We therefore find that the only obvious limitations imposed by virtue of the physical nature of predictive activities on the prediction of A by B are due to:

(i) the finite operating speed of the computer and of measurement (requiring $t_2 - t_1 > \Delta t_a + \Delta t_b + \Delta t_c$),

(ii) A not being sufficiently large to interact with some B′,

(iii) the physical impossibility of B learning of the construction of A.

Neither (ii) nor (iii) are the case if A is itself a predictor and is a finite prediction task, and (i) is a physical requirement devoid of any logical significance.

We conclude that predictors are in principle predictable, and so that there is therefore no logical reason why in a determinist universe any system should not be predictable.

Emmanuel College
Cambridge

A CAUSAL CALCULUS (I) *

I. J. Good

1 *Introduction*

This paper contains a suggested quantitative explication of probabilistic causality in terms of physical probability.[1] The main result is to show that, starting from very reasonable desiderata, there is a unique meaning, up to a continuous increasing transformation, that can be attached to ' the tendency of one event to cause another one '. A reasonable explicatum will also be suggested for the degree to which one event caused another one. It may be possible to find other reasonable explicata for tendency to cause, but, if so, the assumptions made here will have to be changed.

I believe that the first clear-cut application in science will be to the foundations of statistics, such as to an improved understanding of the function of randomisation, but I am content for the present to regard the work as contributing to the philosophy of science, and especially to what may be called the ' mathematics of philosophy '. Light may also be shed on problems of allocating blame and credit. I hope to consider applications to statistics on another occasion.[2]

In a previous note[3] I have tried to give an interpretation of 'an event F caused another event E ' without making reference to time. It was presumably clear from the last three paragraphs, which were added in

* Received 19. 1. 1960

[1] Compare : (i) Hans Reichenbach, ' Die Kausalstruktur der Welt und der Unterschied von Vergangenheit und Zukunft ', *Ber. d. Bayer. Akad. d. Wissensch., math.-nat. Abt.*, 1925, pp. 133-175; or Chapter 3 of his book *Modern Philosophy of Science*, London and New York, 1959; (ii) Norbert Wiener, ' The theory of prediction ', in *Modern Mathematics for the Engineer* (ed. by E. F. Beckenbach), New York, 1956, pp. 165-190. The present work goes further than Reichenbach's in offering a definite explicatum. It bears little resemblance to Wiener's work, which is mathematically much more advanced.

[2] The present paper owes much to correspondence and discussion with Mr E. M. L. Beale, Professor Bruno de Finetti, Professor K. R. Popper, Professor L. J. Savage, Mr Christopher Scott, and especially to Dr Oliver Penrose. The Referees and Editor have also been helpful.

[3] This *Journal*, 1959, 9, 307-310

proof,[1] that I was not satisfied with my attempt.[2] I shall describe the note as the 'previous paper' but it will not be necessary for the reader to refer back.

The present paper is more ambitious in that it is quantitative, but less so in that it assumes, at least at first, that F is earlier than E. (It may be possible to interpret the explicatum more generally.) As in the previous paper I shall take for granted the notion of physical (= material) probability. In order to avoid misunderstanding I must mention my opinion that in so far as physical probability can be measured it can be done only in terms of subjective probability, but this opinion will not affect the arguments below. Likewise the notion of an 'event' will be taken for granted. In some earlier drafts I included material dealing with the meanings of 'event', 'probability', and 'definition', and with the modifications of the analysis required in order to cope with the possibility that the future may affect the past. I have omitted this material here for the sake of brevity.

2 *Notation and General Outline*

Propositions and events will be understood in a very general sense, and will be denoted by the symbols E, F, G, H, and U. These will be combined by means of the logical connectives '.' meaning 'and', '$^-$' meaning 'not', and '**v**' meaning 'or'. A vertical stroke, '|', will mean 'given', as in the expression $P(E \mid H)$, the probability of E given H. Similarly $O(E \mid H)$ will mean that the odds of E given H, i.e. $P(E \mid H)/P(\bar{E} \mid H)$. Sometimes some or all of what is 'given' is omitted from the notation for the sake of brevity. A colon will be used to mean 'provided by' or 'by' or 'from', as in

$$I(E : F \mid G) = \log(P(E \mid F . G)/P(E \mid G)),$$

which can be read from left to right as the amount of information concerning E provided by F given G. Another example of the colon notation is

$$\begin{aligned} W(H : E \mid G) &= \log(P(E \mid H . G)/P(E \mid \bar{H} . G)) \\ &= \log(O(H \mid E . G)/O(H \mid G)) \\ &= I(H : E \mid G) - I(\bar{H} : E \mid G), \end{aligned}$$

[1] The words 'added in proof' were omitted in error, and the effect was peculiar.

[2] I find that Reichenbach made a similar error when defining the expression 'causal relevance'. See Appendix I.

the 'weight of evidence concerning H provided by E given G'.[1]

The general plan of the paper is to suggest explicata for:

(i) $Q(E : F)$, or Q for short, the 'causal support for E provided by F, or the tendency of F to cause E'. The explicatum that the argument forces upon us is the weight of evidence against F if E does not happen, $W(\bar{F} : \bar{E})$, or more explicitly, $W(\bar{F} : \bar{E} \mid U . H)$, where U and H are defined below. In order to formulate enough desiderata it is necessary to introduce some auxiliary notions.

(ii) The strength of a causal chain joining F to E.

(iii) The strength of a causal net joining F to E. (Causal chains and nets will be defined in Sections 8 and 11.)

(iv) $\chi(E : F)$, or χ for short, the contribution to the causation of E provided by F, i.e. the degree to which F caused E. This will be defined as the strength of a causal net joining F to E, when the details of the net are completely filled in, so that there are no relevant events omitted. (I avoid the use of the letter C for either Q or χ, because it has been used to mean corroboration.[2]) An example is given in Appendix II to show that Q and χ cannot be identified.

It would not be appropriate to define χ as the limit of strengths of more and more detailed nets; for, if space and time are continuous, the limiting operation could be done in a biased manner so as to get entirely the wrong result; like a lawyer making a case by special selection of the evidence. We must have the whole truth in order to define χ in principle. (Compare the first Appendix.) If, however, the events fill the relevant parts of space and time, and we let the events become smaller and smaller, then the limit should be unique.

In practical uses of the notion of causality, judgments of approximate irrelevance are always made in order to reduce the complication of the causal net.

It is possible to draw an analogy between a causal net and an electrical resistance network, with a resistance in each link. In this analogy it is necessary to imagine a rectifier placed in each link in order to prevent a flow of causal influence backward in time. It is then tempting to define the degree of causality between the input and output of

[1] See, for example, 'Weight of Evidence, Corroboration, Explanatory Power, Information, and the Utility of Experiments', *J. Roy. Stat. Soc., ser. B*, 1960, **22**, 319-331, where there are further references.

[2] Karl R. Popper, *The Logic of Scientific Discovery*, London, 1959, pp. 379-418, mostly published in this *Journal*, 1954, 1957, and 1958. See also Note **1** above.

the causal net as the effective resistance of the corresponding causal network. This analogy suggests that the causal resistances should be defined in such a manner that they are additive for chains in 'series', and such that their reciprocals are additive for chains in 'parallel'. It turns out that the analogy cannot be pressed as far as this, but it is one of the themes of the paper.

The main part of the paper consists of a list of assumptions, and deductions from them, leading up to the above explicatum for Q. Afterwards a general explicatum will be suggested for χ, but this will not be deduced in the same formal manner. It is fairly convincingly unique for causal nets of the 'series-parallel' type, and has a certain cogency in the general case.

3 *Small Events*

Until near the end of the paper all events will be assumed to occupy small volumes of space (more precisely: have small diameters) and occupy small epochs of time. For the most part 'space' could be interpreted in a more general sense than as ordinary three-dimensional space; for example, it could be phase space or Hilbert space. On the other hand time will be assumed to be well-ordered and one-dimensional. There must be some sort of metric in both space and time, in order that smallness and contiguity should have a meaning. If the metrics of space and time have been mixed up, as in the theory of relativity, then they will be assumed to be unmixed by the use of a fixed frame of reference. (Dr O. Penrose has pointed out that the present work is consistent with the theory of relativity provided that causal influence does not travel faster than light.)

4 *Causal Support, or Tendency to Cause*

Let H denote all true laws of nature, whether known or unknown, and let U denote the 'essential physical circumstances' just before F started. When we talk about 'essential physical circumstances' we imply that the exact state has a probability distribution. An equivalent description is to say that a system is one of an 'ensemble'. (I must admit that there is more than meets the eye in this description, since in quantum mechanics the word 'state' can be given at least eight interpretations, seven of which may be relevant here. See Appendix 3.)

In order to arrive at explicata for Q and χ I have found it necessary to discuss them in an interconnected manner; i.e. there do not appear to be enough desiderata for Q, considered by itself, to circumscribe its possible explicata to a satisfactory extent.

In the present section the ground will be cleared by discussing what Q and χ depend upon. It is convenient to think of this dependence in terms of notation, which seems to bring out the main points better than a purely verbal discussion. For example, the symbols Q and χ are abbreviations for $Q(E:F)$, and $\chi(E:F)$, and these expressions are themselves abbreviations for $Q(E:F \mid U.H)$ and $\chi(E:F \mid U.H)$. To take U and H for granted, and omit them from the notation, is parallel to linguistic usage. If we say that it is bad for eggs to throw them in the air, we take it for granted that there is a law of gravitation, and that there is a large gravitational body nearby.

Events later than F and earlier than E may be relevant to χ but not to Q. Accordingly I shall assume that $Q(E:F)$ depends only on $P(E \mid F)$, $P(E \mid \bar{F})$, $P(E)$, and $P(F)$. It is natural to define $Q(E:F \mid G)$ as the same function of these four probabilities, but made conditional on G.

Even $Q(E:F \mid U.H)$ is an incomplete notation. If the subjective element is to be removed from the expression 'F caused E', then it must be expanded to 'F, as against $\bar{F}_D$, caused E rather than E′', where the suffix, D, to $\bar{F}$ (the negation of F), represents a complete specification of the relative probabilities of the mutually exclusive events whose disjunction is $\bar{F}$. (D represents a probability distribution.) We could use a notation like

$$Q(E/E' : F/\bar{F}_D \mid U.H)$$

or

$$Q(E : F/F_D \mid U.H.(E \vee E')),$$

the degree of causation of E rather than E′ by F rather than $\bar{F}_D$.

The failure to recognise all the variables on which tendency to cause is based was for me one of the stumbling blocks in capturing the notion of probabilistic causality, if indeed I have fully succeeded even now.

It should be held in mind that $E \vee E'$ is regarded as taken for granted in the four probabilities on which Q is assumed to depend, when we are concerned with the causation of E rather than E′. When we take $E \vee E'$ for granted we may write $\bar{E}$ instead of E′.

5 *Assumptions and Deductions Leading to the Explicatum for Q*

In order to make my assumptions clear I shall list them in the form of axioms, A1, A2, . . . ; and the deductions from them will be called theorems T1, T2, . . ., for ease of reference. On a first quick reading the justifications and proofs should *quite definitely* be skipped, but I have not postponed them to a later section. (I did so in an earlier draft, but the cross-referencing made the paper more difficult to read.) The justifications of the most easily acceptable axioms, and the proofs of the easily proved theorems will be omitted.

Let $P(F) = x$, $P(E \mid F) = p$, $P(E \mid \bar{F}) = q$, $P(E) = r$, so that

$$r = xp + (1 - x)q, \quad x = (r - q)/(p - q).$$

Unless $p = q$ (in which case $r = p = q$), x is a function of p, q, and r. Therefore by an assumption of the previous section we have :

A1. *$Q(E : F \mid G)$ is a function of p, q, r, unless perhaps $p = q = r$. We call this function $Q(p, q, r)$ so that the symbol Q has two slightly different meanings. The symbol G will usually be taken for granted and omitted.*

A2. *Q is a real number or ∞ or $-\infty$; but it may be indeterminate for special values of p, q, and r, such as when two of them are equal or one of them is 0 or 1.* (It seems sensible to look for a scalar explicatum rather than a 'vector'.)

The next two axioms deal with monotonicity and continuity.

A3. (i) *Q increases with p if q and r are held constant.*
(ii) *Q decreases when q increases if p and r are held constant.*

A4. *Q is continuous except when it becomes infinite or indeterminate, if there are such points.*

A5. *If $P(F) \neq 1$, meaning, as usual, $P(F \mid U . H) \neq 1$, then Q has the same sign as $p - r$, and therefore also the same sign as $p - q$, and as $r - q$; and if these expressions vanish we may say that F has no tendency to cause E, and we put $Q = 0$.* (This axiom removes a gloss from A1.)

A6. *Any causal net joining F to E, as defined below in Section 11, has a causal strength, S, and a causal resistance, R. These are positive numbers, except that if $p = q = r$, or if p or q is 0 or 1, we may get zero or infinite resistance or strength.* (An important part of the definition of a causal net is that it consists only of events that actually occurred or will have occurred.)

A7. *There is a functional relationship between R and S, $S = f(R)$, $R = g(S)$, where f and g are absolute functions inverse to one another.*

A8. *f and g are continuous decreasing functions.*

A9. *$\chi(E : F)$ is the strength of the complete causal net joining F to E. More precisely, it is the limit, as the sizes of the events tend uniformly to zero, of the strengths of nets ; where each net of the sequence joins F to E, consists of a finite number of events, and omits no events* (temporal) *temporarily between F and E.* It is not claimed that this axiom is formulated with complete rigour, but it is used in only a weak form for the explication of Q (in the proof of T14). It is introduced at this early stage in order to supplement the outline in Section 2. (If we assume that the degree to which F caused E has an objective meaning, with a precise numerical value, we are committed to the idea that there is a complete world, uninfluenced from outside. Outside influence could be allowed for by assuming that the numerical values are not absolutely precise.)

A10. *The strength of the causal net consisting of F and E alone is equal to $Q(E : F)$ when this is positive, and is otherwise zero.* (We shall clearly get nowhere unless we assume some relationship betwen Q and S, and A10 is the simplest reasonable relationship that could be assumed.)

The strength and resistance of a net, $\mathfrak{n}$, joining F to E, are denoted by $S(E : F \mid \mathfrak{n})$ and $R(E : F \mid \mathfrak{n})$.

A11. *Let $\mathfrak{n}$ be a 'chain', $F = F_0 \to F_1 \to \ldots \to F_n = E$. Then $S(E : F \mid \mathfrak{n})$ is some function, $\phi(s_0, s_1, \ldots, s_{n-1})$, of the strengths, $s_0, s_1, \ldots, s_{n-1}$, of the links.* Here

$$s_i = S(F_{i+1} : F_i \mid U_i . H),$$

where U_i represents the essential physical circumstances just before F_i began. (Causal chains are formally defined in Section 8.)

T1. $\phi(s) = s$. (Proof from A10 and A11.)

A12. *ϕ is a symmetrical function, i.e. the arguments of the function can be permuted without changing its value.*

A13. *ϕ is a non-decreasing function of each of its arguments.* (A chain cannot be weakened by strengthening any of its links without changing the strength of any of the others.)

T2. *ϕ vanishes if the chain is cut, i.e. if any of the links is of zero strength.* We may alternatively say in this case that there was no causal chain.

A14. *If two consecutive links are replaced by a single link of equivalent strength, then the strength of the chain is unchanged. Formally,*

$$\phi(s_0, s_1, \ldots, s_{n-1}) = \phi(s_0, s_1, \ldots, s_{i-1}, \phi(s_i, s_{i+1}), s_{i+2}, \ldots, s_{n-1}).$$

A15. *A chain is not weakened by 'omitting' one of its links, i.e.*

$$\phi(s_0, s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_{n-1}) \geq \phi(s_0, s_1, \ldots, s_{n-1}).$$

T3. *A chain is no stronger than its weakest link.* (From A10 and A15.)

Definition. Let the maximum possible causal strength be σ. This is either a positive finite number or $+\infty$.

A16. $S(E:F \mid \mathfrak{N}) \leqslant \sigma$, *for any net*, $\mathfrak{N}$. (This axiom is a mere restatement of the definition.)

T4. $Q(p, q, r) \leqslant \sigma$. (From A10 and A16.)

A17. *If any of the links of a chain is of strength σ, then it can be 'omitted' in the sense of A15, without strengthening the chain.*

T5. $$\phi(s_0, s_1, \ldots, s_{i-1}\, \sigma, s_{i+1}, \ldots, s_{n-1}) = \phi(s_0, s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_{n-1}).$$

(From A15 and A17.)

T6. *If every link of a chain is 'cast-iron', then the chain is cast-iron, i.e.* $\phi(\sigma, \sigma, \ldots, \sigma) = \sigma$. (From T1 and T5.)

A18. *A chain of n links all of the same fixed strength, τ, where $\tau < \sigma$, is as weak as you like if n is large enough. Formally $\phi(\tau, \tau, \ldots, \tau) \to 0$ as the number of arguments tends to infinity.*

A19. *ϕ is a continuous function of all its arguments when they are all less than σ; and, if $s_i \to \sigma$, then the value of the function tends to the value it would have with $s_i = \sigma$.* The reason for the clumsy expression of this axiom is that σ may be $+\infty$.

T7. *If a chain has n links, all of the same strength, τ, where $\tau < \sigma$, then the chain is as strong as you like if n is fixed and τ is close enough to σ. Formally, if n is fixed, then*

$$\phi(\tau, \tau, \ldots, \tau) \to \sigma \text{ when } \tau \to \sigma.$$

(From T6 and A19.)

T8. *There is a function, g, such that, identically,*

$$\phi(s_0, s_1, \ldots, s_{n-1}) = g^{-1}(g(s_0) + g(s_1) + \ldots + g(s_{n-1})).$$

The function, g, is defined for all non-negative arguments not exceeding σ, and is itself non-negative, continuous and strictly decreasing, and $g(0) = \infty$, $g(\sigma) = 0$. We may define g as $+\infty$ when its argument is negative.

Proof. Consider the function $\phi(s, t)$ of just two variables. By A19, A13, A12, and A14, this function may be said to be continuous, monotonic, commutative, and associative. It follows that it is of the form $g^{-1}(g(s) + g(t))$, where g is a strictly monotonic continuous function. The use of the symbol g is justified since A7 and A8 can be satisfied with this function.

The mathematical theorem just invoked was apparently first published by Abel.[1] It was rediscovered several times, such as by Aczél.[2] What it amounts to is that ϕ can be calculated on a suitably calibrated slide-rule.

The results $g(0) = 0$, $g(\sigma) = \infty$, follow from A18 and A19.

Q. E. D.

We may satisfy A6, A7, and A8, which are the only axioms that involve R, by identifying $g(S)$ with R. This identification is no restriction on the explication of Q. As a consequence of this identification we have the following theorem.

T9. *The resistance of a chain is equal to the sum of the resistances of its links.*

A20. *Consider the causal net shown in the diagram below, in which* $P(F) = x$, $P(G_j \mid F) = p_j$, $P(G_j \mid \bar{F}) = q_j$, $P(G_j) = r_j = xp_j + (1 - x)q_j$; $P(E \mid G_1 \vee G_2 \vee G_3) = 1$, $P(E \mid \bar{G}_1 . \bar{G}_2 . \bar{G}_3) = 0$, $P(E) = r$, $j = 1, 2, 3$, *and where* G_1, G_2, G_3 *are independent given* F *and also given* $\bar{F}$. *Then the strength of the net is a function of the strengths of the three separate chains, and this function is continuous, monotonic increasing in each variable, commutative* (cf. A12), *and associative* (cf. A14).

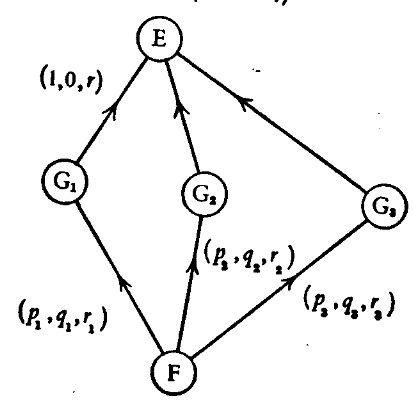

FIG. 1

T10. *The strength of the net of A20, generalised to m chains in parallel, is of the form*

$$h^{-1}(h(s_1) + h(s_2) + \ldots + h(s_m)),$$

[1] Neils Henrick Abel, *Oeuvres Complètes*, tome 1, 1881. The paper was Abel's first publication.

[2] J. Aczél, *Bull. Soc. Math. France*, 1948, **76**, 59–64. See also Aczél, *Acta Phys. Acad. Sci. Hungar.*, 1955, 4, 351–362; L. Jànossy, *loc. cit.* pp. 333–349. Or see *Math. Rev.*, **10**, 685, and **16**, 1127–1128.

where $s_1, s_2, \ldots, s_m$ are the strengths of the individual chains. The function h is defined for all non-negative arguments not exceeding σ, and is itself non-negative, continuous and strictly increasing, and $h(0) = 0$, $h(\sigma) = \infty$. (The theorem of the generalised slide-rule implies this theorem, just as it implied T8 above.)

We are now at liberty to call $h(S)$ the strength of a causal net, in place of S, provided we are content to determine the explicatum of S and Q only up to a continuous increasing transformation. It might be thought for a moment that this change of notation would invalidate T9. But since T8 is now true with $g(x)$ replaced by $g(h^{-1}(x))$, we can simply rename this function '$g(x)$' in order to validate T9. With these conventions we have:

T11. *The strength of the net of A20, generalised to m chains in parallel is the sum of the strengths of the individual chains.* When applying this theorem the independence condition mentioned in A20 should not be overlooked.

It appears that the analogy with electric networks is not bad, although the function $f(x)$ turns out later not to be $1/x$, but another self-inverse function.

A21. *In the net of* A20, *with G_3 omitted, i.e. with only two chains in parallel, we may regard $G = G_1 . G_2$ as a single event, without changing the causal strength of the net.* Note that it would be unreasonable to assume this coalescence property for dependent events, for if we did so we could collapse any net into a single event.

It may be objected that $G_1 . G_2$ is not necessarily a small event. But the strength of a causal net should depend only on its topology, with time-order preserved, and on the various probabilities and conditional probabilities. Hence $G_1 . G_2$ could be a small event in an equivalent network.

Although I think A21 is eminently reasonable, especially in view of later developments, as in Section 9, I believe it to be the weakest part of my argument, and I conjecture that the replacement of this axiom by other assumptions would be the most fruitful method of finding other explicata of tendency to cause, if they exist.

T12. *Identically, if $p_1 \geqslant q_1$, $p_2 \geqslant q_2$, $0 < x < 1$, then*

$$S(p_1 + p_2 - p_1p_2,\, q_1 + q_2 - q_1q_2,\, x(p_1 + p_2 - p_1p_2) + (1 - x)(q_1 + q_2 - q_1q_2))$$
$$= S(p_1, q_1, xp_1 + (1 - x)q_1) + S(p_2, q_2, xp_2 + (1 - x)q_2).$$

This follows at once from T11 by making the identification mentioned in A21.

A22. *$Q(p, q, r)$ is an analytic function when $0 < p < 1$, $0 < q < 1$, $0 < r < 1$, $p \neq q$.*

The only purpose of this axiom is to enable us to extend a formula proved for a large set of values of (p, q, r) to all values except those for which Q may be infinite or indeterminate. I think only an extreme purist would object to A22. It could be avoided by assuming instead that Q is anti-symmetric in the sense

$$Q(p, q, xp + (1 - x)q) = -Q(q, p, xq + (1 - x)p).$$

T13.

$$Q(p, q, r) = u(x) \log \frac{1 - q}{1 - p} = u\left\{\frac{r - q}{p - q}\right\} \log \frac{1 - q}{1 - p},$$

where $u(x)$ is a non-negative analytic function of x.

This theorem, and the next one, will be superseded by T15.

Proof. By A10 and A22, we may replace S by Q in T12, and drop the inequalities $p_1 \geqslant q_1$, $p_2 \geqslant q_2$. Let

$$\psi(\xi, \eta, x) = Q(1 - e^{\xi}, 1 - e^{\eta}, x(1 - e^{\xi}) + (1 - x)(1 - e^{\eta})),$$
$$p_1 = \exp \xi_1,\ q_1 = \exp \eta_1, \text{ etc.}$$

Then

$$\psi(\xi_1 + \xi_2, \eta_1 + \eta_2, x) = \psi(\xi_1, \eta_1, x) + \psi(\xi_2, \eta_2, x).$$

On putting $\eta_1 = \eta_2 = 0$, and provisionally regarding x as a constant, we get a well known functional equation whose only continuous solution is easily seen to be of the form

$$\psi(\xi, 0, x) = \xi . u(x),$$

where $u(x)$ is a function of x only. (The only other solutions are in fact non-measurable.[1]) Likewise $\psi(0, \eta, x) = \eta . w(x)$, where $w(x)$ is a function of x only. Therefore

$$\psi(\xi, \eta, x) = \psi(\xi + 0, 0 + \eta, x)$$
$$= \psi(\xi, 0, x) + \psi(0, \eta, x) = \xi u(x) + \eta w(x).$$

Therefore

$$Q(p, q, xp + (1 - x)q) = u(x) . \log (1 - p) + w(x) . \log (1 - q).$$

T13 now follows from A5 combined with the equation

$$r = xp + (1 - x)q.$$

Q. E. D.

A23. *Consider a radioactive particle in a certain state, which I shall call the 'white' state. In any time interval, t, it has probability $e^{-\alpha t}$ of remaining in the white state throughout the interval if it starts the interval in that state.*

[1] G. Hamel, 'Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichungen $f(x + y) = f(x) + f(y)$', *Math. Annalen*, 1905, 60, 459-462. Or see G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge 1934, p. 96.

If it does not remain in the white state, then it proceeds to another state called here the 'black' state, from which there is no return. Now let F be the event that the particle is in the white state at the start of an interval of duration T and let E be the event that it is in the white state at the end of this interval. Then we assume that, if F and E both occurred, $\chi(E:F)$ *does not depend on the unit in terms of which time is measured.*

A24. *If* F . E *implies* G, *and* $F \to G \to E$ *is a chain, then this chain is of the same strength as* $F \to E$.

T14. $R(p, 0, r) = v(r/p) - k \cdot \log p,$

where $v(x)$ *is a non-negative analytic function of* x, *and* k *is a positive constant.*

Proof. Consider the radioactive particle described in A23. Let $P(F) = x$. The degree to which F caused E is the limit of the strengths of finite chains obtained by breaking up the time interval (0,T) into a 'Riemann dissection' (see A9). Since g is a continuous function (A8) the resistances of these finite chains must also tend to a limit, which we may call the causal resistance from F to E. This must be some function of x, α, and T, say $R^*(x, \alpha, T)$. By A23 we see that for any positive constant, k, the resistance must be equal to $R^*(x, k\alpha, T/k)$. Since this is independent of k it must be of the form of $R^*(x, \alpha T)$.

Now, by a continuity argument, we may generalise T9 to continuous chains, and hence deduce that, for any positive T and U we have

$$R^*(x, \alpha T) + R^*(1, \alpha U) = R^*(x, \alpha T + \alpha U).$$

By giving x the value 1 and subtracting from the equation with arbitrary x, we see that $R^*(x, \alpha T)$ is of the form

$$R^*(x, \alpha T) = v(x) + R^*(\alpha T),$$

where, identically,

$$R^*(\alpha T_1 + \alpha T_2) = R^*(\alpha T_1) + R^*(\alpha T_2),$$

so that $R^*(\alpha T)$ is of the form

$$R^*(\alpha T) = k_1 \alpha T.$$

Now, by repeated use of A24, we see that

$$R(p, 0, xp) = R^*(x, \alpha T),$$

where $p = e^{-\alpha T}$. Thus

$$R(p, 0, r) = v(r/p) - k \cdot \log p.$$

Q. E. D.

T15. $Q(p, q, r) = \log(1 - q) - \log(1 - p),$

$R(p, 0, r) = -\log p,$

where the base of the logarithms may be taken as e. $Q(p, q, r)$ *is mathematically independent of* r, *and may be abbreviated to* $Q(p, q)$. *It can be written in other ways:*

$$Q(E:F \mid G) = \log \frac{P(\bar{E} \mid \bar{F} . G)}{P(\bar{E} \mid F . G)} = \log \frac{O(\bar{F} \mid \bar{E} . G)}{O(\bar{F} \mid G)}$$

$$= W(\bar{F}:\bar{E} \mid G) = -W(F:\bar{E} \mid G),$$

the weight of evidence against F *if* E *does not happen. More precisely,* Q *is uniquely determined only up to a continuous analytic increasing transformation. Among all the explicata there is just one apart from a scale factor (choice of unit), for which theorems* T9 *and* T11 *are true.* We lose no real generality, and we gain simplicity, by choosing this explicatum.

Proof. By T13, T14, and A7, we have the identity

$$f(v(r/p) - \log p) = -u(r/p) \cdot \log(1 - p).$$

Let $v(x) = y$, $-\log p = z$, $\log f(y + z) = \rho(y + z)$. Then $\rho(y + z)$ is of the form

$$\rho(y + z) = \rho_1(y) + \rho_2(z).$$

If $v(x)$ is not a constant, we can differentiate and deduce that $\rho(y)$ is a linear function of y, from which we can soon derive that $\log(1 - p)$ is a power of p. Since this is false it follows that $v(x)$ is a constant, and hence also that $u(x)$ is a constant.

The theorem now follows from the remark that the choice of the base of the logarithms is equivalent merely to the choice of units of measurement of strength and resistance. We may call the units 'natural', 'binary', or 'decimal', according as the base is e, 2, or 10. In this paper I shall use natural units. Possible names would be 'natural causats' and 'natural tasuacs'.

Note that the explicatum for Q was by no means obvious in advance, nor was it obvious that all the desiderata could simultaneously be satisfied.

It is interesting to note that, if, contrary to most of the discussion, we assume E to be earlier than F, and if the universe has the 'Markov' property, defined below, then the tendency of F to cause E is zero. This result may very well have been taken as a desideratum, but was in fact noticed only after the explcatum was obtained. By the Markov property is meant here that, for prediction, a complete knowledge of the immediate past makes the remote past irrelevant.

T16. *The relationship between* R *and* S *is symmetrical, namely*

$$R \geqslant 0, \; S \geqslant 0,$$
$$e^{-R} + e^{-S} = 1,$$

or equivalently,

$$R = -\log(1 - e^{-S}), \; S = -\log(1 - e^{-R}).$$

Further,

$$R(p, q, r) = \log(1 - q) - \log(p - q).$$

This is an immediate corollary of A7 and T15.

Thus the function f is its own inverse, g. It is tempting to permit negative and imaginary values because some of the formalism is faintly reminiscent of Feynmann's formulation of quantum mechanics, but I shall not pursue this matter here.

T17. *If a chain consists of n links whose p's and q's are* (p_i, q_i), *where* $p_i \geqslant q_i$, *then its causal strength is*

$$-\log\left\{1 - \prod_i \frac{p_i - q_i}{1 - q_i}\right\}.$$

This follows from T16 and T9.

Before reading the proofs in the present section the reader will probably prefer to read the next two sections, in which some examples are given.

Appendix. Holmes, Moriarty, and Watson (see section 2)

Sherlock Holmes is at the foot of a cliff. At the top of the cliff, directly overhead, are Dr Watson, Professor Moriarty, and a loose boulder. Watson, knowing Moriarty's intentions, realises that the best chance of saving Holmes's life is to push the boulder over the edge of the cliff, doing his best to give it enough horizontal momentum to miss Holmes. If he does not push the boulder, Moriarty will do so in such a way that it will be nearly certain to kill Holmes. Watson then makes the decision (event F) to push the boulder, but his skill fails him and the boulder falls on Holmes and kills him (event E).

This example shows that Q(E: F) and χ(E: F) cannot be identified, since F had a tendency to prevent E and yet caused it. We say that F was a cause of E because there was a chain of events connecting F to E, each of which was strongly caused by the preceding one.

(to be concluded)

Admiralty Research Laboratory
Teddington, Middlesex

COMMUNICATIONAL EPISTEMOLOGY (I) *

MAGOROH MARUYAMA

IN the history of philosophy the philosopher for long asked himself: 'What is the relationship between THE universe and THE human understanding about it?' This question presupposed the existence of THE universe and of THE human understanding.

THE universe was conceived of in three ways:

(1) As THE material reality outside the human organism, existing independently of the human mind, and perceived through sense organs.

(2) As THE ideal reality, existing independently of the human mind, and of which the human mind can attain understanding, at least in part, by refined thinking.

(3) As a construction of the human mind which is the same for all human beings, though it might be different for other beings.

Those who conceived of the universe in the first way, as THE material reality, can be subdivided into:

(1*a*) Those who interpreted the natural law as independent of the human mind.

(1*b*) Those who interpreted it as a construction of the human mind.

(1*c*) Those who identified the human mind with God, with the natural law, with nature, or, as in mysticism, the human mind with God, natural law and nature.

THE universe of (1*a*), (1*c*) and (2) did not depend on the existence of THE human mind. But the THE-ness of the universe in (1*b*) and (3) depended on the THE-ness of the human understanding. Belief in the possibility and success of interpersonal communication was also dependent on belief in the uniqueness and sameness of the human understanding, except for the subjective qualities of sensations and emotions. This belief was strengthened by the fact that philosophy dealt with conscious and verbally communicable realities[1] and by

* Received 3.xi.59. The writer is indebted to a referee for numerous stylistic improvements.

[1] M. Maruyama, 'Communicable and Incommunicable Realities', *The British Journal for the Philosophy of Science*, 1959, 10, 50-54

sense of emergence, or the appearance of genuine novelty at a higher level of richness or complexity. But this is also a kind of improvement which neither statistical genetics nor selectionist biology can handle, since it is neither quantitative nor adaptive. It is best for selectionists to ignore it, as Darwin warned himself to do when he wrote in his copy of the *Vestiges* 'Never speak of higher or lower in evolution'. Yet the great outlines of the fossil record are there, and demand to be spoken of, especially since the fact that we can speak of them is one of the surprising results of the process they record. But evolution as macro-evolution, as the emergence of life and of higher forms of life, outruns both the concept of gene-substitution, and of improvement in relation to environment. It makes sense only as an achievement—an achievement for which statistical methods can measure the necessary, but not the sufficient conditions.

15

One brief concluding remark. I have side-stepped here altogether the question of prediction and retrospect, of the historical nature of evolutionary explanation: a question which is very close to the philosophical difficulties raised by Fisher's theory. Evolutionary theory is essentially an assessment of the past. Fisher treats it in terms of present and future. Just how closely the philosophical confusions of this kind of argument are related to the attempt to think unhistorically about an historical subject matter, I should not at the moment venture to say.

Department of Philosophy
The Queen's University, Belfast

A CAUSAL CALCULUS (II) *

I. J. Good

6 *Two-state Markov Processes*

The radioactive process described in Axiom 23 can be slightly generalised by permitting return from the black to the white state, with a parameter β corresponding to the α of the white-to-black transition. We have a two-state Markov process with continuous time. The parameters α and β are of course both non-negative. In the special case of the radioactive particle we have $\beta = 0$.

It can be shown that

$$Q(E : F) = \log\left((\alpha + \beta e^{-(\alpha+\beta)T})/(\alpha - \alpha e^{-(\alpha+\beta)T})\right).$$

If the particle ever entered the black state during the time interval, T, the chain would be cut and the degree of causality would be zero. Assuming that this does not happen, we can calculate $\chi(E : F)$ by applying a Riemann dissection to the interval, so as to obtain a causal chain consisting of a finite number of events, and then proceed to the limit as the fineness of the dissection tends to zero. By applying T17 and A9 we find that

$$\chi(E : F) = -\log(1 - e^{-\alpha T}),$$

which is mathematically independent of β.

For large T, both Q and χ are exponentially small, but Q is smaller than χ, and is much smaller if β is large. This is reasonable since, if β is large, the initial state makes little difference to the probability of being in the white state at the end of the interval.

Note that χ is the degree to which being in the white state rather than in the black state at the end of the interval was caused by being in the white state rather than in the black state at the start of the interval. A similar explicit description can of course be given for Q.

7 *Partially Spurious Correlation*

A well known pitfall in statistics is to imagine that a statistically significant correlation or association is necessarily indicative of a causal relationship. The seeing of lightning is not a cause of the hearing of

* The first part of this article appeared in the previous Number.

thunder, though the two are strongly associated. Such associations and correlations are often described as 'spurious', a better description than 'illusory'. They may also be *partially spurious*, and the explicata for Q and χ should help with the analysis of such things. Smoke and dust might be a strong cause of lung cancer, but smoking only a weak cause. Even so, the correlation between smoking and lung cancer may be high if there is more smoking per head in smoky districts. I mention this only as an example, and have not made a special study of this problem.

Note that

$$Q(E : F . G/\overline{F} . \overline{G}) = Q(E : G|\overline{F}) + Q(E : F|G),$$

so that the tendency to cause can be split into components, somewhat in the manner of an analysis of variance. For example, the tendency for lung cancer to be caused by smoking and living in a smoky district as against not smoking and living in a clean district is equal to the tendency through living in a smoky district, given no smoking, plus the tendency through smoking, given that the district is smoky. It is also equal to the causal tendency through living in a smoky district, given that one smokes, plus the tendency through smoking, given that the district is clean. This approach to the analysis of spurious correlation is entirely different from, and more quantitative, than the approach used by Simon.[1]

Let

$$K(E : F) = - I(\overline{E} : F),$$

the *intrinsic* causal tendency of E by F. It is related to Q in essentially the same way that I is related to W, since

$$Q(E : F) = K(E : F) - K(E : \overline{F}),$$
$$Q(E : F/F') = K(E : F) - K(E : F').$$

K does not depend on the negation of F, so its use enables us to avoid the distribution, D, of Section 4. We have

$$K(E : F. G) = K(E : F) + K(E : G \mid F),$$

so that K can be split up into contributions from various sources in a simpler manner than Q. In my opinion both K and Q will probably have useful applications in statistics and physics.

The remainder of this paper is primarily concerned with the extension of the explication of causal strength to general nets, in order that degree of causality should be generally explicated. The next section however contains a formal definition of a causal chain, which strictly

[1] Herbert A. Simon, *Models of Man*, New York and London 1957

was required in what has already been discussed. I postponed it in order not to interrupt the thread of the argument.

8 *Causal Chains*[1]

Let $F = F_0, F_1, \ldots, F_{n-1}, F_n = E$, be $n + 1$ events such that (for $i = 0, 1, \ldots, n - 1$):

(*i*) F_i and F_{i+1} are contiguous in space and time, or approximately so.

(*ii*) No two of the events overlap much in space *and* time.

(*iii*) All the events occurred (or will have occurred, i.e. they 'obtain' but I prefer to write simply 'occurred').

(*iv*) F_{i+1} started later than F_i did.

(*v*) F_i had a positive tendency to cause F_{i+1}.

(*vi*) If F_i is given, then the probability of F_{i+1} is unchanged if one or more of the earlier events did not occur, i.e. we have a Markov chain.

(*vii*) If the chain is embedded in a completely detailed chain containing intermediate events, then condition (*v*) will remain true for the more detailed chain.

Then we say that $F_0, F_1, \ldots, F_n$ or $F_0 \to F_1 \to \ldots \to F_n$ is a causal chain connecting F to E. Perhaps it should be called a 'putative causal chain' if condition (*vii*) has not been established. In practice all causal chains are putative, but there are degrees of putativity.

The failure of condition (*v*) may be said to 'cut the chain'.

A causal net will be formally defined in Section 11. A chain is a special case of a net.

9 *Independent Causal Tendencies*

Let $G_1, G_2, \ldots, G_m$ be independent given H, and also independent given $H . \overline{E}$. Then it is easily proved, with the help of T15, that the tendencies to cause E are additive in the sense of the theorem below. It therefore seems reasonable to say in these circumstances that the G's have *independent tendencies to cause* E *given* H. The events G_1, G_2, and G_3 of A20 exemplify this definition, with $H = F$, and also with $H = \overline{F}$; that they are independent given $\overline{E}$ is trivial since their probabilities are then all zero.

[1] Cf. Hans Reichenbach, *The Direction of Time*, Berkeley and Los Angeles 1956, Index reference under 'Causal chain', and *The Philosophy of Space and Time*, New York and London 1956, Index reference under 'Casual chain' [*sic*]

T18. *If* $G_1, \ldots, G_m$ *have independent tendencies to cause* E *given* H, *then*

$$Q(E : G_1 . G_2 . \ldots . G_m \mid H) = \sum_i Q(E : G_i \mid H).$$

The nets of A20 and T10 also exemplify the following definition :

A *bundle of parallel independent causal chains* from F to E is a class of chains from F to E such that, apart from F and E, each event on each chain is, given F and given $\bar{F}$, probabilistically independent of any collection of events on other chains, and also such that the penultimate events have independent tendencies to cause E, given their pasts.

10 *Series-parallel Networks*

As an extension of T11 it is natural to define the strength of a bundle of independent causal chains as the sum of the strengths of the individual chains.

For a 'chain of bundles', in a self-explanatory sense, we can first calculate the resistance by summing the resistances of the individual bundles, and then obtain the strength from T16. We can extend the process to bundles of chains of bundles and so on. In other words we can construct natural rules for evaluating the causal strength of any 'series-parallel' net. Topologically these are the same as the two-terminal series-parallel networks whose enumeration was considered by MacMahon.[1] Not all networks are of this type.

11 *Causal nets 'Having Independence'*

Let $\mathfrak{n}$ be a class of events all of which occurred. For each event, G, in $\mathfrak{n}$, there is a subclass of earlier events, $G_1, G_2, \ldots, G_k$, which so to speak, 'lead in' to G. By 'lead in' is meant that the probability of G, given which of $G_1, G_2, \ldots, G_k$ occurred and which did not, is independent of any further assumptions of which other events in $\mathfrak{n}$, earlier than G, occurred. (Note that not all the events in $\mathfrak{n}$ are regarded as 'given' even though they all actually occurred. This should cause neither surprise nor confusion to those who are familiar with the idea of a conditional probability.) We may think of k oriented links joining $G_1, G_2, \ldots, G_k$ to G. If the whole class, $\mathfrak{n}$, is connected together by means of such links we describe $\mathfrak{n}$ as a *causal net*. If E is the latest of the events in the net, and can be reached from each other event by passing through a succession of links in the right direction, then the causal net will be said to *lead to* E. If F is the earliest of the events in $\mathfrak{n}$, and each other event can be reached from F by passing through a succession of links in the right direction, then the causal net will be said to *lead from* F. If both conditions are satisfied, the net will be said to *lead from* F *to* E. For example, a net leading to E could have the form of a 'tree', but a net leading from F to E could be a tree only if it were a chain.

[1] P. A. MacMahon, 'The combination of resistances', *The Electrician*, 1892, **28**, 601-602. Or see John Riordan, *An Introduction to Combinatorial Analysis*, New York and London 1958, pp. 139-143.

In this definition we may call $G_1, G_2, \ldots, G_k$ the *immediate predecessors* of G. A causal net will be said to *have independence* if, for each G in the net, the immediate predecessors have independent tendencies to cause G given the past.

For each link $G_i \rightarrow G$, having a 'p' and a 'q',

$$p = p_i = P(G \mid G_i), \; q = q_i = P(G \mid \bar{G}_i),$$

let the *quasiprobability*, π, be defined as

$$\pi = \left[\frac{p - q}{1 - q}\right],$$

in which the 'square' brackets indicate that $\pi = 0$ if $q \geqslant p$. The quasiprobability reduces to p when $q = 0$. We know from T17 that the quasiprobabilities are multiplicative for a chain, and the strength of the chain is the same as if the quasiprobabilities were ordinary probabilities and the q's were all zero. Also, from T15, we have

$$S(p, q) = -\log(1 - \pi),$$

so that for a bundle of the type occurring in T10 the quasiprobabilities again behave like probabilities, in view of the additivity of the strengths of the chains.[1]

Let us now consider an arbitrary finite causal net having independence and leading from F to E. We should like a general procedure for defining the strength of such a net that will include the results for the nets already considered, and which is simple, and which does not lead to a contradiction. I believe that the procedure illustrated in the following example satisfies these conditions. It would of course be more satisfactory if some convincing axioms could be laid down that would uniquely determine the procedure.

[1] The term 'pseudoprobabilities' would conveniently refer (by analogy with the pseudo-random numbers that are often used in Monte Carlo calculations) to the apparent probabilities that occur in a deterministic, but pseudo-indeterministic, set-up.

In the diagram, the quasiprobabilities $\pi_1, \pi_2, \ldots, \pi_8$ are assigned, and pertain to the links of the net. It will be easier to appreciate the example if the π's are at first thought of as ordinary probabilities (with all the q's equal to 0).

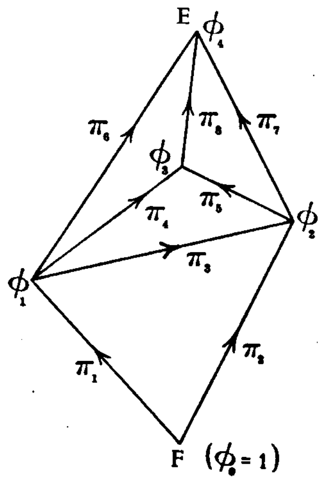

FIG. 2

The ϕ's may be thought of as quasiprobabilities of the *events*. They are defined successively as follows:

$$\begin{aligned} \phi_0 &= 1. \quad \phi_1 = \pi_1. \\ \phi_2 &= 1 - (1 - \pi_2)(1 - \phi_1\pi_3). \\ \phi_3 &= 1 - (1 - \phi_1\pi_4)(1 - \phi_2\pi_5). \\ \phi_4 &= 1 - (1 - \phi_1\pi_6)(1 - \phi_2\pi_7)(1 - \phi_3\pi_8). \\ S(\Pi) &= Q(\phi_4, 0) = -\log(1 - \phi_4). \end{aligned}$$

The reader should perhaps check that this procedure contains the previous ones as special cases.

12 *Causal Nets in General*

It will often be possible to divide up a time-slice preceding E into non-overlapping events whose causal influences on E are approximately but not absolutely independent. Let such a dissection of the time-slice be $F_1, F_2, \ldots, F_m$. We need a definition of the strength of the causal link $F_1 \to E$ that will reduce to the value given previously in the case where F_1 and $F_2 . F_3. \ldots F_m$ are *causally independent* with respect to E, in the same sense as that defined above for nets having

independence. A simple definition having the required property is

$$S(E:F_1) = \log \frac{P(\bar{E} \mid \bar{F}_1 . F_2. \ldots F_m)}{P(\bar{E} \mid F_1 . F_2. \ldots F_m)} = W(\bar{F}_1 : \bar{E} \mid F_2 . F_3. \ldots F_m).$$

This definition reduces to the previous use of the expression $S(E:F)$ in the case of causal independence. But the strengths of the lead-ins do not add up to $S(E:F_1. \ldots F_m)$ unless the F's do have independent causal influences on E. We can cope with this difficulty by the introduction of 'interaction terms' in a sense analogous to the use of this expression in the literature of the design of statistical experiments.[1]

We can think of an extra node in the causal net leading to E corresponding to every subset of the events $F_1, F_2, \ldots, F_m$. For example, there will be a node corresponding to the pair $(F_1 . F_2)$. The strength of the link from the node $(F_1 . F_2)$ to E will then be taken as the 'interaction' term

$$t_{12} = s_{12} - s_1 - s_2,$$

where

$$\begin{aligned} s_{12} &= \log \frac{P(\bar{E} \mid \overline{F_1 . F_2}. \ldots F_m)}{P(\bar{E} \mid F_1 . F_2. \ldots F_m)} \\ &= W(\overline{F_1 . F_2} : \bar{E} \mid F_3 \ldots F_m). \end{aligned}$$

When F_1 and F_2 are independent causes of E, we have $s_{12} = s_1 + s_2$, and the second order interaction term vanishes. The strength of the link *to* $(F_1 . F_2)$, from an earlier event, G, is

$$W(\bar{G} : \overline{F_1 . F_2} \mid G_1 . G_2. \ldots),$$

where $G_1, G_2, \ldots$ are the other immediate predecessors of F_1 and F_2. The definitions of the s's are forced, if we regard conjunctions of the F's as single events. An example of a third-order interaction is

$$t_{1234} = s_{1234} - s_{234} - s_{134} - s_{124} - s_{123} + s_{12} + s_{13} + \ldots + s_{34} - s_1 - s_2 - s_3 - s_4,$$

where the notation is now self-explanatory. In any piece of causal analysis one would try to choose the dissection of the time-slice so as to make the high-order interactions negligible.

Since

$$s_{123 \ldots m} = \sum s_i + \sum t_{ij} + \sum t_{ijk} + \ldots,$$

our enlarged causal net has the property of additivity of strengths of

[1] See, for example, *Design and Analysis of Industrial Experiments*, ed. by O. L. Davies, London and Edinburgh 1954; index reference under 'Interaction'.

lead-ins that we previously had for causally independent lead-ins. It is therefore now potentially possible to apply the method of Section 11 to define the causal strength of an arbitrary finite net from F to E.

13 *Degrees of Causation*

We may now define $\chi(E : F)$ as the limit of the strength of the net joining F to E and containing all intermediate events, when the events are made smaller and smaller. I have not proved that this limit exists. The proof, if possible, would depend on a physical theory, and would be mathematically intricate. Note the implication: whether degrees of causality exist is a matter of physics, even if we take for granted that physical probabilities exist.

In practice one must always over-simplify or simplify in order to be able to judge, estimate, or guess, the value of $\chi(E : F)$. (In the past, χ has been given only a few values, such as 'small', 'moderate', and 'large'.) There is always the possibility that something has been overlooked. Even in a statistical experiment involving randomisation, from which we can apparently deduce that some $\chi(E : F)$ is large, in fact E and F may both have been caused by some preceding event. The table of random numbers might have been seen by the famous lady tea-taster,[1] or there may have been some psychokinesis. We are always thrown back on judgment.

14 *Big Events*

So far the analysis has assumed F and E to be small events. If F is big we may imagine it split up into many small events, and imagine all these to be 'short-circuited' from an earlier 'input node'. By 'short-circuited' is meant that the resistances of all the imaginary links are taken to be zero. We may apply a similar process to a big E by short-circuiting its small parts to a future output node. The previous methods may then be applied even if F does not end before E begins.

Appendix I. Correction of some errors in previous work

Reichenbach[2] says that F is causally relevant to E if $P(E \mid F) > P(E)$ and if there is no set of events earlier or simultaneous with F that 'screens off' E from F. By 'screens off' he means that the probability of E given these

[1] Sir Ronald A. Fisher, *The Design of Experiments*, 5th edn., Edinburgh and London 1949, Chapter 2

[2] Page 204 of the first reference in footnote 1, p. 45 above

other events is unchanged if F is also given. The property is analogous to the Markov property.

It seems to me that this definition is not acceptable as it stands for much the same reason that my previous paper is not acceptable. For let G be any set of events earlier than or simultaneous with F. G might be some exceedingly biased selection of individual molecules, such as those that are proceeding south at a thousand miles per hour. Consider the expression $P(E \mid G) - P(E \mid G \,.\, F)$. Normally this will be positive for some G, say G_1, and negative for some G, say G_2. We now imagine G_1 to be gradually distorted into G_2. The above expression must change sign at some point during this gradual distortion, at which 'time' its value will be zero. Hence the second part of Reichenbach's definition seems to be vacuous. In order to patch up the definition it seems to be necessary to take G as the complete state of the universe at the time F started.

In my previous paper, conditions C7 to C10 were vacuous for much the same reason, though it may be possible to patch the thing up, as stated therin (inserted in proof), by insisting that G should be in some sense a 'natural' event.

Appendix II. The meaning of 'state' in quantum mechanics (see Section 4)

The seven relevant interpretations of 'state' in quantum mechanics are the first seven on the following list. All seven of these meanings, and perhaps others, should be taken into account in a comprehensive discussion of the place of probabilistic causality in quantum mechanics.

(*i*) The class of all past phenomena, classically describable. (*ii*) The class of phenomena extending only a short way into the past. (*iii*) The wave function of a physical system, under observation by another physical system. (*iv*) The joint wave function of the pair of systems. (*v*) The wave function of one system conditional on an assumed wave function of another system. This is the 'relative state' of Hugh Everett III, 'Relative state formulation of quantum mechanics', *Rev. Modern Physics*, 1957, **29**, 454-462. (*vi*) The wave function of the entire universe if this has any meaning. See Everett, *loc. cit.* (*vii*) The wave function of the entire universe together with all other past phenomena. (*viii*) An ensemble of wave functions. See, for this eighth interpretation, R. C. Tolman, *The Principles of Statistical Mechanics*, Oxford 1938, Section 98.

(*concluded*)

Admiralty Research Laboratory
Teddington, Middlesex