WOLFGANG SPOHN

STOCHASTIC INDEPENDENCE, CAUSAL INDEPENDENCE, AND SHIELDABILITY*

ABSTRACT. The aim of the paper is to explicate the concept of causal independence between sets of factors and Reichenbach's screening-off-relation in probabilistic terms along the lines of Suppes' probabilistic theory of causality (1970). The probabilistic concept central to this task is that of conditional stochastic independence. The adequacy of the explication is supported by proving some theorems about the explicata which correspond to our intuitions about the explicanda.

1. INTRODUCTION

The conjecture that there is a close connection between causal and probabilistic concepts is an old one dating back at least to Hume. Some frequentist interpretations of probability assumed outright that causal independence implies stochastic independence; a glance at textbooks of probability theory, in particular older ones, shows how widespread this assumption is or was. Or looking at multivariate analysis, we find people, albeit cautious in their claims, very eager to infer causal assertions from their probabilistic data. Recently there have been efforts to *define* causal concepts in probabilistic terms. Suppes (1970) is most explicit in this respect, giving a probabilistic explication of the notion of an event A being a cause of an event B.

However, it is by no means obvious that such projects are basically sound and satisfactorily practicable. My aim here is to argue that they are, demonstrating this with a somewhat less difficult subject than Suppes': namely the concept of causal independence between variables, factors, or whatever you may call it – and the distinction between direct and indirect or, more appropriately, shieldable and unshieldable causal dependence. (This is less difficult a subject, since we need not bother about positive and negative causal relevance.)

In more detail, my program is the following: In Section 2 I shall first set out the formal framework I am working with and then define and characterize the notion of conditional stochastic independence, which will prove to be central. All this is preparatory and completely standard.

Journal of Philosophical Logic 9 (1980) 73–99. 0022–3611/80/0091–0073$2.70.

Section 3 works up to the desired explication. The strategy I have adopted there consists in accumulating as many probabilistic arguments for and against causal independence as possible; when we have taken into account all these arguments and ascertained that no new ones can possibly turn up, then we may fairly hope to have said all about causal independence that can be said in probabilistic terms. Likewise for shieldability.

In Section 4 I shall derive some properties of the notions thus explicated and argue for the desirability of just these properties, intending thereby to support not only the adequacy of the explications, but also the probabilistic definability of causal concepts in general. The proofs of the theorems contained in the Sections 2 and 4 are deferred to the final section.

I have tried to be mathematically as general as is possible, with two major exceptions: the whole paper is confined to a discrete time structure, and my explication will moreover work only for those probability measures which I shall call strictly indeterministic (cf. Section 2). These restrictions, which are actually interrelated, are technically required for my explication as it stands, which is therefore not yet applicable to most of the physical examples. But the considerations carried through here for a restricted case are definitely vital to the unrestricted case as well, which shows additional problems. These are however not really essential to the gist of my paper. I shall comment on these restrictions in more detail at the end of Section 4.

2. CONDITIONAL STOCHASTIC INDEPENDENCE

As I said, I am concerned here with independence between factors or variables, i.e., such things as the temperature in Munich at July 17th, 1977, 10 p.m., the blood pressure of person X at time t, the color of car X at t, the ruling parties of state X at t, the number of drug addicts in Germany at t, the action of person X at t, etc. Of course, this is to be understood not as the *actual* temperature, blood pressure, etc. (these would be events), but rather as the set of *possible* temperatures, blood pressures, etc. Mathematically, such variables or factors are represented either by random variables or by σ-algebras or measurable spaces. Here I take measurable spaces, since stochastic independence of random variables is defined by the stochastic independence of the σ-algebras generated by them. Hence, we are dealing with a (finite or infinite) collection $(\langle\Omega^i, \mathfrak{A}^i\rangle)_{i\in I}$ of measurable spaces indexed by the index set I. Each Ω^i may in turn be finite or infinite, but to avoid trivialities we assume that each Ω^i has at least two elements.

Of course, we have to combine these measurable spaces to form one big measurable space. This is done by forming the product space: Let Ω be the set of all functions ω defined on I such that for all $i \in I$ $\omega(i) \in \Omega^i$. For every $i \in I$ the projection function π_i from Ω into Ω^i is defined by $\pi_i(\omega) = \omega(i)$ for all $\omega \in \Omega$. For every $J \subseteq I$ we denote by $\mathfrak{A}_J$ the σ-algebra on Ω generated by the functions $(\pi_i)_{i \in J}$; in particular, we set $\mathfrak{A}_{\{i\}} = \mathfrak{A}_i$ and $\mathfrak{A}_I = \mathfrak{A}$. $\langle \Omega, \mathfrak{A} \rangle$ then is the desired product of $(\langle \Omega^i, \mathfrak{A}^i \rangle)_{i \in I}$. Thus, we may translate independence statements about single factors or about set of factors into independence statements about the $\mathfrak{A}_i$ $(i \in I)$ or about the $\mathfrak{A}_J$ $(J \subseteq I)$. Since we shall often make such statements, it is convenient and simplifying notation to talk not of the σ-algebras themselves, but of the indices or index sets representing them. Thus we will need a long list of variables for indices and sets of them: I shall use the letters i, j, k, and l as variables for elements of I, the letters J, K, L, and M as variables for subsets of I, and the letters A, B, C, and D as variables for events, i.e., elements of $\mathfrak{A}$. This is to be understood with or without subscripts throughout the paper. Often I shall talk of an index $i \in I$ itself as a factor or variable.

I will add the temporal structure here, though it does not yet become relevant in this section: to this end, we assume a set T of time indices representing points or intervals of time. Metric properties of time are not relevant here; so we only assume T to be linearly ordered by $\leqslant$. Of course, $t \leqslant t'$ is to mean that t is not later than t'; the expressions $t \geqslant t'$, $t < t'$, $t > t'$, and $t = t'$ are self-explanatory. As mentioned before, we want to deal only with discrete time; therefore we assume that T is order isomorphic to some subset of the set of integers. To complete the temporal structure, we have to associate with each factor $i \in I$ a time index $\tau_i \in T$ at which i will realize; different factors may of course get the same time index (in this respect our framework is more general than the standard framework for stochastic processes). As a variable for elements of T I shall use the letter t with or without subscripts. Moreover, the following notation will prove to be convenient: For $J \subseteq I$ and $t \in T$ $J_{\leqslant t}$ is to denote the set of all $i \in J$ such that $\tau_i \leqslant t$; $J_{\geqslant t}$, $J_{<t}$, $J_{>t}$, and $J_{=t}$, are defined similarly.

Finally, let us not forget to assume a σ-additive probability measure P on $\mathfrak{A}$; then we have gathered all the working material we need.

We have however mentioned a restriction on P; this is specified in

DEFINITION 1. P is *strictly indeterministic* iff for all disjoint $K, L \subseteq I$ and all $A \in \mathfrak{A}_K$ and $B \in \mathfrak{A}_L$ the following holds: if $P(A|B) = 0$, then $P(A) = 0$.

Thus, a strictly indeterministic P does not establish any necessary connection between contingent events belonging to disjoint set of factors (where "necessary" and "contingent" are of course to be understood relative to P). I call this *strictly* indeterministic, because every probability measure is, in a sense, not deterministic. Wherever I suppose P to be strictly indeterministic I shall note this explicitly. (Actually, in a frame as general as ours, this definition will not quite do. But this is merely a technicality and need not worry us in the following. For details see the proof of Theorem 1(e) in Section 5.)

So let us turn to conditional stochastic independence: Stochastic indepen dence between two factors or, more generally, between two sets $K, L \subseteq I$ of factors (relative to P) of course means that $\mathfrak{A}_K$ and $\mathfrak{A}_L$ are stochastically independent, i.e., that for all $A \in \mathfrak{A}_K$ and $B \in \mathfrak{A}_L$ $P(A \cap B) = P(A) \cdot P(B)$. I shall symbolize this by $K \perp L$.

Conditional stochastic independence is a bit harder to define: First suppose, for expositive reasons, that I and every Ω^i is finite. Thus, for every $M \subseteq I \mathfrak{A}_M$ is finite and consists of $\emptyset$, its atoms, and all unions of atoms. That K and L are stochastically independent conditional on M then means that for all $A \in \mathfrak{A}_K, B \in \mathfrak{A}_L$ and all atoms C of $\mathfrak{A}_M$ with $P(C) \neq 0$

$$(2.1) \qquad P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

holds, i.e., that any event in $\mathfrak{A}_K$ is independent of any event in $\mathfrak{A}_L$ given any possible state the factors of M might be in (where this latter clause is just a circumlocution for the nonnull atoms of $\mathfrak{A}_M$).

If we allow some or all Ω^i and I to be infinite, this definition does not work, however, since $\mathfrak{A}_M$ may be atomless for some $M \subseteq I$. We have then to resort to a more general mathematical concept, namely to the conditional probability $P^{\mathfrak{B}}$ of P with respect to some sub-σ-algebra $\mathfrak{B}$ of $\mathfrak{A}$. Using this we may define:

DEFINITION 2. For all $K, L, M \subseteq I$, K is *stochastically independent of L conditional on M* (relative to P), symbolized by $K \perp L/M$, if and only if for all $A \in \mathfrak{A}_K$ and $B \in \mathfrak{A}_L$

(2.2) $P^{\mathfrak{A}_M}(A \cap B) = P^{\mathfrak{A}_M}(A) \cdot P^{\mathfrak{A}_M}(B)$ P-almost surely.

The reader unfamiliar with this concept of conditional probability may either look it up in the literature (e.g., in Loève (1960), ch. VII) or assume finiteness of I and all Ω^i throughout this paper and read every statement of the form (2.2) as equivalent to (2.1).

The most important properties of the relation thus defined are compiled in the following

THEOREM 1. For all $J, K, K', L, L', M, M' \subseteq I$ we have:

(a) $\emptyset \perp K/M$;

(b) if $K \perp L/M$, then $L \perp K/M$;

(c) if $K \perp L/M, K' \subseteq K \cup M, L' \subseteq L \cup M$, and $M \subseteq M' \subseteq M \cup K \cup L$, then $K' \perp L'/M'$;

(d) if $K \perp J/L \cup M$ and $L \perp J/M$, then $K \cup L \perp J/M$;

(e) if $K \cup L \perp J/M$ and $K \cup M \perp J/L$, then $K \cup L \cup M \perp J/L \cap M$, provided that P is strictly indeterministic.

(For the proof of Theorem 1 see Section 5.) For infinite I the following theorem is relevant, too:

THEOREM 2. Let $(M_\delta)_{\delta \in \Delta}$ be a (possibly uncountable) family of subsets of I, $\mathfrak{E}(\Delta)$ the set of all finite subsets of Δ, and $K, L, M \subseteq I$ such that $M_\delta \subseteq M$ for every $\delta \in \Delta$. Then we have:

(a) if for all $\Gamma \in \mathfrak{E}(\Delta)$ $\bigcup_{\delta \in \Gamma} M_\delta \perp K/L$, then $\bigcup_{\delta \in \Delta} M_\delta \perp K/L$;

(b) if for all $\Gamma \in \mathfrak{E}(\Delta)$ $K \perp L/\bigcup_{\delta \in \Gamma} M_\delta$, then $K \perp L/\bigcup_{\delta \in \Delta} M_\delta$;

(c) if for all $\Gamma \in \mathfrak{E}(\Delta)$ $K \perp L/\bigcap_{\delta \in \Gamma} M_\delta$, then $K \perp L/\bigcap_{\delta \in \Delta} M_\delta$;

(d) if for all $\delta \in \Delta$ $K \cup M_\delta \perp L/M \backslash M_\delta$, then $K \cup \bigcup_{\delta \in \Delta} M_\delta \perp L/M \backslash \bigcup_{\delta \in \Delta} M_\delta$, provided that P is strictly indeterministic (where "\" denotes set theoretic difference).

In Theorem 2, as throughout the paper, we allow I to be uncountable.

In most applications, however, an uncountable I will only be needed when time is continuous. Thus, this generality is redundant here; but since it makes no difference to any formulation or proof, we may admit it. Note also, that Theorem 2(d) is just an infinite version of Theorem 1(e).

Obviously, stochastic independence may not be identified with causal independence, simply because by being symmetric (Theorem 1(b)) it disregards temporal relations. But there is also a deeper reason for this. Causal independence is a notion we may not be very clear about, but with which we are intuitively familiar. In fact, if we speak of two things being empirically independent, we usually mean somthing like causal independence. This is documented by our strong inclination to recast phrases such as "is independent of" and "depends on" into more clearly causal phrases such as "has no impact on", "affects", "influences", etc.

In contrast to this, conditional stochastic independence between sets of factors as disclosed by Theorem 1 seems to me to be a rather unintuitive and intricate relation (and this carries over to unconditional stochastic independence). In my view, the main reason for this is that stochastic independence misses an elementary property of intuitive independence notions, which we may call *conjunctivity*: namely that, if x is independent of y and independent of z, then it is also independent of y and z taken together, and conversely that, if x is independent of z and y is independent of z, then x and y taken together are independent of z.

Stochastic independence misses conjunctivity, because $K \perp L$ and $K \perp M$ do *not* imply that $K \perp L \cup M$; it obeys more complicated laws instead. (By the way, this is one important reason for many troubles philosophers have with confirmation and relevance, one of which consists in the counterintuitive fact that some evidence may be irrelevant to each of two hypotheses and yet confirm both taken together. Cf., e.g., Salmon (1975).) Intuitive notions of independence, however, do satisfy conjunctivity; this is supported by the fact that in saying that x is independent of y and z we do not, and need not, distinguish between "y and z taken individually" and "y and z taken together". (The same holds in the converse case.) This is why they are logically simple, and this in turn is the reason for their being intuitively manageable.

This discussion shows that it is an important criterion for the adequacy of our explication of causal independence that it fulfill conjunctivity. We shall check this. (Let me add, that symmetry seems to me not to be intuitively necessary for independence relations; else we would not need the phrase "is independent of", which is not symmetric. And surely, we must not expect causal independence to be symmetric.)

3. CAUSAL INDEPENDENCE AND SHIELDABILITY: EXPLICATION

As just stated, stochastic independence must not be equated with causal independence. Nevertheless, knowing that one factor does or does not stochastically depend on another factor, we are strongly tempted to infer from this the corresponding causal statement, given the right temporal order. There are, however, three more or less well-known ways in which this temptation may lead us astray. Our explication will essentially consist in systematically avoiding these errors. Since my reasoning to this end will become somewhat abstract, it may be of help to illustrate these three types of error by examples:

First, if confidential information is trustworthy, the Educational Department is worried about the high rate of regular Mickey Mouse readers among the eight to twelve year old children, which at present is 60%. So it ordered a study which revealed that among the children who have received an authoritarian education the percentage is even higher, namely 70%, while it is only 50% among the other children. Officials were therefore inclined to suspect a causal connection of educational style to consumption of Mickey Mouse. But then a more thoroughly conducted study brought to light a surprising fact: namely that 90% of those children whose parents were MM fans when young are regular MM readers, *irrespective of* whether they receive an authoritarian education or not, whereas the rate among the other children is only 45%, again independent of educational style. So, officials accepted the fact that the popularity of MM inevitably spreads from one generation to the next (and they were additionally shocked on inferring from the data given that one half of all parents prefer an authoritarian educational style, while as many as five out of six former MM fans do so).

This causal state of affairs is well-known. A more familiar example is the falling barometer which predicts without causing the subsequent storm. And such is the way of epiphenomena in general. The mechanism is always the same: The observation that a factor i stochastically depends on a factor j, formally that $i \perp j$ does not hold (where $\tau_i > \tau_j$), suggests that there is a causal influence from j on i, but this is invalidated when a third factor k with $\tau_k < \tau_j$ is found conditional on which i and j are not stochastically dependent any more. Thus, the fact that $i \perp j/k$ shows the influence to be only *spurious*, and the illusion is explained by k influencing both i and j correspondingly. (For similar examples see, e.g., Salmon (1971), pp. 33ff., 53ff., and Suppes (1970), pp. 21–27. There is the corresponding phenomenon of spurious correlation in multivariate analysis; cf. Van de Geer (1971, p. 106.))

The second example is similar to the first one: Since 1963, every time the Rolling Stones have released a new album, a fan club has regularly determined how may Germans think them to be the greatest rock'n'roll band on earth; the rate was always rather low, somewhere around 2%. But among those born between 1943 and 1953 the rate was considerably higher, oscillating between 10 and 15%. Thus, it seems that date of birth is causally relevant for Rolling Stones fanship. But the fan club also found out that always about 80% of those who were enthusiastic about their latest album held that superlative to be true, *irrespective of* date of birth. This is no reason for us now to deny a causal influence from date of birth to fanship, but it deprives the first of its direct causal access to the second, it screens off the first from the second, as Reichenbach (1956, p. 192), termed it.

Such a phenomenon is well-known too. Observe that its probabilistic structure is the same as before. Again, we have a factor i (maintaining the superlative or not) and a factor j (born between 1943 and 1953 or not), such that not $i \perp j$, and again we discover a third factor k (being enthusiastic about the latest LP or not), such that $i \perp j/k$. The difference lies in the temporal structure, for now the intervening k is temporally between the other two. Hence our differing conclusion that k only shows the influence of j on i to be *shieldable*, but not to be spurious. (For this type of argument, cf. also Suppes (1970, pp. 28ff).)

The third case is in a sense just the opposite of the first two: A popular and often cited example in philosophical literature is that regarding a certain kind of neurosis the remission time after psychoanalytic treatment does not significantly differ from the spontaneous one. From this it is concluded that psychoanalytic treatment has no influence on remission time (and thus is of no use whatsoever for the patient). Some Freudian, full of wrath about this silly propaganda, has got to the bottom of the matter, and this is what he has found out: One has to take into account the income of the patient. For, if a patient with low income underwent treatment, his remission time is, on the average, definitely shorter than the spontaneous one, whereas for wealthy patients remission time after treatment is considerably longer. This shows that there *is* an influence of psychoanalytic treatment on remission, though it is brought to light only by deeper analysis.

Cases like this have the following structure: There are two factors i (remission time) and j (psychoanalytic treatment or not), the first being stochastically and therefore presumably also causally independent of the second. However, a third factor k (income of the patient), for which $i \perp j/k$ is false, refutes this conjecture and suggests instead that there is a *hidden* influence of j on i. Note that the temporal position of k is not relevant to this argument if it is only earlier than i. This is particularly clear in a hypothetical example of Van de Geer (1971, p. 106f.): "We do not find a correlation between the amount of rainfall and the amount of wheat produced, measured over consecutive years, whereas the partial correlation after elimination of the effect of daily temperature is positive. That is, for years with equal temperature, there is a correlation between rainfall and amount of wheat produced, but this relation is contaminated by variation in temperature, since higher rainfall accompanies lower temperature, which is disadvantageous for wheat production." And in this case, there is no telling which one is earlier, rain or temperature.

In contrast to the first two types of reasoning, I could not find that this third type has received proper attention, if any at all, in the relevant philosophical literature. But it is of equal importance, and in particular the interplay of these three types will turn out to be most interesting. (This neglect is, I think, largely due to the fact that the long debate about the symmetry of explanation and prediction has been settled with the acceptance of one half of the symmetry, i.e., that explanations could always serve as predictions under appropriate pragmatic circumstances, or that a cause, if known, would be a reason for expecting the effect. Hidden causes and hidden causal dependence do not fit into this picture.)

Thus, there are three ways – I know of no others – in which our inferences from probabilistic premises to causal dependence or independence between factors may be modified by further probabilistic data. What we shall do in the rest of this section, is simply to carry this through to its logical consequence, to a point where no further modification of this sort is possible. This sounds simple, and in fact, it starts simply. Unfortunately, yet inevitably, as it seems to me, it will become somewhat intricate in the end. I apologize for this in advance.

So let us return to the formal framework introduced in Section 2 and let us assume in the sequel that the underlying probability measure is strictly indeterministic, as it already was in our examples. At the end of Section 4 I shall indicate, where things would go wrong without this assumption.

We have first to add a further bit of notation: That $K \subseteq I$ is causally independent of $L \subseteq I$ (relative to P) will be symbolized by $K \perp_c L$; instead of "$\{i\} \perp_c \{j\}$" we shall simply write "$i \perp_c j$". Similarly, $K \perp_s L$ is to mean that K is shieldable from L (relative to P). Moreover, I hope to improve the readability of the next pages by the following convention: In connection with a statement of the form $K \perp L/M$ the expression "$[t]$" is to denote the set $I_{\leqslant t} \backslash (K \cup L)$. Thus, "$i \perp j/[\tau_i]$", for instance, says the same as "$i \perp j/I_{\leqslant \tau_i} \backslash \{i, j\}$". As long as any expression of the form $[t]$ is only related to that stochastic independence statement in which it occurs or which is spoken of in the very moment, this notation cannot lead to any misunderstanding.

Another advisable move is to simplify matters in two respects, as we have already done in our examples: Firstly, we shall restrict the discussion to causal independence between single factors. Only after this has been cleared up can we explain causal independence between sets of factors. Secondly, we shall provisionally assume that every time index is associated with only one factor; we can thereby defer the question of how to determine causal dependence between different factors having the same time index.

So, let i and j be any factors in I. Of course, we make the natural assumption that, if $\tau_i < \tau_j$, then i is *causally* independent of j; we are not attempting to account for the weird possibilities physics is facing today. And clearly, we would not be irritated by discovering that i is nevertheless *stochastically* dependent on j; j may furnish us with additional information about the earlier i.

We therefore assume for the next paragraphs that $\tau_i > \tau_j$. When would we say that i causally depends on j, i.e., that $i \perp_c j$ is false? The first guess would be: "When $i \perp j$ is false." But we already know from our Mickey Mouse example that this answer is deficient; there might be a factor $k \in I$ with $\tau_k < \tau_j$ such that $i \perp j/k$, showing the supposed influence of j on i to be spurious. To exclude this possibility, we obviously have to require for causal dependence between i and j that not only not $i \perp j$, but also not $i \perp j/k$ for every $k \in [\tau_j] = I_{<\tau_j}$. But this is not yet sufficient, since the argument for spurious influence may still be raised in a more subtle way: There might be two factors $k, l \in [\tau_j]$ such that $i \perp j/\{k, l\}$. In this case, k and l are, so to speak, individually not strong enough to uncover spuriousness, but jointly they are. Continuing this line of reasoning, we seem to be driven to the condition that not $i \perp j/L$ for all subsets L of $[\tau_j]$.

However, this result looks a bit strong. So let us think over the opposite case. When would we say that i is causally independent of j, i.e., that $i \perp_c j$ holds? Again, we should not answer: "When $i \perp j$." For there might be a third factor, as in our psychoanalysis example, revealing a hidden influence; that is, restricting the discussion for the moment to those possibly interfering variables which are earlier than j, we might find a $k \in [\tau_j]$ such that $i \perp j/k$ does not hold. Therefore to require for $i \perp_c j$ that not only $i \perp j$, but also $i \perp j/k$ for every $k \in [\tau_j]$, is again not sufficient. There might still be two factors $k, l \in [\tau_j]$ such that $i \perp j/\{k, l\}$ is false, and then we could not deny the presence of a particularly well hidden influence of j on i. And as before, we end up with the condition that $i \perp j/L$ for all $L \subseteq [\tau_j]$ is necessary for $i \perp_c j$.

The upshot of these two lines of reasoning is this: Whatever might be suggested about the causal independence of i from j by the truth or falsity of $i \perp j/L$ for some $L \subseteq [\tau_j]$, it is refuted by the falsity or truth, respectively, of $i \perp j/M$ for some larger $M \subseteq [\tau_j]$. Hence, only the largest subset of $[\tau_j]$, i.e., $[\tau_j]$ itself, is decisive, that is; Whether $i \perp_c j$ holds or not depends solely on whether $i \perp j/[\tau_j]$ holds or not.

So far we have only considered factors earlier than j as possibily interferring. But it is obvious from our second and third example that the factors temporally between j and i are relevant as well. So let us suppose that not $i \perp j/[\tau_j]$; at the present state of discussion this suggests that not $i \perp_c j$. Not surprisingly, there might then be a $k \in I$ with $\tau_j < \tau_k < \tau_i$ such that $i \perp j/[\tau_j] \cup \{k\}$. This, however, does not constitute an argument for the influence being spurious, because of the temporal relations assumed; rather, the situation is like that of our Rolling Stones example, and we have to conclude that the influence of j on i is still present, but screened off or shielded by k, i.e., that $i \perp_s j$.

Sticking for the moment to the concept of shieldability, we still have to face the possibility that there is another factor $l \in I$ with $\tau_j < \tau_l < \tau_i$ such that not $i \perp j/[\tau_j] \cup \{k, l\}$. In this case we might perhaps still claim that k screens off the influence of j on i, but it does not do so effectively, since by taking l into account the hidden influence of j on i appears again; the screen erected by k is not *stable*, and therefore it would be inappropriate to say that i is shieldable from j. Again we are forced to consider larger and larger conditioning sets, coming thereby to the conclusion that the only stable screen which can in no case be overthrown consists of all factors in I temporally between j and i. Since $[\tau_j]$ plus this maximal screen is equal to $[\tau_i]$, we have arrived at a first result: namely that

(3.1) $\qquad i \perp_s j$ iff $i \perp j/[\tau_i]$.

Now it is easy to complete the explication of causal independence. Before considering the factors temporally between j and i we conjectured that $i \perp_c j$, if $i \perp j/[\tau_j]$. But it is clear from the psychoanalysis example and the above discussion, that any L with $[\tau_j] \subseteq L \subseteq [\tau_i]$, for which $i \perp j/L$ is false, gives rise to an argument for the presence of a hidden influence of j on i and against $i \perp_c j$. The point now is that an argument basing on such a set L can no longer be refuted, since stochastic independence conditional on some larger set only indicates shieldable, but not spurious causal dependence, as we have just seen. Therefore we have to require that there is no such L, i.e.,

(3.2) $\qquad i \perp_c j$ iff $i \perp j/L$ for every L with $[\tau_j] \subseteq L \subseteq [\tau_i]$.

Observe that according to (3.1) and (3.2) $i \perp_c j$ implies $i \perp_s j$, as ought to be expected. If there is no influence whatsoever of j on i, there should all the more be no unshieldable influence of j on i. Of course, $i \perp_c j$ is stronger than $i \perp_s j$. We might say, and I think this is also intuitively convincing, that causal independence holds when the empty screen is already stable.

Let us pause for a moment and reconsider. Do we really have to face this almost endless moving to and fro between arguments for a spurious or shieldable influence and arguments for a hidden influence? I think, in principle there is no getting round that. In reality, this to and fro may be unlikely to occur more than two or three times, and relying on this it may be admissible to jump quickly from incomplete probabilistic data to causal conclusions. But the point is that this would be bound to be a risky jump and not a safe step. It is essential to recognize that there are not just the three types of argument we exemplified, but that they develop the characteristic interplay shown above, which drove us to consider ever larger conditioning sets and finally to (3.1) and (3.2).

But we may ask the opposite question: Have we really considered all conceivable probabilistic arguments for and against causal independence and shieldability? Strictly speaking, the answer is no for two reasons. Firstly, we have completely neglected the factors subsequent to i. But rightly so. If $i \perp j/L$ does not hold for some $L \subseteq I$ which is not a subset of $[\tau_i]$, we would of course only conclude that the factors in L which are later than i are informationally relevant to i, but we would see no reason for changing our views about the causal relations between i and j.

Secondly, one might object, that we have not really considered *all* factors earlier than i, but only those contained in I. But this is no argument against our explication; this explication is of course relative to the given, well-defined conceptual framework $(\langle \Omega^i, \mathfrak{A}^i \rangle)_{i \in I}$, and I see no way to avoid this relativity, I would not know how to get along with an indefinite set of really *all* factors. However, it might constitute an objection against a specific choice of a conceptual framework. In any application this choice might be so inadequate that unacceptable causal relations ensue within this choice. In fact, we have to reverse this: that the ensuing causal relations conform sufficiently to our preconceptions, be they pretheoretic or highly theoretic in origin, is an important criterion for the adequacy of our conceptual framework. One such preconception, e.g., is the principle of causality, which is still regulative for the most of science (and a weak form of which may be formulated within our frame as saying that each factor causally depends on its past). Or to give a more concrete example: Whenever we have discovered what we think to be the symptoms of some desease, we seek a factor (e.g., some bacterium or virus) causally responsible for these symptoms, though we may have not the slightest idea, what this factor consists in; and we do this simply because this factor must be there according to our medical knowledge.

But let us return to our explication. We have still to free our findings (3.1) and (3.2) from the simplifying restrictions stated at the beginning of this discussion. So let us first admit that several factors are associated with the same time index. We then face two problems. The minor one is this: may a factor causally depend on itself? This is a strange question, and the simplest way to cope with it is, I think, to declare it meaningless. Thus, I shall stipulate that $K \perp_c L$ makes sense only if $K \cap L = \emptyset$; the same applies to $\perp_s$, of course. However, other conventions might be adopted as well; nothing hinges on this issue.

But the main problem naturally is: Should we allow one factor to be influenced by another factor carrying the same time index? This is a delicate question, but I will not go into it with due thoroughness. Let it suffice to say the following: As far as I know, the question is still at issue, no unanimity has yet emerged. Moreover, notice that two factors may get the same time index without being strictly simultaneous; this is simply due to the coarse, i.e., discrete time structure we have assumed, and it suggests a liberal handling of the case. In view of this, I take the liberty of choosing the technically more convenient alternative, which turns out to be the liberal one. More precisely, for $\tau_i \geqslant \tau_j$ I take the factors k with $\tau_k \leqslant \tau_j$ as relevant for judging an influence of j on i as spurious, and those factors k with $\tau_j < \tau_k \leqslant \tau_i$ as relevant for judging an influence of j and i as shieldable. If we recall how the notation $[t]$ was defined, we see that (3.1) and (3.2) are still appropriate for all $i, j \in I$ with $i \neq j$ and $\tau_i \geqslant \tau_j$. I must confess, however, that the theorems stated in Section 4 depend heavily on this ruling of the matter; without it the results holding when no two factors have the same time index could not be extended to the more general case considered now.

As the last step of our analysis we have to explain our relations $\perp_c$ and $\perp_s$ generally for *sets* of factors. So let us first take up causal independence and deal with the special case of a single factor being independent of a set of factors. To simplify matters further, we shall focus on the statement $i \perp_c \{j, k\}$. This brings in something new only if both j and k are not later than i *and* if $\tau_j \neq \tau_k$, since in case that $\tau_j = \tau_k$ exactly the same reasoning as before may be carried through for $\{j, k\}$ instead of j. Thus let us assume that $\tau_k < \tau_j$. Then two different ways of generalizing (3.2) to this case present themselves: namely that $i \perp_c \{j, k\}$ is to mean that $i \perp \{j, k\}/L$ either (a) for every L with $[\tau_k] \subseteq L \subseteq [\tau_i]$ or (b) for every L with $[\tau_j] \subseteq L \subseteq [\tau_i]$. Unfortunately, neither possibility will do.

(a) is too strong. For suppose that $i \perp \{j, k\}/[\tau_k]$ does not hold. Would we accept this as conclusive evidence against $i \perp_c \{j, k\}$? No; the apparent influence if j and k on i may be due to j and then still prove spurious when new factors between k and j are added to the conditioning set. Only if $i \perp k/[\tau_k]$ does also not hold, would we conclude that at least part of the influence of j and k on i is due to k and that this part may only be screened off, but not proved spurious.

And (b) is too weak. For suppose that $i \perp k/[\tau_k]$ is false. Then, according to our earlier considerations, there is an influence of k on i; and condition (b) only asserts that this part of the influence of j and on i may be screened off, but it does not prevent it altogether.

These considerations clearly show how to weaken (a) and to strengthen (b): namely by maintaining $i \perp_c \{j, k\}$ just in case that $i \perp \{j, k\}/L$ for every L with $[\tau_j] \subseteq L \subseteq [\tau_i]$ (this is identical with (b)) *and* $i \perp k/L$ for every L with $[\tau_k] \subseteq L \subseteq [\tau_i]$ (this is weaker than (a)). Generalized to arbitrary sets of factors, this means that for all $i \in I$ and $K \subseteq I$

(3.3) $i \perp_c K$ iff for all $t \leqslant \tau_i$: $i \perp K_{\leqslant t}/L$ for every L with $[t] \subseteq L \subseteq [\tau_i]$.

The converse case of a set of factors being causally independent from a single factor is quite similar. Again, let us first analyse the simple statement that $\{k, i\} \perp_c j$, where we may assume $\tau_k > \tau_i \geqslant \tau_j$ without loss of generality. As before, there are two obvious ways of generalizing (3.2): namely by understanding $\{k, i\} \perp_c j$ as saying that $\{k, i\} \perp j/L$ either (a) for every L with $[\tau_j] \subseteq L \subseteq [\tau_k]$ of (b) for every l with $[\tau_j] \subseteq L \subseteq [\tau_i]$. But (a) is too strong, since it requires i to be stochastically independent conditional on factors partly lying in the future of i. And (b) is too weak, since it does not exclude all possibilities for a hidden influence of j on k to show up. Again the remedy is clear: We have to require that (b) holds as well as a weakened form of (a), namely that $k \perp j/L$ for every L with $[\tau_j] \subseteq L \subseteq [\tau_k]$. In full generality this means that for all $j \in I$ and $J \subseteq I$

(3.4) $J \perp_c j$ iff for all $t' \geqslant \tau_j$: $J_{\geqslant t'} \perp j/L$ for every L with $[\tau_j] \subseteq L \subseteq [t']$.

If we finally combine (3.3) and (3.4), our efforts are rewarded by a first result:

DEFINITION 3. For all $J, K \subseteq I$ such that $J \cap K = \emptyset$, J is *causally independent of* K, i.e., $J \perp_c K$, if and only if for all $t, t' \in T$ with $t \leqslant t'$ $J_{\geqslant t'} \perp K_{\leqslant t}/L$ for every L with $[t] \subseteq L \subseteq [t']$ (that is, for every L with $I_{\leqslant t} \backslash (J_{\geqslant t'} \cup K_{\leqslant t}) \subseteq L \subseteq I_{\leqslant t'} \backslash (J_{\geqslant t'} \cup K_{\leqslant t})$).

For generalizing the concept of shieldability as fixed in (3.1), we may proceed in a similar fashion: The first step is quite straightforward and yields for $i \in I$ and $K \subseteq I$:

(3.5) $\quad i \perp_s K$ iff $i \perp K_{\leqslant \tau_i}/[\tau_i]$.

Here we have neglected $K_{>\tau_i}$ and used the largest screen available, as we have done in (3.1).

The other step does not run so smoothly. So let us consider the simple $\{k, i\} \perp_s j$, where of course $\tau_k, \tau_i \geqslant \tau_j$. If $\tau_k = \tau_i$, then (3.1) stated for $\{k, i\}$ instead of i is still adequate; the former reasoning applies here as well. If $\tau_k > \tau_i$, however, we again face two obvious generalizations of (3.1): (a) $\{k, i\} \perp_s j$ iff $\{k, i\} \perp j/[\tau_k]$, and (b) $\{k, i\} \perp_s j$ iff $\{k, i\} \perp j/[\tau_i]$. As before, both are inappropriate. (a) assumes i to be stochastically independent of j conditional on future factors (relative to i), and (b) provides only a possibly unstable screen for k. Since we have to employ the maximal screen for every temporal segment of the shieldable set $\{k, i\}$, the only way out of this quandary is to require that $\{k, i\} \perp_s j$ iff $i \perp j/[\tau_i]$ and $k \perp j/[\tau_k]$. The immediate generalization of this to arbitrary $J \subseteq I$ is:

(3.6) $\quad J \perp_s j$ iff for all $t \geqslant \tau_j$ $J_{=t} \perp j/[t]$.

Combining (3.5) and (3.6), we get our second result:

DEFINITION 4. For all $J, K \subseteq I$ such that $J \cap K = \emptyset$, J is *shieldable from* K, i.e., $J \perp_s K$, if and only if for all $t \in T$ $J_{=t} \perp K_{\leqslant t}/[t]$ (i.e., $J_{=t} \perp K_{\leqslant t}/I_{\leqslant t} \backslash (J_{=t} \cup K_{\leqslant t})$).

You may complain that Definitions 3 and 4 are much too complicated to be grasped immediately. Granted, but this is, I think, partly due to the fact that causal independence (and shieldability) of arbitrary sets of factors is intuitively not so clear anyway. Moreover, I hope to have provided intuitive access at least to (3.1) and (3.2), and the definitions are merely consequential to (3.1) and (3.2), indeed uniquely consequential in my view. Additional support for our explication and further insight may be gained by turning to the theorems holding for the concepts so defined.

4. CAUSAL INDEPENDENCE AND SHIELDABILITY: PROPERTIES

The central properties of $\perp_c$ and $\perp_s$ as defined by Definitions 3 and 4 are stated in:

THEOREM 3. Suppose P to be strictly indeterministic and let $(J_\delta)_{\delta \in \Delta}$ and $(K_\delta)_{\delta \in \Delta}$ be two (possibly uncountable) families of subsets of I, $J = \bigcup_{\delta \in \Delta} J_\delta$, and $K = \bigcup_{\delta \in \Delta} K_\delta$. Then we have:

(a) if for all $j \in J$ and $k \in K$ $\tau_j < \tau_k$, then $J \perp_c K$;

(b) if $J \perp_c K$, $J' \subseteq J$, and $K' \subseteq K$, then $J' \perp_c K'$;

(c) if for all $\delta \in \Delta$ $J_\delta \perp_c K$, then $J \perp_c K$;

(d) if for all $\delta \in \Delta$ $J \perp_c K_\delta$, then $J \perp_c K$;

and the same holds for $\perp_s$ instead of $\perp_c$.

Again, the uncountable case is somewhat superfluous in view of the discrete time assumed; but that does not matter. The most obvious facts about the connection between $\perp_c$ and $\perp_s$ are stated in:

THEOREM 4. For all $J, K \subseteq I$ we have:

(a) if $J \perp_c K$, then $J \perp_s K$;

(b) if $J \perp_s K$ and if for all $i \in I \setminus (J \cup K)$, $j \in J$ and $k \in K$ $\tau_k \leqslant \tau_j$ and either $\tau_i \leqslant \tau_k$ or $\tau_j < \tau_i$, then $J \perp_c K$.

Theorem 5 will establish a more subtle connection between $\perp_c$ and $\perp_s$.

How desirable are the properties of $\perp_c$ and $\perp_s$ expressed in Theorems 3 and 4? Did we expect more? Well, the properties 3(a) and 3(b) are utterly indispensible for both $\perp_c$ and $\perp_s$. 3(c) and 3(d) relate to our discussion at the end of Section 2; they state that conjunctivity does indeed hold for $\perp_c$ and for $\perp_s$, as demanded by that discussion. Thus, $\perp_c$ and $\perp_s$ have the simple properties that are intuitively required.

In fact, in the light of Theorem 3 we could have chosen another approach to Definitions 3 and 4. We could have extended (3.1) and (3.2) to sets of factors simply by assuming conjunctivity outright, i.e., by defining $J \perp_c K$ to mean that for all $j \in J$ and all $k \in K$ $j \perp_c k$ (and likewise for $\perp_s$). Theorem 3 then says that this definition is equivalent to ours. This way would have been easier to grasp, but it cannot give us a *proof* of conjunctivity. Ultimately, I think, it is a matter of taste which way one chooses.

Theorem 4(a) is of course a must, and 4(b) may be read in an intuitively quite compelling way, namely as saying: if J can be screened off from K and if there are no factors in the temporal region occupied by J and K except those of J and K themselves, then it must be just the empty screen which screens off J from K; and as remarked previously, the stability of the empty screen amounts to causal independence.

Is there some property missing? If any at all, it will probably be transitivity of causal dependence. (I can think of nothing else.) At least it should be true, one might say, that if i is not shieldable from j and j is not shieldable from k, then i causally depends on k. But even this weak transitivity does not hold (as simple numerical counterexamples show). That this transitivity cannot be expected within a probabilistic framework is well-known (cf., e.g., Suppes (1970, pp. 58f.)), and there is a simple general reason for this: Define the *causal range* of a factor i to be the set of the time indices of all the factors which causally depend on i. Within a strictly deterministic framework it may be reasonable to expect the causal range of a factor to extend infinitely into the future. But within a probabilistic framework this is not plausible at all, simply because it may happen that more and more randomizing elements superpose such that the causal influence of a factor blurs more and more, until it fades entirely. Hence we must reckon at least some factors to have a finite causal range, and this contradicts transitivity of causal dependence.

So far we have always stressed that stochastic and causal independence, though interrelated, are very different relations. The following consideration will bring them closer together again:

It is well known from the construction of product measures, that, roughly speaking, a probability measure on some product σ-algebra partitions this σ-algebra into stochastically independent sub-σ-algebras, between which there is no probabilistic dependence whatsoever (and that it can be reconstructed from its restrictions to these sub-σ-algebras). To be precise, let us define:

DEFINITION 5. $(J_\delta)_{\delta \in \Delta}$ is a $\perp$-*partition of* I if and only if $(J_\delta)_{\delta \in \Delta}$ is a partition of I (i.e., $J_\delta \neq \emptyset$, $J_\delta \cap J_\gamma = \emptyset$ for $\delta \neq \gamma$, and $\bigcup_{\delta \in \Delta} J_\delta = I$) and if for all $\delta \in \Delta$ $J_\delta \perp I \backslash J_\delta$.

Of course, it is tempting to look for the causal analogue, and it is clear how it is to be defined with respect to causal independence as well as to shieldability:

DEFINITION 6. $(J_\delta)_{\delta \in \Delta}$ is a $\perp_c(\perp_s)$-*partition of* I if and only if $(J_\delta)_{\delta \in \Delta}$ is a partition of I and if for all $\delta \in \Delta$ $J_\delta \perp_c I \backslash J_\delta (J_\delta \perp_s I \backslash J_\delta)$.

Naturally, we are interested in the relationships between these different kinds of partitions. This interest is exhaustively answered within our restricted frame by:

THEOREM 5. If P is strictly indeterministic, the following three statements are equivalent:

(a) $(J_\delta)_{\delta \in \Delta}$ is a $\perp$-partition of I;

(b) $(J_\delta)_{\delta \in \Delta}$ is a $\perp_c$-partition of I;

(c) $(J_\delta)_{\delta \in \Delta}$ is a $\perp_s$-partition of I.

Hence all three kinds of partitioning amount to the same thing. Actually, the equivalence of (b) and (c) should have been expected, as the following consideration shows: Define two factors i and j to be *causally connected* if and only if there are factors $k_1, \ldots, k_n$ such that $k_1 = i$, $k_n = j$, and either not $k_{m+1} \perp_c k_m$ or not $k_m \perp_c k_{m+1}$ for every $m = 1, \ldots, n-1$. Obviously causal connectedness is an equivalence relation within I, and with the help of Theorem 3 it can be easily checked that the equivalence classes of this relation form a $\perp_c$-partition, in fact the finest $\perp_c$-partition, of I. In this reading Theorem 5 affirms the intuitively felt fact that it makes no difference to causal connectedness whether it is based on bare causal dependence or only on direct, unshieldable causal dependence. This also reconciles us perhaps a bit to the fact that even weak transitivity of causal dependence in the above sense does not hold.

As to equivalence of (a) and (b), intuition is more neutral. I think; all the more, it is welcome. It establishes a simple and useful connection between causal and probabilistic concepts, partly justifying, partly explaining their careless interchange.

It may be illuminating to apply our results to two of the most widely discussed situations in probability theory: sequences of independent random variables and discrete Markov processes. Let N be the set of all positive integers. In both cases we have to deal with a sequence $(X_n)_{n \in N}$ of random variables defined for some probability space. This means that we may put $I = T = N$. In the case of independent random variables each X_n is stochastically independent of all others, i.e., in our terms we have $n \perp N \setminus \{n\}$ for all $n \in N$. Hence, Theorem 5 applies, and we may infer that in this case, and only in this case, the X_n are also causally independent from each other. This in a sense justifies the frequentists mentioned in the introduction, who, attempting to establish a frequentist view of probability with the help of the law of large numbers for sequences of independent random variables, needed some nonprobabilistic grounds for assuming stochastic independence of the random variables and hence inferred stochastic from causal independence only in the case sanctioned now.

In the case of a discrete Markov process, the distribution of any X_n conditional on its immediate past is the same as the one conditonal on its entire past, i.e., in our terms we have $n + 1 \perp \{1, \ldots, n-1\}/n$ for all $n \in N$. This means according to our explication that in a Markov process each X_{n+1} is shieldable from $\{X_1, \ldots, X_{n-1}\}$, and that is exactly what Markov processes are designed for, as is also indicated by the usage of paraphrasing Markov processes as "memoryless" or "nonhereditary" or as characterized by the "absence of after-effect". No general statement is possible about causal independence within a Markov process, but usually every X_n will causally depend on every X_m with $m < n$.

Let us finally discuss the two restrictions governing this paper and first take up strict indeterminateness of P. Its mathematical origin is clear; without it Theorem 1(e) would not hold, and from there it spreads to the causal theorems, i.e., to be precise, to Theorem 3(d) for $\perp_c$ and for $\perp_s$ and, less importantly, to Theorem 3(c) for $\perp_s$ and Theorem 5 only in case there are several factors with the same time index.
But there is, I think, also good intuitive reason for assuming it. If we were to generalize our explication to probability measures that are not strictly indeterministic or, as we might say, weakly deterministic, all sorts of awkward things would happen, all of which relate to the invalidity of Theorem 3(d). The basic case is this: Suppose there are three factors i, j, and k such that $\tau_k < \tau_j < \tau_i$, $i \perp j/k$, and $i \perp k/j$. This means according to our explication that i is causally independent of j (because of k) and that i is shieldable from k (by j). Normally we would say that, if k is shielded by something causally independent, this must be a trivial case of shielding where k itself is causally independent. And this is exactly what follows if P is strictly indeterministic. But for weakly deterministic P, $i \perp \{j, k\}$

may nevertheless be false. Imagine, e.g., that $B_1, \ldots, B_n$ and $C_1, \ldots, C_n$ are the atoms of $\mathfrak{A}_j$ and $\mathfrak{A}_k$, respectively, that for all $m = 1, \ldots, n$, $P(B_m | C_m) = P(C_m | B_m) = 1$, and that i somehow stochastically depends on j, and on k in the same way. This case is not clear regarding its causal relations, and it therefore gives no hint as to how to modify our definitions for weakly deterministic P.

But this unclarity is not surprising, I think, since it seems to me that this modification, whatever it may look like in detail, cannot be carried through within the standard probabilistic framework. For how would we judge the causal relations in the example above? We should somehow know the probabilities of the states of i given that C_p and nevertheless B_q (where $q \neq p$) has occured. If these probabilities are the same as those conditional on C_p alone irrespective of B_q, we would suppose j to be causally irrelevant to i; and if these probabilities only depend on B_q, i is presumably shieldable from k. But the point is that standard probability measures cannot give us these probabilities, because probabilities conditional on null events are not defined. This is the deeper reason for our restriction to strictly indeterministic probability measures: Only then do we have all the conditional probabilities needed for safely inferring causal relations. But this also leads me to suspect that we have to resort to something like Renyi's conditional probability spaces (cf. Renyi (1973)) for handling the weakly deterministic case.

What is there to say about the restriction to a discrete time structure? Note firstly that it is pragmatically implied by the strict indeterminateness of P; that is, in almost all interesting cases with continuous time the probability measure involved will be weakly deterministic. This is so simply because scientists have found that applications of a continuous stochastic process require, in general, that almost all of its paths be almost everywhere continuous; and such a process can only be described by a weakly deterministic probability measure. This is not to say that to cover continuous time we only have to solve the weakly deterministic case. If this were so, our theorems would have to hold for continuous time as well. But they do not; discreteness of time proves to be essential for Theorem 4(b) and for Theorem 5.

Again, the mathematical source of trouble is clear: By repeated application of Theorem 1(d) we can roll up any discretely ordered set of factors; for instance, if we have $J_{=t} \perp K/J_{<t}$ for all, discretely many t, we may infer $J \perp K$. But this does not carry over to the continuous case.

Intuitively, I cannot find anything wrong with our explication of causal independence nor, concerning shieldability, with (3.1) and (3.5). But (3.6) cannot be sustained if time is continuous. Consider, e.g., a continuous Markov process $(X_t)_{t \in T}$. According to (3.5) every X_t $(t > t_0)$ is shieldable from $\{X_t | t \leqslant t_0\}$, as it should be. But if (3.6) would be correct, then the whole $\{X_t | t > t_0\}$ too would be shieldable from $\{X_t | t \leqslant t_0\}$, and this is obviously wrong.

However, I will not speculate now about feasible adjustments of (3.6) to the continuous case. It is all too clear, that all the problems touched upon in the last paragraphs call for a separate, careful study.

5. PROOFS

The proofs given here are somewhat succinct. In particular, not all applications of Theorem 1 are explicitly mentioned. But I think the existing lacunae may be filled without difficulty. Let me also note at the outset that in every proof the notation of the corresponding theorem is employed unless otherwise stated.

As to Theorem 1: I did not find it stated in just this form, but it will only require reformulating well-known things. So I may confine myself to giving some hints starting from what may be found in usual textbooks such as Loève (1960): Denoting the conditional probability of P relative to $\mathfrak{A}_M$ simply by P^M instead of $P^{\mathfrak{A}_M}$, a first thing to note is that for all $K, L, M \subseteq I$ the following holds, as Loève (1960, Section 25.3A), shows:

(5.1) $$P^M(A \cap B) = P^M(A) \cdot P^M(B), P\text{-a.s. for all } A \in \mathfrak{A}_K \text{ and } B \in \mathfrak{A}_L$$

iff

(5.2) $$P^{L \cup M}(A) = P^M(A), P\text{-a.s. for all } A \in \mathfrak{A}_K.$$

Then, by stating premises and conclusions in the form of (5.2), and, if necessary, by applying the smoothing properties of conditional expectation (cf. Loève (1960, Section 25.2)), Theorem 1(a), (b), (c), and (d) prove to be trivial.

Proof of Theorem 1(e). Using the equivalence of (5.1) and (5.2), the premises say that for all $A \in \mathfrak{A}_J$, $P^{K \cup L \cup M}(A) = P^M(A)$ P-a.s. and $P^{K \cup L \cup M}(A) = P^L(A)$ P-a.s. Now, *if* there exists an $\mathfrak{A}_{L \cap M}$-measurable version of $P^M(A)$, we have $P^M(A) = P^{L \cap M}(A)$ P-a.s. and thus $P^{K \cup L \cup M}(A) = P^{L \cap M}(A)$ P-a.s. as desired. We have therefore to show that there is an $\mathfrak{A}_{L \cap M}$-measurable version of $P^M(A)$:

Let $D = \{P^M(A) \neq P^L(A)\}$. Obviously, we have $D \in \mathfrak{A}_{L \cup M}$ and $P(D) = 0$. Let $\Omega_{L \cap M}$ be the set of the restrictions of all $\omega \in \Omega$ to $L \cap M$, $N = (L \cup M) \backslash (L \cap M)$, and for each $\omega \in \Omega_{L \cap M}$ D_ω the section of D at ω; clearly, $D_\omega \in \mathfrak{A}_N$ for each $\omega \in \Omega_{L \cap M}$. Moreover, (assuming $P^{L \cap M}$ to be regular – this is an additional premise) we have $0 = P(D) = \int [P^{L \cap M}(D_\omega)](\omega)\, \mathrm{d}P(\omega)$ (cf. Loève (1960, Section 8.2 and 8.3)); hence $[P^{L \cap M}(D_\omega)](\omega) = 0$ for P-almost all $\omega \in \Omega_{L \cap M}$.

Now, strict indeterminateness of P comes in. It implies that for each nonnull $E \in \mathfrak{A}_N$, $P^{L \cap M}(E) > 0$ P-a.s. As indicated immediately after Definition 1, we must slightly strengthen strict indeterminateness so as to have a null set $F \in \mathfrak{A}_{L \cap M}$ such that for all nonnull $E \in \mathfrak{A}_N$, $P^{L \cap M}(E) > 0$ outside F. (There would have been no point in introducing this more complicated definition in the non-technical parts of the paper.) P being strictly indeterministic in this strengthened sense, we may infer that $P(D_\omega) = 0$ for P-almost all $\omega \in \Omega_{L \cap M}$.

Take now any such D_ω, and let for each $x \in [0, 1]$, $B_x \in \mathfrak{A}_{L \backslash M}$ be the section of $\{P^L(A) < x\}$ at ω and $C_x \in \mathfrak{A}_{M \backslash L}$ the section of $\{P^M(A) > x\}$ at ω. Thus, for each $x \in [0, 1]$, $B_x \cap C_x \subset D_\omega$ and $P(B_x \cap C_x) = 0$. Hence, by strict indeterminateness again, either $P(B_x) = 0$ or $P(C_x) = 0$ for each $x \in [0, 1]$. This implies that the section of $P^M(A)$ at ω is P-a.s. constant and might therefore be chosen completely constant. From this it follows at last that there is an $\mathfrak{A}_{L \cap M}$-measurable version of $P^M(A)$. □ (I.e., end of proof.)

So we may turn to Theorem 2; in its proof we shall denote $\bigcup_{\delta \in \Gamma} M_\delta$ by M_Γ for all $\Gamma \subseteq \Delta$.

Proof of Theorem 2(a). The premise says that for all $\Gamma \in \mathfrak{E}(\Delta)$ and $A \in \mathfrak{A}_{M_\Gamma}$, $P^{K \cup L}(A) = P^L(A)$ P-a.s. But it is clear from the construction of infinite product σ-algebras that this holds then even for all $A \in \mathfrak{A}_{M_\Delta}$. □

Proof of Theorem 2(b). The premise says that for all $\Gamma \in \mathfrak{E}(\Delta)$ and $A \in \mathfrak{A}_K$, $P^{L \cup M_\Gamma}(A) = P^{M_\Gamma}(A)$ P-a.s. Now we may conceive of $(P^{M_\Gamma}(A))_{\Gamma \in \mathfrak{E}(\Delta)}$ as a martingale indexed by the directed index set $\mathfrak{E}(\Delta)$. Theorems for martingales with a directed index set (cf. Neveu (1972, pp.

95–100)) then tell us that this martingale converges to $P^{M_\Delta}(A)$ in mean. But by assumption, the $P^{L \cup M_\Gamma}(A)$ form the same martingale. Therefore it also converges to $P^{L \cup M_\Delta}(A)$ in mean. Hence $P^{L \cup M_\Delta}(A) = P^{M_\Delta}(A)$ P-a.s. □

Theorem 2(c) may be proved like 2(b) by using the corresponding theorems for reversed martingales (cf. Neveu (1972, pp. 115–119)). Finally, Theorem 2(d) may be easily obtained from Theorem 1(c), 1(e), and 2(c).

Moving to the causal theorems, we see first that Theorem 3(a) is completely trivial for $\perp_c$ and $\perp_s$ as defined in Definitions 3 and 4. Theorem 3(b) for $\perp_s$ follows immediately from Theorem 1(c). The only non-trivial part of Theorem 3(c) for $\perp_s$ is the case where all J_δ assemble at one single time index; but in this case 3(c) for $\perp_s$ is a straight application of 2(d). And Theorem 3(d) for both, $\perp_s$ and $\perp_c$, is an equally direct consequence of 2(d).

Proof of Theorem 3(b) for $\perp_c$. Let $t, t' \in T$ such that $t \leqslant t'$. Then the premise says that $J_{\geqslant t'} \perp K_{\leqslant t}/L$ for every relevant L. (It is not necessary always to write out the somewhat tedious condition for these L.) From this we want to conclude that $J'_{\geqslant t'} \perp K'_{\leqslant t}/M$ for every M relevant to the conclusion of 3(b). But since for every such M there is an L relevant to the premise such that $L \subseteq M \subseteq L \cup J_{\geqslant t'} \cup K_{\leqslant t}$, we get this immediately by applying Theorem 1(c). □

Proof of Theorem 3(c) for $\perp_c$. By definition, we have as premise that for all $t^*, t^{**} \in T$ with $t^* \leqslant t^{**}$, all $\delta \in \Delta$, and all relevant L

$$(J_\delta)_{\geqslant t^{**}} \perp K_{\leqslant t^*}/L. \tag{5.3}$$

Choose now $t', t'' \in T$ and $M \subseteq I$ arbitrarily such that $t' \leqslant t''$ and $I_{\leqslant t'} \backslash (J_{\geqslant t''} \cup K_{\leqslant t'}) \subseteq M \subseteq I_{\leqslant t''} \backslash (J_{\geqslant t''} \cup K_{\leqslant t'})$. Then we have to show that $J_{\geqslant t''} \perp K_{\leqslant t'}/M$: Abbreviating $(J_\delta)_{=t}$ by $J_{\delta, t}$ for all $t \geqslant t''$, we have

$$J_{\geqslant t''} = \bigcup_{\substack{\delta \in \Delta \\ t \geqslant t''}} J_{\delta, t}.$$

According to Theorem 2(a) it suffices therefore to prove that for all finite sequences $J_{\delta_1, t_1}, \ldots, J_{\delta_n, t_n}$

$$\bigcup_{r=1}^{n} J_{\delta_r, t_r} \perp K_{\leqslant t'}/M,$$

and this may be done in the following way:

Let us take any such sequence, where we may assume without loss of generality that $t_1 \leqslant \ldots \leqslant t_n$, and let us for the moment write J_r instead of the tedious J_{δ_r, t_r} $(r = 1, \ldots, n)$. By (5.3) we have $J_1 \perp K_{\leqslant t'}/M$, of course. Planning a proof by induction, we now suppose that

$$\bigcup_{r=1}^{m} J_r \perp K_{\leqslant t'}/M.$$

An appropriate specialization of (5.3) (namely $t^* = t'$ and $t^{**} = t_{m+1}$) gives us

$$J_{m+1} \perp K_{\leqslant t'}/M \cup \bigcup_{r=1}^{m} J_r.$$

Both together imply

$$\bigcup_{r=1}^{m+1} J_r \perp K_{\leqslant t'}/M$$

by Theorem 1(d). This completes the proof. □

Theorem 4(a) is completely trivial.

Proof of Theorem 4(b). Since time is discrete, we may assume that there is a smallest $t \in T$ such that $J_{=t} \neq \emptyset$. (Otherwise $K = \emptyset$, and nothing has to be proved.) So let $t_1, t_2, \ldots$ be just those $t \in T$ for which $J_{=t} \neq \emptyset$, where $t_m < t_n$ for $m < n$. Because of our assumptions about J and K our premise may be written in the form:

$$J_{=t_n} \perp K/(I_{\leqslant t_1} \backslash (J \cup K)) \cup \bigcup_{m=1}^{n-1} J_{=t_m}$$

for all n. By successively applying Theorem 1(d), we get from this

$$\bigcup_{m=1}^{n} J_{=t_m} \perp K/I_{\leqslant t_1} \backslash (J \cup K)$$

for all n, and from this $J \perp K/I_{\leqslant t_1} \backslash (J \cup K)$. (If J extends over infinitely many time indices, we have to use Theorem 2(a) for the last step.) And using Theorem 1(c) this implies that $J \perp_c K$. □

Proof of Theorem 5. (a) ⇒ (b) is obvious on account of Theorem 1(c); (b) ⇒ (c) is assured by Theorem 4(a). The final part, (c) ⇒ (a), is a bit harder: Let us assume that T is order isomorphic to the set Z of all integers, i.e., that $T = \{t_n | n \in Z\}$, where $t_m < t_n$ iff $m < n$. (If T is finite or order isomorphic to the set of either all positive or all negative integers, the same proof applies, as will be evident.) The premise is that for all $\delta \in \Delta$, $J_\delta \perp_s I \backslash J_\delta$. Hence with Theorem 3(b) for all $\gamma, \delta \in \Delta$ with $\gamma \neq \delta$, $J_\gamma \perp_s J_\delta$, and from this with Theorem 3(c) $I \backslash J_\delta \perp_s J_\delta$. According to Definition 4 this means that for all $\delta \in \Delta$ and $n \in Z$

$$(5.4) \qquad (J_\delta)_{=t_n} \perp (I \backslash J_\delta)_{\leqslant t_n} / (J_\delta)_{<t_n}$$

and

$$(5.5) \qquad (I \backslash J_\delta)_{=t_n} \perp (J_\delta)_{\leqslant t_n} / (I \backslash J_\delta)_{<t_n}.$$

Take now any $\delta \in \Delta$ and define $K_n = (J_\delta)_{\leqslant t_n}$ and $L_n = (I \backslash J_\delta)_{\leqslant t_n}$. Then (5.4) and (5.5) say that for all $r \in Z$:

$$(5.6) \qquad K_r \backslash K_{r-1} \perp L_r / K_{r-1}$$

and

$$(5.7) \qquad L_r \backslash L_{r-1} \perp K_r / L_{r-1}.$$

Our next step is to prove by induction that for all $m, n \in Z$ with $m > n$

$$(5.8) \qquad K_m \backslash K_n \perp L_m / K_n.$$

For $m = n + 1$ (5.8) says the same as (5.6). So assume that (5.8) holds for some $m > n$. From this and (5.6) for $r = m + 1$ we get with Theorem 1(c) and (d) that

$$(5.9) \qquad K_{m+1} \backslash K_n \perp L_m / K_n.$$

(5.7) for $r = m + 1$ implies that $L_{m+1} \backslash L_m \perp K_{m+1} \backslash K_n / K_n \cup L_m$ and hence, with (5.9) and Theorem 1(b) and (d), that $K_{m+1} \backslash K_n \perp L_{m+1} / K_n$, i.e., that (5.8) is true for $m + 1$ too. Thus we have (5.8) for all $m, n \in Z$ with $m > n$. From this we easily obtain with Theorem 1(c) that for all $m, n, r \in Z$ with $n \leqslant r < m$, $K_m \backslash K_r \perp L_m / K_n$. Hence, with Theorem 2(c), $K_m \backslash K_r \perp L_m$ for all $m, r \in Z$, and, with 2(a), $K_m \perp L_m$ for all $m \in Z$. Thus finally, $K_m \perp L_n$ for all $m, n \in Z$, and with 2(a) first $J_\delta \perp L_n$ for all $n \in Z$ and then $J_\delta \perp I \backslash J_\delta$. □

Universität München

NOTE

* I am indebted to the referee for his most valuable report and to the editor for taking more trouble than usual, I fear.

BIBLIOGRAPHY

Loève, M.: (1960), *Probability Theory*, Princeton: Van Nostrand.

Neveu, J.: (1972), *Martingales à Temps Discret*, Paris: Masson.

Reichenbach, H.: (1956), *The Direction of Time*, Berkeley: University of California Press.

Renyi, A.: (1973), *Wahrscheinlichkeitsrechnung*, Berlin: Deutscher Verlag der Wissenschaften.

Salmon, W. C.: (1971), *Statistical Explanation and Statistical Relevance*, Pittsburgh: University Press.

Salmon, W. C.: (1975), 'Confirmation and Relevance', in: G. Maxwell, R. M. Anderson Jr. (eds.), *Minnesota Studies in the Philosophy of Science*, vol VI, Minneapolis: University of Minnesota Press, pp. 3–36.

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

Van de Geer, J. P.: (1971), *Introduction to Multivariate Analysis for the Social Sciences*, San Francisco: Freeman.