$ to be finite. Our talk of "agents" and "choices" has to be taken in a very broad sense, though. Mother Nature is an agent, too, and her "actions" consist in chance events. The "actions" we talk about, then, are not all of them actions in the usual sense of the word.

Let us call tree-universes for which alternatives are defined, TA-universes.

D12: A *TA-universe* is a quadruple $\mathfrak{U}\mathfrak{A} = \langle I, r, S, A\rangle$ such that
(1) $\langle I, r\rangle$ is a tree-universe.
(2) $S = \{s_1, \ldots, s_n\}$ is a non-empty set of agents.
(3) For all $s \in S$ and $i \in I$:
  (a) $X, Y \in A(s, i) \wedge X \neq Y \supset X \cap Y = \Lambda$
  (b) $\bigcup A(s, i) = W_i$
  (c) $X_1 \in A(s_1, i) \wedge \ldots \wedge X_n \in A(s_n, i) \supset \bigvee j(X_1 \cap \ldots \cap X_n = \{j\})$.
Conditions (a) and (b) say that $A(s, i)$ is a partition of W_i – $X \in A(s, i) \supset X \neq \Lambda$ follows from (c). (c) states that the actions of all the agents together uniquely determine one successor of i. This implies that every agent can execute each of his alternatives, no matter what the other agents are doing. The alternatives of the group $\{s_{k1}, \ldots, s_{kr}\}$ of agents ($k_q \in \{1, \ldots, n\}$, $q \in \{1, \ldots, r\}$) then is:

D13: $A(\{s_{k1}, \ldots, s_{kr}\}, i) := \{X_1 \cap \ldots \cap X_r\colon X_1 \in A(s_{k1}, i) \wedge \ldots \wedge X_r \in A(s_{kr}, i)\}$.

This yields $A(S, i) = \{\{j\}\colon irj\}$.

An (infinite) *strategy* of s in i is defined by the choice of an alternative $X_i \in A(s, i)$, the choice of an alternative X_j for all $j \in X_i$, the choice of an alternative X_k for each $k \in X_j$ and all such X_j, and so on. It is a segment $\mathfrak{U}'$ of the tree $\mathfrak{U}$, beginning with i and defined by a relation r' for which $\{k: jr'k\} \in A(s, j)$. A *strategy in the wider sense* is a segment $\mathfrak{U}'$ of $\mathfrak{U}$, beginning with i and defined by a relation r' such that for all j in the domain of r' there are alternatives $X_1, \ldots, X_r$ from $A(s, j)$ with $\{k: jr'k\} = X_1 \cup \ldots \cup X_r$. So it does not commit s to just one alternative in every WS j but leaves it open for some or all j, what alternative s choses from a subset of $A(s, j)$. A strategy is *finite* if in every branch of $\mathfrak{U}'$ there is a WS k after which r' coincides with r, i.e. if it guides s only up to k. In what follows the word "strategy" is always understood in the sense of finite strategies in the wider sense.

We can also represent strategies as sets of worlds. If a strategy corresponds to the segment $\mathfrak{U}'$ of $\mathfrak{U}$ it is then taken as the set of worlds of $\mathfrak{U}$ belonging to $\mathfrak{U}'$. Let

D14: $f(s, w, t) := \{w': w'(t + 1) \in \iota X(X \in A(s, w(t)) \wedge w(t + 1) \in X)\}$. $f(s, w, t)$ is the set of worlds going through the WSs of that alternative X of s in $w(t)$, to which $w(t + 1)$ belongs. We have then $f(S', w, t) = \bigcap_{s \in S'} f(s, w, t)$ for all groups S' of agents, and $f(S, w, t) = W^{w(t+1)}$. If $U, U', \ldots$ are sets of worlds we can define the set $R(s, i)$ of strategies of s in i as

D15: $R(s, i) := \{U: \Lambda \neq U \subset W^1 \wedge \bigwedge ww't(w \in U \wedge z(i) \leqslant t \wedge w' \in f(s, w, t) \supset \bigvee w''(w'' \in U \wedge w''(t + 1) = w'(t + 1))) \wedge \bigwedge w(w \in U \supset \bigvee t(z(i) \leqslant t \wedge W^{w(t)} \subset U)) \wedge \bigwedge w(\bigwedge t(z(i) \leqslant t \wedge \bigvee w'(w' \in U \wedge w'(t) = w(t)) \supset w \in U)\}$.

The last condition is to ensure that there are no more sets in $R(s, i)$, than subtrees $\mathfrak{U}$. It could be dispensed with, if, instead of finite strategies we considered only bounded ones as such $U \in R(s, i)$ for which $\bigvee t \bigwedge w(w \in U \supset W^{w(t)} \subset U)$ holds. W^i is a strategy of s in i, too, one that does not commit s to anything. We call it the *empty strategy*. If U is a non-empty strategy of s in i there is another strategy U' of s in i with $U - U' \neq U' - U$. In this sense the agent can omit doing what he does in U. He can do something else which may have different effects. This does not imply, however, that he can do something that excludes any course of events compatible with his doing what he does in U. In the latter sense we might speak of refraining. The non-empty strategies in $R(s, i)$ are then possible courses of action for the agent s in i, which he may omit, but not all of them are such that he can refrain from them.

The non-empty strategy U of s in i uniquely corresponds to the event $E(U) := \{w_\tau: w \in U \wedge \tau_1 = z(i) \wedge \bigwedge t(\tau_2 \leqslant t \equiv W^{w(t)} \subset U)\}$. Let $R^*(s, i)$ be the set of non-empty strategies of s in i, represented as events. Strategies for groups of agents are to be defined correspondingly. Then we have:

$$(*) \qquad E \in R^*(S, w(t)) \equiv \Lambda \neq E \wedge \bigwedge w'\tau(w'_\tau \in E \supset w' \in W^{w(t)} \wedge \tau_1 = t \wedge \neg D(E, w', \tau_2 - 1)) \wedge \bigwedge w(\bigwedge t(z(i) \leqslant t \supset \bigvee w'(w' \in E^0 \wedge w'(t) = w(t)) \supset w \in E^0).$$

VII. CAUSING AND BRINGING ABOUT

If an agent realizes a non-empty strategy, that is an action of s – again only in a very broad sense of this word. s brings it about, that an event E' occurs, if s does something, which causes E'. If we write $A(s, w, E)$ for "In w the event E is an action of s" and $B(w, s, E')$ for "In w s brings it about that E'", we obtain the definitions:

D16: (a) $A(w, s, E) := \bigvee\tau(w_\tau \in E \wedge E \in R^*(s, w(\tau_1)))$
(b) $B(w, s, E') := \bigvee E(A(w, s, E) \wedge K(w, E, E'))$.

The first one is alright because every agent can omit doing each of his actions as we have seen – an action is something the agent could have left undone. The second definition is acceptable since from $E \in R^*(s, w(t))$ we obtain $\bigwedge w'(w' \in W^{w(t)} \supset \neg DB(E, w'))$ – $R^*(s, w(t))$ does not contain the empty strategy. Therefore the actions from $R^*(s, w(t))$ are possible causes in the sense of **D10**, they are not determined from their beginning.[8]

If $S' \subset S$ is a set of agents, the relations $A(w, S', E)$ and $B(w, S', E')$ can be defined correspondingly. For $S' \subset S'' \subset S$ we have

$R^*(S', i) \subset R^*(S'', i)$ and therefore $A(w, S', E) \supset A(w, S'', E)$. This looks worse than it actually is, for if we have $E \in R^*(S', i)$ and $E \in R^*(S'', i)$ the agents that are relevant for determining the strategy E are the same in both cases. We say that an agent s in S' is relevant for E, if $E \in R^*(S' - \{s\}, i)$ does not hold. For $E \in R^*(s, i)$ and $E \in R^*(s', i)$ we always have $s = s'$. The relations $B(w, S', E')$ and $B(w, S'', E')$, however, may hold even for different sets of relevant agents. But this is as it should be, since we have not excluded causal overdetermination in the definition **D10** of causation.

D16 immediately yields:

(1) If s brings it about that E' occurs, there is an action E of s that causes E'. Bringing about it a causing by actions. But we also have
(2) Every cause of E' contains an action of agents by which they bring E' about.

Assuming $K(w, E, E')$ and $w_\tau \in E$, let E_1 be the event $\{w_{\tau'}: \tau_1' = \tau_1 \wedge D(E, w, \tau_2') \wedge \neg D(E, w, \tau_2' - 1)\}$, i.e. the event that E happens in w and comes to its point of determination. Then $E_1 \in R^*(S, w(\tau_1))$. This follows directly from (*) at the end of the preceding section, for $\neg D(E_1, w, \tau_1)$ is a consequence of $\neg D(E, w, \tau_1)$, which follows from **D10**. If there is a time t in which all occurrences of E that start in $w(\tau_1)$ have come to their point of determination, which is the normal case, we may also choose the less exclusive event $\{w'_{\tau'}: \bigvee w'\tau''(w'_{\tau''} \in E \wedge w' \in W^{w(\tau_1)} \wedge \tau_1' = \tau_1 = \tau_1'' \wedge D(E, w', \tau_2') \wedge \neg D(E, w', \tau_2' - 1))\}$ instead of E_1, i.e. the event that E starts in $w(\tau_1)$ and comes to its point of determination. From $K(w, E, E')$ we then obtain $K(w, E_1, E')$. That is: E causes E' iff there is a set of agents who, by doing what they do in the course of E, bring it about that E' occurs. Thus our causal relation has an "actionistic" character – if only in a very wide, formal sense, since we also count chance events as "actions" of Nature. This character is already indicated by the basic idea of **D10**, according to which causes are events that need not have happened together with our interpretation of branching as a matter of choice.[9]

VIII. CONDITIONALS

Let us turn back to the question raised at the end of Section V how conditionals can be interpreted given tree-universes. Our aim is to state truth conditions for a unified type of if-then-sentence which can be read as an indicative conditional or as a counterfactual according to the possibility or impossibility of the if-part. We write $C(w, t, X, Y)$ for "In w and t: If it is the case that X, then it is the case that Y". X, Y, . . . are to be SAs, i.e. subsets of $W \times T$. Conditional relations do not hold only between events but also between types of events, that may occur several times in the same world, between states or eternal SAs.[10] The truth value of conditionals depends not only on worlds but also on the time of their utterance. The sentence "If it will rain, John will stay at home" may be true today but false tomorrow when other circumstances obtain.

An indicative conditional "If X then Y" is usually uttered only if it is both possible that X holds and that X does not hold. "Possible" can be understood in the sense of an alethic or of a doxastic modality; X may be contingent or it may be unknown whether X holds. Ignoring epistemic matters the normal condition for indicative conditionals is $W^{w(t)} \cap X_t \neq \Lambda \neq W^{w(t)} \cap \bar{X}_t$. In this case we say that $C(w, t, X, Y)$ holds iff $W^{w(t)} \cap X_t \subset Y_t$, i.e. iff $N(w, t, \bar{X} \cup Y)$. So "If X, then Y" is true iff Y holds in all the actually possible worlds in which X holds. As a rule, also, we only say "If X then Y" in cases where it is uncertain whether Y holds, i.e. for $W^{w(t)} \cap Y_t \neq \Lambda \neq W^{w(t)} \cap \bar{Y}_t$". But if Y is certainly not the case, i.e. for $W^{w(t)} \subset \bar{Y}_t$ the statement "If X then Y" is wrong. And if Y is certain, i.e. for $W^{w(t)} = Y_t$, we should only say something like "Y, even if X", and this will then be true for all X satisfying the normal condition. The result is:

$$\text{(a)} \qquad W^{w(t)} \cap X_t \neq \Lambda \neq W^{w(t)} \cap \bar{X}_t \supset (C(w, t, X, Y) \equiv W^{w(t)} \cap X_t \subset Y_t).$$

A counterfactual "If X were the case, then Y would be the case" is uttered only when it is certain that X is not the case. This, again, can be taken epistemically or alethically. In the latter case the normal condition is $W^{w(t)} \subset \bar{X}_t$. We distinguish two cases: If it has never been possible in w that X would hold in t, i.e. for $W^{w(0)} \subset \bar{X}_t$ $C(w, t, X, Y)$ is to be true. This is a borderline case so that we are in no danger to offend any intuitions. If, on the other hand, X_t has been possible at some time in w, $C(w, t, X, Y)$ is to hold if in the latest among these times, t'', the sentence "If X will hold in t, then Y will also hold in t'' was true under normal conditions for indicative conditionals, i.e. for $W^{w(t'')} \cap X_t \subset Y_t$. It is sufficient to demand that there is such a time t', however, for if $t' < t''$ and $W^{w(t')} \cap X_t \subset Y_t$ we also have $W^{w(t'')} \cap X_t \subset Y_t$ because of $W^{w(t'')} \subset W^{w(t')}$; $t' \leqslant t$ holds in view of $W^{w(t)} \cap X_t = \Lambda$. $W^{w(t)} \subset \bar{X}_t$ implies $W^{w(t')} \cap \bar{X}_t \neq \Lambda$ for all $t' \leqslant t$. The idea, then, is this: Evaluating a counterfactual "If X were the case, then Y would be the case" we have to go back to a situation, i.e. a time, when it was still possible that X would hold in t. This situation should be as similar to the present one as possible, and therefore should be the last point t' in which it was possible that X should hold in t. If in this moment we can say "If X will hold in t, then Y will hold in t'' the counterfactual is true, otherwise false.

Now we normally say "If X were the case, then Y would be the case" only if it is also certain, that Y does not obtain, i.e. for $W^{w(t)} \subset \bar{Y}_t$. For $W^{w(t)} \subset Y_t$ we should rather say "Even if it would be the case that X, it would still be the case that Y". But this case, as well as the one in which Y is uncertain in t, can be treated in the same way as the normal one. So we get:

(b) $$W^{w(t)} \subset \bar{X}_t \supset (C(w, t, X, Y) \equiv W^{w(0)} \subset \bar{X}_t \lor Vt'(t' \leqslant t \land \Lambda \neq W^{w(t')} \cap X_t \subset Y_t)).$$

If it is certain that X obtains we neither say "If X is the case, . . . " nor "If X were the case, . . . ". Therefore we can complete our stipulations any way we like, for instance like this:

(c) $$W^{w(t)} \subset X_t \supset (C(w, t, X, Y) \equiv W^{w(t)} \subset Y_t).$$

This would be unacceptable if we wanted to read $C(w, t, X, Y)$ as "Y because X" if the premis of (c) holds, but we have no ambition to include such sentences in our analysis.

Together (a) to (c) yield the definition:

D17: $C(w, t, X, Y) := W^{w(0)} \subset \bar{X}_t \lor Vt'(t' \leqslant t \land \Lambda \neq W^{w(t')} \cap X_t \subset Y_t)$.

$F(X) := \{\langle w, t\rangle: Vt'(t < t' \land \langle w, t'\rangle \in X)\}$ is the SA that X will be the case, $P(X) := \{\langle w, t\rangle: Vt'(t' < t \land \langle w, t'\rangle \in X)\}$ the SA that X was the case. With these operations we can also formulate conditionals like "If X will be the case, then Y will be the case", "If X was the case, Y will be the case", etc.

Mostly conditionals are interpreted by similarity relations between worlds.[11] The logic of conditionals resulting from **D17** is the standard one (Lewis', not Stalnaker's) if only weakly centered comparative similarity systems are used, i.e. relations $w' \leqslant_w w''$ (w' is at most as similar to w as w'' is), for which we have not $w \leqslant_w w' \supset w' = w$, but only $w' \leqslant_w w$. This difference is important: Strongly centered systems make "If A, then B" true in case "A" and "B" are both true, and such a principle is not acceptable for indicative conditionals. In our framework we can define weakly centered systems as follows: Let $S_w = W^{w(0)}$ and for all $w', w'' \in S_w$: $w' \leqslant_{w,t} w''$ iff $\bigwedge t'(t' \leqslant t \wedge w' \in W^{w(t')} \supset w'' \in W^{w(t')})$. A world w'' from S_w, then, as seen from t, is more "similar" to w than w' if it shares a longer common history with w in the interval $[0, t]$ than w' does. We assume that t is not the last of all times, so that $t + 1$ is in T. Otherwise the relation would coincide with $\bigwedge t(w' \in W^{w(t)} \supset w'' \in W^{w(t)})$, and that yields a strongly centered system. Finally we stipulate: If w' is not in S_w, $w' \leqslant_{w,t} w''$ holds for all w'', and if w'', but not w' is in S_w then $w' <_{w,t} w''$. The relations $\leqslant_{w,t}$ are transitive and connex. They only have the formal properties of similarities; intuitively worlds diverging earlier may be more similar if they differ only in minor aspects.

If $C(A, B)$ is the equivalent of "If A, then B" in the language of conditional logic the standard truth condition for it is:

$$(*) \qquad V_w(C(A, B)) = t \text{ iff } S_w \subset [\neg A] \vee \bigvee w'(w' \in [A] \cap S_w \wedge \bigwedge w''(w' \leqslant_{w,t} w'' \wedge w'' \in [A] \supset w'' \in [B])).$$

Here V is the interpretation function and $[A]$ is the set of A-worlds, i.e. the set $\{w: V_w(A) = t\}$. This corresponds to **D17** if we turn to interpretation functions $V_{w,t}$ – the truth-conditions in $(*)$ depend on t – and consider only eternal SAs as arguments of $C(w, t, X, Y)$. X is an eternal SA if there is a set U of worlds with $X = U \times T$. If U and U' are sets of worlds **D17** says $C(w, t, U, U') \equiv S_w \subset U \vee \bigvee t'(t' \leqslant t \wedge \bigwedge \neq W^{w(t')} \cap U \subset U')$. For $U = [A]$ and $U' = [B]$ this is equivalent to $(*)$.[12]

While the notion of an overall similarity between worlds remains extremely vague in the usual treatments it receives a more precise sense in our definition by reference to tree-universes. Our intuitions about the branching of worlds in a model supposed to represent a part of reality are, of course, not any clearer than those concerning the validity of conditionals about this part of the world. As we determine the set of possible worlds referred to in an interpretation of a fragment of natural language with a view to what sentences we take to be analytically true or analytically false, we have to determine the branchings of the world with reference to the conditionals we take to be true. We have to ask: What will be the case, or what is actually possible given this and that, and what would be the case or have been possible if this or that would have happened. Nevertheless the statement, that a sentence is analytically true if it is true in all possible worlds, is not useless for our understanding of the notion 'analytical truth'. And the same holds for the interpretation of conditionals by **D17**. It is not the task of logic to say which conditionals are actually true. But the utility of tree-universes does also not consist merely in defining a logic for conditionals. They rather allow us to study conditional dependences in simple models, and in this they are more useful than a reference to unspecified similarity relations.

As conditionals are only a side-issue in this paper I shall not discuss the intuitive correctness of **D17**. It should be emphasized, however, that the use of the indicative or subjunctive mood depends on our assumptions about the facts. We say "If Oswald didn't shoot Kennedy, then someone else did", if we are not sure about Oswald being responsible, although now it is either necessary that he did it or necessary that he did not do it, so that our objective normal condition for indicative conditionals is not satisfied. The use of the indicative presupposes the subjective uncertainty of the antecedent condition and the acceptability of our conditional depends upon the further assumption that Kennedy has indeed been shot. The counterfactual "If Oswald had not shot Kennedy, then someone else would have", on the other hand, presupposes that Oswald did the shooting. This presupposition is connected with our conviction that Kennedy has been shot, and therefore this conviction will be suspended on the counterfactual assumption. This explains why we do not accept the counterfactual. This example does not, then, show that indicatives and counterfactuals have different truth conditions. All it shows is that they have different presuppositions that may result in different epistemic attitudes towards them.

IX. CAUSES AND NECESSARY CIRCUMSTANCES

A workable theory of causation should be able to distinguish causes from necessary circumstances, we said at the end of Section V. In our example of a short-circuit causing a fire the presence of oxygen was such a necessary condition. Let us take circumstances to be events here. This is no strong restriction of generality, for the continuance of a state in an interval, its beginning or ending are events, too. We mostly talk about necessary circumstances at a time when the cause is already completed and it is certain that the effect thas occurred or will occur. Then every necessary condition for its occurrence is sure to be fulfilled. But if E'^0, the effect's occurrence, is necessary in t, and F is any event, for which F^0 is also necessary in t, the statement that F occurs if E' does, is true. Therefore this conditional is not an appropriate expression of a necessary condition. We rather have to use the counterfactual "If F would not occur, E' would not occur either". We say, then, that F, as seen from w and t, is a *necessary condition* for E' iff $C(w, t, \overline{F^0}, \overline{E'^0})$.

Necessary conditions, however, are not always necessary circumstances. E', for instance, is a necessary condition for E' itself, and a cause of E' can also be a necessry condition for E'. A necessary circumstance is an event that is already completed, or at least determined, when E' begins. It is a *precondition* for E', i.e. a necessary condition for which there is a time t' such that $D(F, w, t') \wedge \neg D(E', w, t')$. This excludes the two unwanted cases, for if E is a cause of E', E' is determined when E is.

The presence of oxygen is a precondition for the fire, since a fire can only develop if there is oxygen. Let us assume now that the short-circuit has occurred in a defective part of a cable. Then it is also a necessary circumstance for its causing the fire that some inflammable material was close to this part of the cable. Let this be the event F. F is no precondition for the fire, since the fire could also have arisen at another place in the building, by arson, e.g. A necessary circumstance for the causation of one event by another, then, is not always a precondition for the effect, but rather a necessary condition for the first event's causing the second. To exclude unwanted cases we again have to postulate that the circumstance be determined before the effect. If we write $NC(w, t, F, E, E')$ for "As seen from w and t F is a necessary circumstance for E's causing E'," we then obtain the definition:

D18: $NC(w, t, F, E, E') := C(w, t, \overline{F^0}, \{w: \neg K(w, E, E')\}) \wedge Vt'(D(F, w, t') \wedge \neg D(E', w, t'))$.

Every precondition for E' is also a necessary condition for E' being the effect of some cause, but the inverse does not hold.

This definition has to be tested on further examples, but the main thing was to show that our framework is rich enough to define truth conditions even for so complex sentences as "If F would not have occurred, E would not have caused E'".

NOTES

[1] Strictly speaking, X_t is not a SA according to our terminology, but sets U of worlds correspond to the *eternal* SAs $U \times T$, i.e. to SAs that hold always or never in each world.

[2] For an example of self-causation, cf. Lewis (1986), pp. 172 sq.

[3] For physical examples cf. Hesse (1961), pp. 279 sq. The standard argument for admitting backward causation is the possibility of foreseeing future events: Since our visual experience is caused by the events we witness, foreseeing would be a case of backward causation. This argument is a *petitio principii*, however, for if seeing is a causal process, foreseeing can only be a sort of seeing, if there is backward causation. Otherwise we have to conceive of it as a kind of precognition. My present knowledge that the sun will rise tomorrow is not causally dependent on tomorrow's sunrise – though dependent on it, since I cannot know what will not happen – and the same thing holds for precognitions.

[4] In view of Simpson's paradox this holds only for appropriate probabilities.

[5] Cf. his postscript to "Causation" in (1986), pp. 196 sq.

[6] According to regularity theories a SA A is a cause of B, if A and B are true and there is a non-empty set L of laws and a set C of singular conditions such that B is implied by L, C, A, but not by L, C or C, A alone. Now if B does not follow from L, A and C minus one of its elements, C_1, then C_1 is also a cause of B relative to L and C with A substituted for C_1.

[7] Nuel Belnap has developed very closely related ideas independently in two papers from 1989.

[8] **D16**(b) does not correspond exactly to the definitions given in my (1986). – If we want to talk about modes of actions which can be realized by the same agent at different times and by different agents at the same time, we first have to introduce *types of events*. The occurrences of a type T of events are events, and different occurrences of T in the same world have to be distinct. If we want to think of T as a set of world

segments an equivalence relation has to be defined on T such that the equivalence classes $[w_\tau]_T$ are events and $w_\tau, w_{\tau'} \in T$ implies: $\tau = \tau'$ or there is at least one point t between τ and τ'. $F(s)$ is a mode of action of agent s if it is a type of events for which $\bigwedge w\tau(w_\tau \in F(s) \supset [w_\tau]_{F(s)} \in R^*(s, w(\tau_1)))$ holds. Then we can take a mode of action to be a function F such that $F(s)$ is a mode of action of s for all s. s does F in w_t if $w_\tau \in F(s)$. – An agent can do F by doing something more specific. s does F in w_τ *by* doing F' if $w_\tau \in F(s)$ and there is an interval τ' with $w_{\tau'} \in F'(s)$ and $[w_{\tau'}]_{F'(s)}{}^\circ \subset [w_\tau]_{F(s)}{}^\circ$. ($[w_{\tau'}]_{F'(s)}{}^\circ$ is a strategy out of $R(s, w(\tau_1')))$.). A. Goldman first pointed out the importance of such *by*-constructions for action theory in (1971).

[9] The term "actionistic" was coined by G. H. von Wright. His analysis of causation in (1974) is different from mine even though there are parallels – he, too, uses tree-universes and time dependent alethic modalities. He first defines causal laws, essentially as statements to the effect that every event of a type T_1 is immediately followed by an event of type T_2, and that this has always been necessary. Singular causal statements are instances of such causal laws. I, on the other hand, have not demanded that they be subsumable under a causal law. As we have seen in Sect. IV, the effects of an event depend on circumstances obtaining when it occurred. With v. Wright the actionistic idea comes into play only in his condition for the testability of causal laws. He does not introduce agents and their alternatives but distinguishes one successor state for each WS as the normal one, the one which will result if all the agents refrain from interfering with the course of nature. The word "agent" is to be understood in the normal, narrower sense here. In our terminology this means: Nature never has a real alternative, and an alternative X of s in i is a *genuine* action only if the normal state of W_i is not in X. Every agent who has a real alternative in i, then, also has the alternative to do nothing, not to interfere with nature. The introduction of normal states (if there are more than one, only one of them would be in each alternative of Nature, and the other agents would have one alternative in which all the normal states are included) is useful in a theory of action for a distinction between doing something and letting something happen. The condition of testability for the causal law that T_1-events are necessarily followed by T_2-events, says that there has to be a possible occurrence of T_1 which normally would not come about but can be brought about by the intervention of agents. If they bring it about they can test whether this T_1-event, too, is followed by a T_2-event, or whether the regularity just holds for the normal course of events. There is, in v. Wright's (1974), no statement to the effect that every instance of causation is also an instance of something brought about by agents. This, however, would not hold for us either, if we had only talked about genuine agents and genuine actions. As Max Urchs pointed out to me, St. Jaskowski has defined causation in (1951) in a similar way as I have done. Jaskowski, in turn, refers to R. Ingarden for the basic idea.

[10] As in (1986) states can be defined as sets of WS. If Z is a subset of I it corresponds to the SA $\{\langle w, t\rangle : w(t) \in Z\}$.

[11] Cf. Lewis (1973).

[12] We can also use selection functions instead of similarity relations. If $f(w, t, X)$ is $X_t \cap S_w \cap \bigcap\{W^{w(t')}: t' \leqslant t \wedge W^{w(t')} \cap X_t \neq \Lambda\}$ we have $f(w, t, X) \subset Y_t \equiv C(w, t, X, Y)$. f satisfies the usual conditions: $f(w, t, X) \subset X_t \cap S_w$, $X_t \subset Y_t \wedge f(w, t, X) \neq \Lambda \supset f(w, t, Y) \neq \Lambda$, $f(w, t, X) \cap Y_t \neq \Lambda \supset f(w, t, X \cap Y) = f(w, t, X) \cap Y_t$ and $w \in f(w, t, X \times T) = W^{w(t)}$. We also have: $w' \in f(w, t, W \times T) \supset f(w, t, X) = f(w', t, X)$ and $\bigcup f(w, t, X) = S_w$. Because of $w' \in W^{w(0)} \supset W^{w(0)} = W^{w'(0)}$ the concept $N^+(w, t, X) := C(w, t, \bar{X}, X) \equiv S_w \subset X_t$ has the properties of an

S5-necessity. These additional principles supply truth-conditions for sentences with iterated applications of the conditional operator. They correspond to those proposed in Kutschera (1976), 3.2 and 3.3. – As far as I see, the closest relative to the interpretation of conditionals proposed here is that given by Richmond Thomason and Anil Gupta in (1981). They, too, use tree-universes and have a Principle of Past Predominance according to which worlds are more similar if they share a longer common history. Antecedent similarity has also been emphasized by Frank Jackson, Brian Ellis, David Lewis and Wayne Davis. In "Counterfactual dependence and time's arrow" (1979, repr. in (1986), pp. 32 sq.) Lewis gives an "Analysis 1" of counterfactuals (p. 39), which is very closely related to ours. His reasons for discarding it are: (1) It does not work for antecedents like "If kangaroos had no tails . . . " that do not refer to any specific time. (2) It makes the asymmetry of counterfactual dependence between future and past an analytical truth. The first objection cannot be raised against our definition, however, and to me the dependence of the present on the past but not on the future is not an empirical matter. What Lewis says in this paper about assessing similarities and how smaller or greater miracles subtract from them seems to have just one clear consequence: we may choose as standards of similarity whatever fits the intended result, and this means just that similarities are no help whatever for their evaluation. Counterfactuals may be vague, but surely they are far from being as vague as the notion of similarity of worlds.

REFERENCES

Belnap, N.: Backwards and forwards in the modal logic of agency, Manuscript 1989.

Belnap, N.: Before refraining: Concepts for agency, *Erkenntnis* **34** (1971), 137–69.

Goldman, A.: The individuation of action, *The Journal of Philosophy* **68** (1971), 761–74.

Hesse, M. B.: *Forces and Fields*, London (Thomas Nelson and Sons) 1961.

Jaskowski, St.: On the modal and causal functions in symbolic logic, *Studia Philosophica* **4** (1951), 71–92.

Kutschera, F. v.: *Einführung in die intensionale Semantik*, Berlin (de Gruyter) 1976.

Kutschera, F. v.: Bewirken, *Erkenntnis* **24** (1986), 253–81.

Kutschera, F. v.: Sebastian's strolls, to appear 1993 in *Grazer Philosophische Studien*.

Lewis, D.: *Counterfactuals*, Oxford (Basil Blackwell) 1973.

Lewis, D.: *Philosophical Papers*, vol. II, Oxford (University Press) 1986.

Thomason, R. H. and Gupta, A.: A theory of conditionals in the context of branching time, in W. L. Harper, R. Stalnaker, G. Pearce (eds.): *Ifs*, Dordrecht (Reidel) 1981, 299–322.

Wright, G. H. von: *Causality and Determinism*, New York (Columbia University Press) 1974.

Philosophische Fakultät I,
Universität Regensburg,
Universitätstrasse 31,
Postfach 397,
8400 Regensburg,
Germany