PROBABILISTIC INFERENCE AND THE CONCEPT OF TOTAL EVIDENCE

by

Patrick Suppes

TECHNICAL REPORT NO. 94

March 23, 1966

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Probabilistic Inference and the Concept of Total Evidence[1]

Patrick Suppes

1. Introduction. My purpose is to examine a cluster of issues centering around the so-called statistical syllogism and the concept of total evidence. The kind of paradox that is alleged to arise from uninhibited use of the statistical syllogism is of the following sort.

(1) The probability that Jones will live at least fifteen years given that he is now between fifty and sixty years of age is r. Jones is now between fifty and sixty years of age. Therefore, the probability that Jones will live at least fifteen years is r.

On the other hand, we also have:

(2) The probability that Jones will live at least fifteen years given that he is now between fifty-five and sixty-five years of age is s. Jones is now between fifty-five and sixty-five years of age. Therefore, the probability that Jones will live to at least fifteen years is s.

---

[1]The writing of this paper has been supported by a grant from the Carnegie Corporation of New York. This paper will be published in Aspects of Inductive Logic, edited by K. J. J. Hintikka and the author, North-Holland Publishing Co., Amsterdam, The Netherlands.

The paradox arises from the additional reasonable assertion that $r \neq s$, or more particularly that $r > s$. The standard resolution of this paradox by Carnap (1950, p. 211), Barker (1957, pp. 76-77), Hempel (1965, p. 399) and others is to appeal to the concept of total evidence. The inferences in question are illegitimate because the total available evidence has not been used in making the inferences. Taking the premises of the two inferences together, we know more about Jones than either inference alleges, namely, that he is between fifty-five and sixty years of age. Parenthetically, I note that if Jones happens to be a personal acquaintance what else we know about him may be beyond imagining, and if we were asked to estimate the probability of his living at least fifteen years we might find it impossible to lay out the total evidence that we should use, according to Carnap et al., in making our estimation.

There are at least two good reasons for being suspicious of the appeal to the concept of total evidence. In the first place, we seem in ordinary practice continually to make practical estimates of probabilities, as in forecasting the weather, without explicitly listing the evidence on which the forecast is based. At a deeper often unconscious level the estimations of probabilities involved in most psychomotor tasks--from walking up a flight of stairs to catching a ball--do not seem to satisfy Carnap's injunction that any application of inductive logic must be based on the total evidence available. Or, at the other end of the scale, many actually used procedures for estimating parameters in stochastic processes do not use the total experimental evidence available, just because it is too unwieldy a task (see, e.g., the discussion on pseudo-maximum-likelihood estimates in Suppes and Atkinson (1960, ch. 2)). It might be argued that these differing sorts of practical examples have as a common feature just their deviation from the ideal of total evidence, but their robustness of range if nothing else suggests there is something wrong with the idealized applications of inductive logic with an explicit listing of the total evidence as envisioned by Carnap.

Secondly, the requirement of total evidence is totally missing in deductive logic. If it is taken seriously, it means that a wholly new principle of a very general sort must be introduced as we pass from deductive to inductive logic. In view of the lack of a sharp distinction between deductive and inductive reasoning in ordinary talk, the introduction of such a wholly new principle should be greeted with considerable suspicion.

I begin my critique of the role of the concept of total evidence with a discussion of probabilistic inference.

2. Probabilistic inference. As a point of departure, consider the following inference form.

(3)
$$\frac{\begin{array}{c} P(A|B) = r \\ P(B) = \rho \end{array}}{P(A) \geq r\rho}$$

In my own judgment (3) expresses the most natural and general rule of detachment in probabilistic inference. (As we shall see shortly, it is often useful to generalize (3) slightly and to express the premises also as inequalities.

$$\text{(3a)}\qquad \frac{\begin{array}{r} P(A|B) \geq r \\ P(B) \geq \rho \end{array}}{P(A) \geq r\rho}$$

The application of (3a) considered below is to take $r = \rho = 1 - \epsilon$.) It is easy to show two things about (3); first, that this rule of probabilistic inference is derivable from elementary probability theory (and Carnap's theory of confirmation as well, because a confirmation function $c(h,e)$ satisfies all the elementary properties of conditional probability), and secondly, no contradiction can be derived from two instances of (3) for distinct given events B and C, but they may, as in the case of deductive inference, be combined to yield a complex inference.

The derivation of (3) is simple. By the theorem on total probability, or by an elementary direct argument

$$\text{(4)}\qquad P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}),$$

whence because probabilities are always non-negative, we have at once from the premises that $P(A|B) = r$ and $P(B) = \rho$, $P(A) \geq r\rho$. Secondly, from the four premises

$$\begin{aligned} P(A|B) &= r \\ P(B) &= \rho \\ P(A|C) &= s \\ P(C) &= \sigma, \end{aligned}$$

we conclude at once that $P(A) \geq \max(r\rho, s\sigma)$, and no contradiction results. Moreover, by considering the special case of $P(B) = P(C) = 1$, we move close to (1) and (2) and may prove that $r = s$. First we obtain, again by an application of the theorem on total probability and observation of the fact that $P(\overline{B}) = 0$ if $P(B) = 1$, the following inference form as a special case of (3)

$$\text{(5)} \qquad \frac{\begin{array}{c} P(A|B) = r \\ P(B) = 1 \end{array}}{P(A) = r} \,.$$

The proof that $r = s$ when $P(B) = P(C) = 1$ is then obvious.

$$\text{(6)} \qquad \left\{ \begin{array}{lrl} (1) & P(A|B) = r & \text{Premise} \\ (2) & P(B) = 1 & \text{Premise} \\ (3) & P(A|C) = s & \text{Premise} \\ (4) & P(C) = 1 & \text{Premise} \\ (5) & P(A) = r & 1, 2 \\ (6) & P(A) = s & 3, 4 \\ (7) & r = s & 5, 6 \end{array} \right.$$

The proof that $r = s$ seems to fly in the face of statistical syllogisms (1) and (2) as differing predictions about Jones. This matter I want to leave aside for the moment and look more carefully at the rule of detachment (3), as well as the more general case of probabilistic inference.

For a given probability measure P the validity of (3) is unimpeachable. In view of the completely elementary--indeed, obvious--character of the argument establishing (3) as a rule of detachment, it is in many ways hard to understand why there has been so much controversy over whether a rule of detachment holds in inductive logic. Undoubtedly the source of the controversy lies in the acceptance or rejection of the probability measure P. Without explicit relative frequency data, objectivists with respect to the theory of probability may deny the existence of P, and in similar fashion confirmation theorists may also if the language for describing evidence is not explicitly characterized. On the other hand, for Bayesians like myself, the existence of the measure P is beyond doubt. The measure P is a measure of partial belief, and it is a condition of coherence or rationality on my simultaneously held beliefs that P satisfy the axioms of probability theory (forceful arguments that coherence implies satisfaction of the axioms of probability are to be found in the literature, starting at least with de Finetti (1937)). It is not my aim here to make a general defense of the Bayesian viewpoint, but rather to show how it leads to a sensible and natural approach to the concept of total evidence.

On the other hand, I emphasize that much of what I have to say can be accepted by those who are not full-fledged Bayesians. For example, what I have to say about probabilistic inference will be acceptable to anyone who is able to impose a common probability measure on the events or premises in question.

For the context of the present paper the most important thing to emphasize about the rule of detachment (3) is that its application in an argument requires no query as to whether or not the total evidence has been considered. In this respect it has exactly the same status as the rule of detachment in deductive logic. On the other hand it is natural from a logical standpoint to push for a still closer analogue to ordinary deductive logic by considering Boolean operations on events.

It is possible to assign probabilities to at least three kinds of entities: sentences, propositions and events. To avoid going back and forth between the sentence-approach of confirmation theory and the event-approach of standard probability theory, I shall use event-language but standard sentential connectives to form terms denoting complex events. For those who do not like the event-language, the events may be thought of as propositions or elements of an abstract Boolean algebra. In any case, I shall use the language of logical inference to talk about one event implying the other, and so forth.

First of all, we define $A \rightarrow B$ as $\bar{A} \vee B$ in terms of Boolean operations on the events A and B. And analogous to (3), we then have, as a second rule of detachment:

$$\begin{array}{r} P(B \rightarrow A) \geq r \\ P(B) \geq \rho \\ \hline \therefore P(A) \geq r + \rho - 1 \end{array} \tag{7}$$

The proof of (7) uses the general addition law rather than the theorem on total probability.

$$\begin{aligned} P(B \rightarrow A) &= P(\bar{B} \vee A) \\ &= P(\bar{B}) + P(A) - P(\bar{B} \,\&\, A) \\ &\geq r \,, \end{aligned}$$

whence, solving for $P(A)$,

$$\begin{aligned} P(A) &\geq r - P(\bar{B}) + P(\bar{B} \,\&\, A) \\ &\geq r - (1 - \rho) \\ &\geq r + \rho - 1 \,, \end{aligned}$$

as desired. The general form of (7) does not seem very enlightening, and we may get a better feeling for it if we take the special but important case that we want to claim both premises are known with near certainty, in particular, with probability equal to or greater than $1 - \epsilon$. We then have

$$
(8) \qquad \begin{array}{r} P(B \rightarrow A) \geq 1 - \epsilon \\ P(B) \geq 1 - \epsilon \\ \hline \therefore P(A) \geq 1 - 2\epsilon \end{array}
$$

It is worth noting that the form of the rule of detachment in terms of conditional probabilities does not lead to as much degradation from certainty as does (8), for

$$
(9) \qquad \begin{array}{r} P(A|B) \geq 1 - \epsilon \\ P(B) \geq 1 - \epsilon \\ \hline \therefore P(A) \geq (1 - \epsilon)^2 \end{array},
$$

and for $\epsilon > 0$, $(1 - \epsilon)^2 > 1 - 2\epsilon$. It is useful to have this well-defined difference between the two forms of detachment, for it is easy, on casual inspection, to think that ordinary-language conditionals can be translated equivalently in terms of conditional probability or in terms of the Boolean operation corresponding to material implication. Which _is_ the better choice I shall not pursue here, for application of either rule of inference does not require an auxiliary appeal to a court of total evidence.

Consideration of probabilistic rules of inference is not restricted to detachment. What is of interest is that classical sentential rules of inference naturally fall into two classes, those for which the probability of the conclusion is less than that of the individual premises, and those for which this degradation in degree of certainty does not occur. Tollendo ponens, tollendo tollens, the rule of adjunction (forming the conjunction), and the hypothetical syllogism all lead to a lower bound of $1 - 2\epsilon$ for the probability of the conclusion given that each of the two premises is assigned a probability of at least $1 - \epsilon$. The rules that use only one premise, e.g., the rule of addition (from A infer $A \vee B$), the rule of simplification, the commutative laws and de Morgan's laws assign a lower probability bound of $1 - \epsilon$ to the conclusion given that the premise has probability of at least $1 - \epsilon$.

We may generalize this last sort of example to the following theorem.

Theorem 1. *If* $P(A) \geq 1 - \epsilon$ *and* A *logically implies* B *then* $P(B) \geq 1 - \epsilon$.

Proof: We observe at once that if A logically implies B then $\bar{A} \cup B = X$, the whole sample space, and therefore $A \subseteq B$, but if $A \subseteq B$, then $P(A) \leq P(B)$, whence by hypothesis $P(B) \geq 1 - \epsilon$.

It is also clear that Theorem 1 can be immediately generalized to any finite set of premises.

Theorem 2. *If each of the premises* $A_1, \ldots, A_n$ *has probability of at least* $1 - \epsilon$ *and these premises logically imply* B *then* $P(B) \geq 1 - n\epsilon$.

Moreover, in general the lower bound of $1 - n\epsilon$ *cannot be improved on, i.e., equality holds in some cases whenever* $1 - n\epsilon \geq 0$.

Proof: By hypothesis for $i = 1, \ldots, n$, $P(A_i) \geq 1 - \epsilon$. We prove by induction that under this hypothesis $P(A_1 \& \ldots \& A_n) \geq 1 - n\epsilon$. The argument for $n = 1$ is immediate from the hypothesis. Suppose it holds for n. Then by an elementary computation

$$\begin{aligned} P(A_1 \& \ldots \& A_n \& A_{n+1}) &= 1 - (1 - P(A_1 \ldots \& A_n)) - (1 - P(A_{n+1})) \\ &\quad + P((\overline{A_1 \& \ldots A_n}) \& \bar{A}_{n+1}) \\ &\geq 1 - (1 - P(A_1 \ldots \& A_n)) - (1 - P(A_{n+1})) \\ &\geq 1 - n\epsilon - \epsilon \\ &\geq 1 - (n + 1)\epsilon , \end{aligned}$$

as desired. (Details of how to handle quantifiers, which are not explicitly treated in the standard probability discussions of the algebra of events, may be found in Gaifman (1965) or in the article by Krauss and Scott in this volume. The basic idea is to take as the obvious generalization of the finite case

$$P((\exists x)Ax) = \sup\{P(Aa_1 \vee Aa_2 \vee \ldots \vee Aa_n)\} ,$$

where the sup is taken over all finite sets of objects in the domain. Replacing sup by inf we obtain a corresponding expression for $P((\forall x)Ax)$. Apart from details it is evident that however quantifiers are handled, the assignment of probabilities must be such that Theorem 1 is satisfied, i.e., that if A logically implies B then the probability assigned to B must be at least as great as the probability assigned to A, and this is all that is required for the proof of Theorem 2.)

The proof that the lower bound $1 - n\epsilon$ cannot in general be improved upon reduces to constructing a case for which each of the n premises has probability $1 - \epsilon$, but the conjunction, as a logical consequence of the premises taken jointly has probability $1 - n\epsilon$, when $1 - n\epsilon \geq 0$. The example I use is most naturally thought of as a temporal sequence of events $A_1 \ldots , A_n$. Initially we assign

$$P(A_1) = 1 - \epsilon$$

$$P(\overline{A}_1) = \epsilon \; .$$

Then

$$P(A_2|A_1) = \frac{1-2\epsilon}{1-\epsilon}$$

$$P(A_2|\overline{A}_1) = 1 \; ,$$

and more generally

$$P(A_n|A_{n-1} A_{n-2} \ldots A_1) = \frac{1-n\epsilon}{1-(n-1)\epsilon}$$

$$P(A_n|A_{n-1} A_{n-2} \ldots \overline{A}_1) = 1$$

$$\vdots$$

$$P(\overline{A}_n|\overline{A}_{n-1} \overline{A}_{n-2} \ldots \overline{A}_1) = 1 \; ,$$

in other words for any combination of preceding events on trials 1 to $n-1$ the conditional probability of A_n is 1, except for the case $A_{n-1} A_{n-2} \ldots A_1$. The proof by induction that $P(A_n) = 1 - \epsilon$ and $P(A_n A_{n-1} \ldots A_1) = 1 - n\epsilon$ is straightforward. The case for $n = 1$ is trivial. Suppose now the assertion holds for n. Then by inductive hypothesis

$$P(A_{n+1} A_n \ldots A_1) = P(A_{n+1}|A_n \ldots A_1)P(A_n \ldots A_1)$$

$$= \frac{1-(n+1)\epsilon}{1-n\epsilon} \cdot 1 - n\epsilon$$

$$= 1 - (n + 1)\epsilon ,$$

and by the theorem on total probability

$$P(A_{n+1}) = P(A_{n+1}|A_n \ldots A_1)P(A_n \ldots A_1) + [P(A_{n+1}|A_n \ldots \bar{A}_1)P(A_n \ldots \bar{A}_1) + \ldots + P(A_{n+1}|\bar{A}_n \ldots \bar{A}_1)P(\bar{A}_n \ldots \bar{A}_1)] .$$

By construction all the conditional probabilities referred to in the bracketed expression are 1, and the unconditional probabilities in this expression by inductive hypothesis simply sum to $n\epsilon$, i.e., $1 - (1 - n\epsilon)$, whence

$$P(A_{n+1}) = \frac{1-(n+1)\epsilon}{1-n\epsilon} \cdot (1 - n\epsilon) + n\epsilon = 1 - \epsilon ,$$

which completes the proof.

It is worth noting that in interesting special cases the lower bound of $1 - n\epsilon$ can be very much improved. For example, if the premises $A_1, \ldots , A_n$ are statistically independent, then the bound is at least $(1 - \epsilon)^n$.

The intuitive content of Theorem 2 reflects a common-sense suspicion of arguments that are complex and depend on many premises, even when the logic seems impeccable. Overly elaborate arguments about politics, personal motives or circumstantial evidence are dubious just because of the uncertainty of the premises taken jointly rather than individually.

A natural question to ask about Theorem 2 is whether any non-deductive principles of inference that go beyond Theorem 2 arise from the imposition of the probability measure P on the algebra of events. Bayes' theorem provides an immediate example. To illustrate it with a simple artificial example, suppose we know that the composition of an urn of black (B) and white (W) balls may be exactly described by one of two hypotheses. According to hypothesis H_r, the proportion of white balls is r, and according to H_s, the proportion is s. Moreover, suppose we assign a priori probability ρ to H_r and $1 - \rho$ to H_s. Our four premises may then be expressed so:

$$P(W|H_r) = r$$

$$P(W|H_s) = s$$

$$P(H_r) = \rho$$

$$P(H_s) = 1 - \rho \ .$$

Given that we now draw with replacement, let us say, two white balls, we have as the likelihood of this event as a consequence of the first two premises

$$P(WW|H_r) = r^2$$

$$P(WW|H_s) = s^2 \ ,$$

and thus by Bayes' theorem, we may infer

(10) $$P(H_r|WW) = \frac{r^2\rho}{r^2\rho + s^2(1 - \rho)} ,$$

and this is clearly not a logical inference from the four premises. Logical purists may object to the designation of Bayes' theorem as a principle of inference, but there is little doubt that ordinary talk about inferring is very close to Bayesian ideas, as when we talk about predicting the weather or Jones' health, and such talk also has widespread currency among statisticians and the many kinds of people who use statistical methods to draw probabilistic inferences.

The present context is not an appropriate one in which to engage upon a full-scale analysis of the relation between logical and statistical inference. I have only been concerned here to establish two main points about inference. First, in terms of standard probability theory there is a natural form of probabilistic inference, and inference from probabilistically given premises involves no appeal to the concept of total evidence. Second, all forms of such probabilistic inference are not subsumed within the forms of logical inference, and two examples have been given to substantiate this claim, one being the rule of detachment as formulated for conditional probability and the other being Bayes' theorem.

3. The statistical syllogism re-examined. There is, however, a difficulty about the example of applying Bayes' theorem that is very similar to the earlier difficulty with the statistical syllogism. I have not stated as explicit premises the evidence WW that two white balls were drawn, and the reason I have not provides the key for re-analyzing the statistical syllogism and removing all air of paradox from it.

The evidence WW has not been included in the statement of the premises of the Bayesian example because the probability measure P referred to in the premises is the measure that holds <u>before</u> any taking of evidence (by drawing a ball) occurs. The measure P does provide a means of expressing the a posteriori probability <u>after</u> the evidence is taken as a conditional probability, but the hypothetical or conditional nature of this assertion has been too little appreciated. Using just the measure P there is no way to express that <u>in fact</u> two white balls were drawn, rather than, say, a white ball and then a black ball. Using conditional probabilities we can express the a posteriori probabilities of the two hypotheses under any <u>possible</u> outcomes of one, two or more drawings. What we cannot express in these terms is the <u>actual</u> evidence, and it is a mistake to try. (It should be apparent that these same remarks apply to Carnapian confirmation functions.) Commission of this mistake vitiates what appears to be the most natural symbolic formulation of the statistical syllogism--the inference form (5) as a special case of (3).

We can symbolize statistical syllogism (1) as follows, where $e(x)$ is the life expectancy of person x and $a(x)$ is the age of person x, and let j = Jones:

(11)
$$\begin{array}{l} P(e(j) \geq 15 \mid 50 < a(j) < 60) = r \\ 50 < a(j) < 60 \\ \hline \therefore P(e(j) \geq 15) = r \, . \end{array}$$

Now let us schematize this inference in terms of <u>hypothesis</u> and <u>evidence</u> as these notions occur in Bayes' theorem

(12) $$\frac{\begin{array}{l}P(\text{hypothesis}|\text{evidence}) = r \\ \text{evidence}\end{array}}{\therefore P(\text{hypothesis}) = r\ ,}$$

and the incorrect character of this inference is clear. From the standpoint of Bayes' theorem it asserts that once we know the evidence, the a posteriori probability $P(H|E)$ is equal to the a priori probability $P(H)$, and this is patently false. The difficulty is that the measure P cannot be used to assert that $P(50 < a(j) < 60) = 1$, which is, I take it, a direct consequence of the assertion that $50 < a(j) < 60$. (I shall expand on this point later.) The measure P is the measure used to express the conditional probabilities about Jones' life expectancy generated by any possible evidence. The inference-form expressed by (11) is illegitimate because the same probability measure does not apply to the two premises and the conclusion, as the scheme (12) makes clear when compared to Bayes' theorem.

Because there seems to be something genuine even if misleading about the statistical syllogism, it is natural to ask what are nonparadoxical ways of symbolizing it. One way is simply to adopt the symbolism used in Bayes' theorem, and then the conclusion is just the same as the first premise, the assertion of the a posteriori probability P(hypothesis| evidence). A related approach that makes the inference seem less trivial is the following. First, we symbolize the major premise in universal form, rather than with particular reference to Jones, for example:

The probability that a male resident of California will live at least fifteen years given that he is now between fifty and sixty years is r, or symbolically, where $m(x)$ is male resident of California,

$$P(e(x) \geq 15 \mid m(x) \,\&\, 50 < a(x) < 60) = r \ ,$$

and secondly, given as second premise the event A that

$$m(j) \,\&\, 50 < a(j) < 60 \ ,$$

we may write the conclusion in terms of a new probability measure P_A conditionalized on A: $P_A(e(j) \geq 15) = r$. Moreover, it is clear that no paradox arises from (2) because the evidence expressed in the second premise of (2) represents an event B distinct from A, and the conclusion $P_B(e(j) \geq 15) = s$, is consistent with the conclusion $P_A(e(j) \geq 15) = r$ of (1).

There is still another way of putting the matter which may provide additional insight into the inferential kernel inside the dubious statistical syllogism. We may think of the premises as all the a posteriori probabilities given all the different possible kinds of evidence. As an additional final premise, some evidence A is asserted. On this basis a new measure P_A is generated and the probability of the hypothesis is then asserted in terms of this new measure P_A, as the conclusion of the inference.

At this point it might seem easy to insist that delicate questions of consistency or coherence about the probability measure P do indeed differentiate deductive and inductive logic, but this is not at all the case.

The problem of temporal order of knowledge is as characteristic of deductive as of inductive logic. In discussing deductive canons of inference we tacitly assume the statements whose inferential relations are being considered are all asserted or denied at a given time or are timeless in character. It is not a paradox of deductive logic that the joint assertion of two statements true at different times leads to a paradox--for example, it rained yesterday, and it did not rain yesterday. The same thing, I have argued, is to be said about the statistical syllogism. The same probability measure does not apply to the first and second premise; the measure referred to in the first premise is temporally earlier than the one implicit in the second premise and the conclusion.

In the next section I turn to these temporal problems and their relation to the complex task of defining rationality, but before doing this I want briefly to pull several strands together and summarize in slightly different fashion the place given to the concept of total evidence by the view of probability advocated here.

According to this view it is automatic that if a person is asked for the probability of an event at a given time it will follow from the conditions of coherence on all his beliefs at that time that the probability he assigns to the event automatically takes into account the total evidence that he believes has relevance to the occurrence of the event. The way in which the total evidence is brought in is straightforward and simple through the theorem on total probability. To be quite clear how this theorem operates it may be useful to take a somewhat detailed look at the gradual expansion of the probability of an event A in terms of given evidence B and C. For purposes of generality we may assume that the probability of B and C is not precisely 1 and therefore deal with the general case. (In the interest of compactness of notation, I here let juxtaposition denote intersection or conjunction.) First we have

$$(13) \qquad P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) .$$

We also have

$$(14) \qquad P(A) = P(A|C)P(C) + P(A|\overline{C})P(\overline{C}) .$$

And in terms of both B and C we have the more complex version:

$$(15) \quad P(A) = P(A|BC)P(BC) + P(A|B\overline{C})P(B\overline{C}) + P(A|\overline{B}C)P(\overline{B}C) + P(A|\overline{BC})P(\overline{BC}) .$$

In the special case that $P(B) = P(C) = 1$, we then have

$$P(A) = P(A|B) = P(A|C) = P(A|BC) .$$

We have in the general case, as indications of the relations between ''partial'' and '' total'' evidence,

$$P(A|B)P(B) = P(A|BC)P(BC) + P(A|B\overline{C})P(B\overline{C})$$

and

$$P(A|\overline{B})P(\overline{B}) = P(A|\overline{B}C)P(\overline{B}C) + P(A|\overline{BC})P(\overline{BC}) .$$

The point of exhibiting these identities is to show that no separate concept of total evidence need be added to the concept of a probability measure on an individual's beliefs. It also may seem that these identities show that the notion of conditional probability is not even needed. The important point however is that the serious use contemplated here or the notion of conditional probability is in terms of passing from the probability measure expressing partial beliefs at one time to a later time. It is just by conditionalizing in terms of the events that actually occurred that this passage is made a good deal of the time by most information-processing organisms.

I conclude this paper with a more detailed look at the processes by which beliefs are changed.

4. Rational changes in belief. It seems important to recognize that the partial beliefs, or probability beliefs as we may term them, that an individual holds as a mature adult are not in any realistic way, even for an ideally rational individual, to be obtained simply be conditionalization, that is, in terms of conditional probability, from an overall probability measure which the individual was born with or acquired early in life. The patent absurdity of this idea does not seem to have been adequately reflected upon in Bayesian discussions of these matters. The static adherence to a single probability measure independent of time is characteristic of de Finetti (1937) and Savage (1954), but even a superficial appraisal of the development of a child's information-processing capacities makes it evident that other processes than conditionalization are required to explain the beliefs held by a mature adult. Moreover, even an adult who does not live in a terribly static and simple world will need other processes than conditionalization to explain the course of development of his beliefs during his years as an adult.

Acceptance of the view that a person's beliefs at time t are to be expressed by a probability measure special for that time raises certain problems that go beyond ordinary talk about probabilities. Under this view what probability are we to assign to events that a person knows have occurred? I can see no other course but to assert that such events have probability 1. Thus if I know that Jones is between fifty and sixty years of age then this event has probability 1 for me. We shall take as an axiom linking knowing and probabilistic beliefs as well as an axiom linking beliefs and probabilistic beliefs exactly this assertion, that is, if a person knows or believes that an event occurred then the probability of this event for that person is 1. For example, since I now believe with certainty in my own mind that Julius Caeser stayed several weeks in Gaul the probability of this event for me is 1.

The kind of problem that is solved by this axiom, and that is troublesome in a more detailed look at the standard Bayesian viewpoint, is the following. It is customary to remark, as has already been indicated, that the probability of an event should be conditionalized on that which has occurred. But it is also natural to ask what currently is the probability to be attached to an event that has occurred. Thus if the event B has occurred what probability is to be assigned to it (here I engage in the standard simplification of talking about the event occurring and not using the more careful locution of talking about knowing or believing that the event has occurred). It is not sufficient to say that we may refer to the probability of B given B because in this vein we can talk about the conditional probability of any event given that event. It is necessary to assign probability 1 to all such conditional probabilities independent of whether or not event B has actually occurred.

To put in another way an argument already given, once the continual adjustment of probability to the current state of affairs is supposed, and there is much in ordinary thinking and language that supports this idea, the problem of seeming to need to introduce a separate assumption about using all the evidence available simply disappears. As has been shown, it is an immediate consequence of the theorem on total probability that when we discuss the probability of an event A the relevance of any information we have about the event is immediately absorbed in a calculation of this probability, as a direct consequence of the theorem on total probability.

Unfortunately, there is another problem of total evidence, related to but not the same as, the one we started with. This new problem would seem to occupy a central position in any analysis of how we change our beliefs from one time to another. The problem is that of characterizing what part of the welter of potential information impinging on an organism is to be accepted as evidence and how this evidence is to be used to change the organism's basic probability measure. From the standpoint of psychology and physiology, a satisfactory empirical answer seems a great distance away. The fact of our empirical ignorance about matters conceptually so close to central problems of epistemology is philosophically important, but of still greater philosophical importance is the fact that our general concept of rationality seems intimately involved in any answer, empirical or not, that we give. Until we can say how an organism of given capacities and powers should process the information impinging on it we cannot have a very deep running concept of rationality.

The point that I want to make here needs some delineation. The problem is not one of giving many instances in which there is great agreement on what is the rational way to process information. If a man is in a house that is burning, he is irrational calmly to continue to listen to music, or if a man is driving down a highway at high speed, he is irrational if he becomes absorbed in the beauty of the landscape and looks for forty or fifty seconds at an angle nearly perpendicular to the road itself. These simple instances can be multiplied to any extent desired, but what is not easy to multiply is the formal characterization of the principles that should be applied and that govern the variety of cases in which we all have a clear intuitive judgment.

To a large extent work in inductive logic and statistical inference tend to obscure the fundamental character of this problem of giving principles by which information is to be judged important and to be responded to. The reason for this neglect in inductive logic or theoretical statistics is that once the formal language or the random variables are selected, then the problem of information-processing is reduced to relatively simple proportions. The selection of the language or the selection of the random variables, as the case may be, is the largest single decision determining how information will be processed, and to a very large extent the simple rule of conditionalizing, so that the measure held at a later time arises from an earlier probability measure as a conditional probability measure, then furnishes the appropriate way to proceed. It might almost be said that the rule of pure prudence is always to derive beliefs held at time t from beliefs held at time t' by conditionalization on the probability measure characterizing partial beliefs at the earlier time. Although such a principle of pure prudence may seem attractive at first glance, in my own judgment, it is a piece of pure fantasy. A myriad of events are occurring at all times and are noticeable by a person's perceptual apparatus. What is not the least bit clear is what sort of filter should be imposed by the individual on this myriad of events in order to have a workable simplifying scheme of decision and action. The highly selective principles of attention that must necessarily be at work do not seem to be characterizable in any direct way from the concept of conditional probability.

The two interrelated processes that any adequate theory of rationality must characterize are the process of information selection and the process of changing beliefs according to the significance of the information that is selected. In other words, to what should the rational man pay attention in his ever-changing environment and how should he use the evidence of that to which he does attend in order to change his beliefs and thereby modify his decisions and actions.

What seems particularly difficult to describe are rational mechanisms of selection. In a first approximation the classical theory of maximizing expected utility as developed by Ramsey, de Finetti, Savage and others uses the mechanism of conditional probabilities to change beliefs once the information to be attended to has been selected. This classical mechanism is certainly inadequate for any process of concept formation, and thus for any very deep running change in belief, and as Jeffrey (1965) points out, it is not even adequate to the many cases in which changes in belief may be expressed simply as changes in probability but not explicitly in terms of changes in conditional probability, because the changes in probability are not completely analyzable in terms of the explicitly noticed occurrence of events. For example, the probability that I will assign to the event of rain tomorrow will change from morning to afternoon, even though I am not able to express in explicit form the evidence I have used in coming to this change. In another article in this volume

I have attempted to show how inadequate the Bayesian approach of conditional probability is in terms of even fairly simple processes of concept formation, and I do not want to go over that ground again here, except to remark that, it is clear on the most casual inspection that all information processing that an organism engages in cannot be conceived of in terms of conditional probabilities.

It is even possible to question whether <u>any</u> changes can be so expressed. The measure P effectively expressing my beliefs at time t cannot be used to express what I actually observe immediately after time t, for P is already "used up" so to speak in expressing the a priori probability of each possible event that might occur, and cannot be used to express the unconditional occurrence of that which in fact did happen at time t. Pragmatically, the situation is clear. If an event A occurs and is noticed, the individual then changes his belief pattern from the measure P to the conditional measure P_A, and our reformulated version of the statistical syllogism is exemplified. What has not been adequately commented upon in discussions of these matters by Bayesians is that the probability measure P held at time t cannot be used to express what actually happened immediately after t, but only to express, at the most, how P would change <u>if</u> so and so did happen.

In any case, genuine changes in conceptual apparatus cannot in any obvious way be brought within this framework of conditional probability. As far as I can see, the introduction of no genuinely new concept in the history of scientific or ordinary experience can be adequately accounted for in terms of conditional probabilities. A fundamental change in the algebra of events itself, not just their probabilities, is required to account for such conceptual changes. Again, it is a problem and a responsibility of an adequate theory of normative information processing to give an account of how such changes should take place.

What I would like to emphasize in conclusion is the difficulty of expressing in systematic form the mechanisms of attention a rationally-operating organism should use. It is also worth noting that any interesting concept of rationality for the behavior of men must be keyed to a realistic appraisal of the powers and limitation of the perceptual and mental apparatus of humans. These problems of attention and selection do not exist for an omniscient God or an information-processing machine whose inputs are already schematized into something extraordinarily simple and regular. The difficulty and subtlety of characterizing the mechanisms of information selection and at the same time a recognition of their importance in determining the actual behavior of men make me doubt that the rather simple Carnapian conception of inductive logic can be of much help in developing an adequate theory of rational behavior. Even the more powerful Bayesian approach provides a very thin and inadequate characterization of rationality, because only one simple method for changing beliefs is admitted. It is my own view that there is little chance of defining an adequate concept of rationality until analytical tools are available to forge a sturdy link between the selection and use of evidence and processes of concept formation.

REFERENCES

Barker, S. *Induction and hypotheses*. Ithaca, New York: Cornell Univer. Press, 1957.

Carnap, R. *The logical foundations of probability*. Chicago: Univer. of Chicago Press, 1950.

de Finetti, B. La Prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. Poincare*, 1937, *7*, 1-68. English translation in H. E. Kyburg, Jr., and H. E. Smokler (Eds.), *Studies in subjective probability*. New York: Wiley, 1964. Pp. 93-158.

Gaifman, H. Concerning measures in first-order calculi. *Israel Journal of Mathematics*, 1964, *2*, 1-18.

Hempel, C. G. *Aspects of scientific explanation*. New York: The Free Press, 1965.

Jeffrey, R. *The logic of decisions*. New York: McGraw-Hill, 1965.

Savage, L. J. *The foundations of statistics*. New York: Wiley and Co., 1954.

Suppes, P. and Atkinson, R. C. *Markov learning models for multiperson interaction*. Stanford: Stanford Univer. Press, 1960.