The Measurement of Belief

BY

PATRICK SUPPES

Reprinted from

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY

SERIES B (METHODOLOGICAL)

Volume 36, No. 2, 1974

(pp. 160–191)

PRINTED FOR PRIVATE CIRCULATION

1974

The Measurement of Belief

By PATRICK SUPPES

Institute for Mathematical Studies in the Social Sciences, Stanford University

[Read before the ROYAL STATISTICAL SOCIETY and the BRITISH PSYCHOLOGICAL SOCIETY at a meeting organized by the RESEARCH SECTION and the MATHEMATICAL and STATISTICAL PSYCHOLOGY SECTION on Wednesday, February 13th, 1974, Professors J. GANI and M. HAMILTON in the Chair]

SUMMARY

This paper criticizes some of the claims of the standard theories of subjective probability. The criticisms are especially oriented toward the structural axioms that cannot be regarded as axioms of pure rationality and the general results that yield exact measurement of subjective probabilities. Qualitative axioms for upper and lower probability are introduced to provide a theory of inexact measurement of subjective probability. Only minor modifications of de Finetti's qualitative axioms yield the desired theory. The paper concludes with a comparison of the measurement of belief to the measurement results characteristic of Euclidean geometry, and also examines briefly some possibilities for using learning models as simplified abstract processes for constructing belief.

Keywords: BELIEF; MEASUREMENT; LOWER PROBABILITY; UPPER PROBABILITY; LEARNING

1. INTRODUCTION

ALMOST everyone who has thought about the problems of measuring beliefs in the tradition of subjective probability or Bayesian statistical procedures concedes some uneasiness with the problem of always asking for the next decimal of accuracy in the prior estimation of a probability or of asking for the parameter of a distribution that determines the probabilities of events. On the other hand, the formal theories that have been developed for rational decision-making under uncertainty by Ramsey (1931), de Finetti (1931, 1937), Koopman (1940a, b), Savage (1954) and subsequent authors have almost uniformly tended to yield a result that guarantees a unique probability distribution on states of nature or whatever other collection of entities is used for the expression of prior beliefs.

In the next section I examine some of these standard theories and address the question of how we can best criticize the claims they make. Among other points, I consider the claim that the idealizations expressed in the axioms can be regarded as theories of pure rationality.

In the third section I examine two constructive possibilities that yield inexact measurements of belief. Because the issues are almost entirely conceptual and not technical at the present stage of investigation, I confine myself to comparing some elementary axiom systems and raise the question of their suitability as a basis for empirical investigation. The first system is relatively trivial, but is designed to make a certain conceptual point. The second system is considerably more interesting and presents, I think, a useful approach to inexact measurement of subjective probability, with a representation theorem formulated in terms of upper and lower probabilities.

In the final section I compare the measurement of belief to the classical theory of measurement embodied in Euclidean geometry and challenge the view that

idealizations of exact measurement are as useful and harmless in the case of the theory of beliefs as they are in the case of geometry. I also briefly compare the situation with that which exists in quantum mechanics and meteorology and argue for the conclusion that the inexact results of these sciences are a more appropriate model than that of geometry. More importantly, I try to state in this section some unfinished ideas about processes for constructing beliefs.

I do not have as much to say about empirical matters in this article as I would like. A common view I share is that the conceptual and formal analysis of belief structures has currently far outstripped the empirical study of beliefs, and probably what is needed most at the present time are several relentless programs of empirical investigation guided and motivated by the insights afforded from various formal concepts and theories that are mathematically now well understood.

2. Weaknesses of the Standard Theories

Because the standard theories mentioned earlier reach essentially the same formal results, namely, the existence of a unique probability distribution on states of nature, criticisms of one will pretty much apply to criticisms of the lot. For this reason, it may pay to concentrate on Savage's (1954) axioms, because of their familiarity to a wide audience and because they have been much discussed in the literature. I emphasize, however, that what I have to say about Savage's axioms will apply essentially without change to other standard theories.

Because Savage's axioms are rather complicated from a formal standpoint, I shall not state them explicitly here, but shall try to describe their intuitive content. The axioms are about preference among decisions, and decisions are mappings or functions from the set of states of nature to the set of consequences. To illustrate these ideas, let me use an example I have used before (Suppes, 1956).

A certain independent distributor of bread must place his order for a given day by ten o'clock of the preceding evening. His sales to independent grocers are affected by whether or not it is raining at the time of delivery, for if it is raining, the grocers tend to buy less on the accumulated evidence that they have fewer customers. On a rainy day the maximum the distributor can sell is 700 loaves; on such a day he makes less money if he has ordered more than 700 loaves. On the other hand, when the weather is fair, he can sell about 900 loaves. If the simplifying assumption is made that the consequences to him of a given decision with a given state of nature (s_1—rain or s_2—no rain) may be summarized simply in terms of his net profits, the situation facing him is represented in Table 1. The distributor's problem is to make a decision.

TABLE 1

	d_1—*buy* 700 *loaves*	d_2—*buy* 800 *loaves*	d_3—*buy* 900 *loaves*
s_1—rain	\$21.00	\$19.00	\$17.00
s_2—no rain	\$21.00	\$24.00	\$26.50

Clearly, if he knows for certain that it is going to rain, he should make decision d_1, and if he knows for certain that it is not going to rain, he should make decision d_3. The point of Savage's theory, expanded to more general and more complex situations,

is to place axioms on choices or preferences among the decisions in such a way that anyone who satisfies the axioms will be maximizing expected utility. This means that the way in which he satisfies the axioms will generate a subjective probability distribution about his beliefs concerning the true state of nature and a utility function on the set of consequences such that the expectation of a given decision is defined in a straightforward way with respect to the subjective probability distribution on states of nature and the utility function on the set of consequences. As one would expect, Savage demands, in fact in his first axiom, that the preference among decisions be transitive and that given any two decisions one is at least weakly preferred to the other. Axiom 2 extends this ordering assumption to having the same property hold when the domain of definition of decisions is restricted to a given set of states of nature; for example, the decision-maker might know that the true state of nature lies in some subset of the whole set. Axiom 3 asserts that knowledge of an event cannot change preferences among consequences, where preferences among consequences are defined in terms of preferences among decisions. Axiom 4 requires that given any two sets of states of nature, that is, any two events, one is at least as probable as the other, that is, qualitative probability among events is strongly connected. Axiom 5 excludes the trivial case in which all consequences are equivalent in utility and, thus, every decision is equivalent to every other. Axiom 6 says essentially that if event A is less probable than event B (A and B are subsets of the same set of states of nature), then there is a partition of the states of nature such that the union of each element of the partition with A is less probable than B. As is well known, this axiom of Savage's is closely related to the axiom of de Finetti and Koopman, which requires the existence of a partition of the states of nature into arbitrarily many events that are equivalent in probability. Finally, his last axiom, Axiom 7, is a formulation of the sure-thing principle.

My first major claim is that some of Savage's axioms do not in any direct sense represent axioms of rationality that should be satisfied by any ideally rational person but, rather, they represent structural assumptions about the environment that may or may not be satisfied in given applications.

Many years ago, at the time of the Third Berkeley Symposium (1955), I introduced the distinction between structure axioms and rationality axioms in the theory of decision-making (Suppes, 1956). Intuitively, a structure axiom as opposed to a rationality axiom is existential in character. In the case of Savage's seven postulates, two (5 and 6) are structure axioms, because they are existential in character.

Savage defended his strong Axiom 6 by holding it applicable if there is a coin that a decision-maker believes is fair for any finite sequence of flips. There are, however, several objections to this argument. First of all, if it is taken seriously then one ought to redo the entire foundations and simply build it around Bernoulli sequences with $p = 0{\cdot}5$ and get arbitrarily close approximations to the probability of any desired event. (See the second system of axioms in the next section.) More importantly, without radical changes in human thinking, it is simply not natural on the part of human beings to think of finite sequences of flips of a coin in evaluating likelihoods or probabilities, qualitative or quantitative, of significant events with which they are concerned.

Consider the case of a patient's deciding whether to follow a surgeon's advice to have major surgery. The surgeon, let us suppose, has evaluated the pros and cons of the operation, and the patient is now faced with the critical decision of whether to take the risk of major surgery with at least a positive probability of death, or whether

to take the risk of having no surgery and suffering the consequences of the continuing disease. I find it very unlikely and psychologically very unrealistic to believe that thinking about finite sequences of flips of a fair coin will be of any help in making a rational decision on the part of the patient.

On the other hand, other axioms like those on the ordering of preferences or qualitative probability seem reasonable in this framework and are not difficult to accept. But the important point is this. In a case in which uncertainty has a central role, in practice, decisions are made without any attempt to reach the state of having a quantitative probability estimate of the alternatives or, if you like, a computed expected utility.

It is, in fact, my conviction that we usually deal with restricted situations in which the set of decisions open to us is small and in which the events that we consider relevant are small in number. The kind of enlarged decision framework provided by standard theories is precisely the source of the uneasiness alluded to in the first sentence of the Introduction. Intuitively we all move away from the idea of estimating probabilities with arbitrary refinement. We move away as well from the introduction of an elaborate mechanism of randomization in order to have a sufficiently large decision space. Indeed, given the Bayesian attitude towards randomization, there is an air of paradox about the introduction *à la* Savage of finite sequences of tosses of a fair coin.

Another way of putting the matter, it seems to me, is that there is a strong intuitive feeling that a decision-maker is not irrational simply because a wide range of decision possibilities or events is not available to him. It is not a part of rationality to require that the decision-maker enlarge his decision space, for example, by adding a coin that may be flipped any finite number of times. I feel that the intrinsic theory of rationality should be prepared to deal with a given set of states of nature and a given set of decision functions, and it is the responsibility of the formal theory of belief or decision to provide a theory of how to deal with these restricted situations without introducing strong structural assumptions.

A technical way of phrasing what I am saying about axioms of pure rationality is the following. For the moment, to keep the technical apparatus simple, let us restrict ourselves to a basic set S of states of nature and a binary ordering relation of qualitative probability on subsets of S, with the usual Boolean operations of union, intersection and complementation having their intuitive meaning in terms of events. I then say that an axiom about such structures is an axiom of pure rationality only if it is closed under submodels. Technically, closure under submodels means that if the axiom is satisfied for a pair $\langle S, \geqslant \rangle$ then it is satisfied for any non-empty subset of S with the binary relation $\geqslant$ restricted to the power set of the subset, i.e. restricted to the set of all subsets of the given subset. (Of course, the operations of union, intersection and complementation are closed in the power set of the subset.) Using this technical definition, we can easily see that of Savage's seven axioms, five of them satisfy this restriction, and the two already mentioned as structure axioms do not.

Let me try to make somewhat more explicit the intuition which is behind the requirement that axioms of pure rationality should satisfy the condition of closure under submodels. One kind of application of the condition is close to the axiom on the independence of irrelevant alternatives in the theory of choice. This axiom says that if we express a preference among candidates for office, for example, and if one candidate is removed from the list by death or for other reasons, then our ordering of preferences among the remaining candidates should be unchanged. This axiom

satisfies closure under submodels. The core idea is that existential requirements that reach out and make special requirements on the environment do not represent demands of pure rationality but rather structural demands on the environment, and such existential demands are ruled out by the condition of closure under submodels.

A different, but closely related, way of defining axioms of pure rationality is that such an axiom must be a logical consequence of the existence of the intended numerical representation. This criterion, which I shall call the *criterion of representational consequence*, can be taken as both necessary and sufficient, whereas the criterion of closure under submodels is obviously not sufficient. On the other hand, the extrinsic character of the criterion of representational consequence can be regarded as unsatisfactory. It is useful for identifying axioms that are not necessary for the intended representation and thus smuggle in some unwanted arbitrary structural assumption. As should be clear, Savage's Axioms 5 and 6 do such smuggling.

I am quite willing to grant the point that axioms of rationality of a more restricted kind could be considered. One could argue that we need special axioms of rationality for special situations, and that we should embark on a taxonomy of situations providing appropriate axioms for each of the major classes of the taxonomy. In the present primitive state of analysis, however, it seems desirable to begin with a sharp distinction between rationality and structure axioms and to have the concept of pure rationality universal in character.

Returning now to my criticisms of Savage's theory, it is easy to give finite or infinite models of Savage's five axioms of rationality for which there exists no numerical representation in terms of utility and subjective probability. In the language I am using here, Savage's axioms of pure rationality are insufficient for establishing the existence of representing numerical utility and subjective probability functions.

Moreover, we may show that no finite list of additional elementary axioms of a universal character will be sufficient to guarantee the existence of appropriate numerical functions. By *elementary axioms* I mean axioms that can be expressed within first-order logic. First-order logic essentially consists of the conceptual apparatus of sentential connectives, one level of variables and quantifiers for these variables, together with non-logical predicates, operation symbols and individual constants. Thus, for example, the standard axioms for groups or for ordered algebraic fields are elementary, but the least upper-bound axiom for the field of real numbers is not. It is possible to formulate Savage's Axiom 5 in an elementary way, but not his Axiom 6.

In the case of infinite models, the insufficiency of elementary axioms, without restriction to their being of a universal character, follows from the upward Lowenheim–Skolem–Tarski theorem, plus some weak general assumptions. This theorem asserts that if a set of elementary axioms has an infinite model (i.e. a model whose domain is an infinite set, as is the case for Savage's theory), then it has a model of every infinite cardinality. Under quite general assumptions, e.g. on the ordering relation of preference or greater subjective probability, it is impossible to map the models of high infinite cardinality into the real numbers, and thus no numerical representation exists.

In the case of finite models, the methods of Scott and Suppes (1958) apply to show that no finite set of universal elementary axioms will suffice. The system consisting of Savage's five axioms of pure rationality has finite models, but by the methods indicated we can show there is no finite elementary extension by means of universal axioms of rationality that will be strong enough to lead to the standard

numerical representation. (The essential idea of Scott and Suppes' work is to show that if for every positive integer n there is a finite model M such that every submodel of n elements satisfies the theory in question, but the model M does not, then the theory is not axiomatizable by a finite list of elementary axioms that are universal in form.)

The results I have outlined indicate the nature of some of the general restrictions that obtain in the hope of finding elementary axioms of pure rationality sufficient to lead to an appropriate numerical representation of the decision situation.

On the other hand, in the case of finite models, necessary and sufficient conditions can be given, and using the criterion of closure under submodels as a criterion of pure rationality, we then have formally adequate axioms of pure rationality in the finite case, even if the conditions are not fixed in number, but are represented by a potentially infinite schema.

The simplest and most elegant version of such axioms is probably that given by Scott (1964) for the de Finetti framework of qualitative subjective probability in which decisions and consequences are not explicitly considered. (His axioms improve on the earlier ones given by Kraft *et al.*, 1959.) Because I want to comment on their character from the standpoint developed in this paper, Scott's axioms are embodied in the following definition, in which the notation A^c is used for the characteristic function of a set A, and $\emptyset$ for the empty set.

Definition 1. Let X be a non-empty finite set and $\geqslant$ a binary relation on the set of all subsets of X. Then a structure $\langle X, \geqslant \rangle$ is a (finite) qualitative belief structure if and only if for all subsets A and B of X

Axiom 1. $A \geqslant B$ or $B \geqslant A$;

Axiom 2. $A \geqslant \emptyset$;

Axiom 3. $X > \emptyset$;

Axiom 4. For all subsets $A_0, \ldots, A_n, B_0, \ldots, B_n$ of X, if $A_i \geqslant B_i$ for $0 \leqslant i < n$, and for all x in X

$$A_0^c(x) + \ldots + A_n^c(x) = B_0^c(x) + \ldots + B_n^c(x),$$

then $B_n \geqslant A_n$.

Axiom 4 only requires that any element of X, that is, any atomic event, belong to exactly the same number of A_i and B_i, for $0 \leqslant i \leqslant n$. To illustrate the force of Scott's Axiom 4, we may see how it implies transitivity. First, necessarily for any three characteristic functions

$$A^c + B^c + C^c = B^c + C^c + A^c;$$

that is, for all elements x of X

$$A^c(x) + B^c(x) + C^c(x) = B^c(x) + C^c(x) + A^c(x).$$

By hypothesis, $A \geqslant B$ and $B \geqslant C$, whence by virtue of Axiom 4,

$$C \leqslant A,$$

and thus, by definition $A \geqslant C$, as desired. Scott proves that for any finite structure $\mathscr{X} = \langle X, \geqslant \rangle$ satisfying the axioms of Definition 1 there is a probability measure P such that for A and B subsets of X

$$A \geqslant B \quad \text{if and only if } P(A) \geqslant P(B).$$

A first point to note is that the probability measure P is not unique, nor apparently can its uniqueness up to a given set of transformations be characterized in an interesting way, a situation that is true for many finite geometries when the set of transformations is as general as possible consistent with the finite number of relationships expressed.

The more profound difficulty with Scott's axioms as a theory of belief is the combinatorial explosion that occurs in verifying the axioms when the number of events is large. To check connectedness, for example, we need only consider pairs of events, and to check transitivity, only triples of events. But, it is fundamental for the kind of axiom schema (Scott's Axiom 4) required to express necessary and sufficient conditions in the finite case that n-tuples of events of arbitrary n must be studied as the number of events increases. As a possible empirical theory of belief, or as a rational one, this seems impractical, and even for fairly small experiments, the effort to determine whether there is a representing probability measure requires the use of a moderate-sized computer facility. Certainly the experiments do not themselves check all the possible n-tuples of comparison. Again, I will not enter into detailed computations, but in conducting some unpublished experiments on measuring beliefs some years ago, already I found that in considering a space with ten atoms, a small number for complex matters, the combinatorial explosion of possible comparisons of pairs of events (not necessarily atomic) was impressive. (Talk about atoms is just another way of talking about the points in a sample space.) If we deal with 30 or 40 or 50 atoms, the numbers are out of hand, even when we take maximal advantage of relationships implied by the axioms.

3. Inexact Measurement

In thinking about these problems once again, I asked myself what are the simplest axioms that would minimize the number of comparisons needed, and that would still yield some results on the underlying measure if it is there. You may find the following axioms amusing. Although I do not propose them as a serious set to be used in extensive studies of actual beliefs, I do advance them as one modest conceptual model of how far we can go in simplifying the comparisons we ask for, and yet obtain some kind of results different from those of simple order if the axioms are satisfied.

The intuitive idea of the restricted system is to have five classes of events: Those that are certain (C), those that are more likely than not (M), those that are less likely than not (L), those that are as likely as not (E) and those that are impossible (I). However, only two of these five classes of events need be taken as primitive. For example, taking the class of certain events and the class of events that are more likely than not as primitive, we can define the other three in the following manner, where if A is an event, then *not* A is of course the event that occurs if A does not, i.e. the complement of A: *A is impossible* if and only if *not* A is certain; *A is less likely than not* if and only if *not* A is more likely than not; *A is as likely as not* if and only if A is neither certain, impossible, more likely than not, nor less likely than not.

Let X be a non-empty set and let events be subsets of X. Then the axioms of what I shall call *weak qualitative probability structures* are the following:

Axiom 1. X is certain.

Axiom 2. If A implies B and A is certain, then B is certain.

Axiom 3. If A implies B and A is more likely than not, then B is more likely than not.

Axiom 4. If A implies B but B does not imply A and A is as likely as not, then B is more likely than not.

Axiom 5. If A is certain, then *not* A is impossible.

Axiom 6. If A is more likely than not, then *not* A is less likely than not.

(A completely formal version of these axioms can easily be given.)

From the axioms we can easily prove the following sorts of elementary theorems: If A implies B and B is less likely than not, then A is less likely than not; if A is as likely as not, then *not* A is as likely as not; if A is as likely as not, B is as likely as not, and A and B are mutually exclusive, then the disjunction A or B is certain. (The proof of the last assertion uses Axiom 4.)

In many cases the situation described by these axioms is about the appropriate degree of crudeness of what a person knows about his beliefs. Even in the present framework we can add axioms that will force the situation to be much tighter. These axioms are of course structural axioms and in general will not be satisfied in a given situation. For example, we can require that every atom be less likely than not, but still not impossible, and also that if an event is less likely than not, then there is some second event such that the disjunction of the two is as likely as not. When these structural assumptions are added, we can show that a system with three atoms is impossible, and a system of four atoms requires that they be equally probable. The three-atom case is easy to see. By way of contradiction, let x, y and z be the numerical probabilities of the three atoms. By hypothesis $x<\frac{1}{2}$, and thus also by hypothesis either $x+y=\frac{1}{2}$ or $x+z=\frac{1}{2}$, but in the first case then $z=\frac{1}{2}$, contrary to assumption, and in the second case, $y=\frac{1}{2}$, also contrary to assumption. I shall not explore the situation in more detail, because it seems to me that these particular structural axioms are not especially interesting. I merely state them as an indication of the kind of results we can get by some relatively innocent-appearing structural assumptions. Notice that even with the structural atoms, we are not able to prove that there is an ordering of events in terms of less probable and more probable.

For weak qualitative probability structures, we can prove a representation theorem.

Theorem 1. If X is finite or countable, and $\langle X, C, M\rangle$ is a weak qualitative probability structure, then there is a probability measure defined on the power set of X such that

(i) $P(A)=1$ if and only if A is certain,
(ii) $P(A)>\frac{1}{2}$ if and only if A is more likely than not.

From the definitions given above it follows that $P(A)<\frac{1}{2}$ if and only if A is less likely than not, $P(A)=\frac{1}{2}$ if and only if A is as likely as not and $P(A)=0$ if and only if A is impossible.

As a final system of axioms, I want to introduce purely in terms of belief or subjective probability what I consider the appropriate finitistic analogue of Savage's axioms. These constitute an extension of de Finetti's qualitative conditions and lead to simple approximate measurement of belief in arbitrary events. The axioms require something that I partly criticized earlier, namely, the existence of some standard set of events whose probability is known exactly. They would, for example, be satisfied by flipping a fair coin n times for some fixed n. They do not require that n be indefinitely large and therefore n may be looked upon as somewhat more realistic. I give the axioms here in spite of my feeling that, from the standpoint of a serious decision like that on surgery mentioned earlier, they may be unsatisfactory.

They do provide a combination of de Finetti's ideas and a finite version of the standard structural axiom on infinite partitions.

The concept of upper and lower probabilities seems to be rather recent in the literature, but it is obviously closely related to the classical concepts of inner and outer measure, which were introduced by Caratheodory and others at the end of the nineteenth century and the beginning of this century. Koopman (1940b) explicitly introduces lower and upper probabilities but does nothing with them from a conceptual standpoint. He uses them as a technical device, as in the case of upper and lower measures in mathematical analysis, to define probabilities. The first explicit conceptual discussions seem to be quite recent (Smith, 1961; Good, 1962). Smith especially enters into many of the important conceptual considerations, and Good states a number of the quantitative properties it seems natural to impose on upper and lower probabilities. Applications to problems of statistical inference are to be found in Dempster (1967). However, so far as I know, a simple axiomatic treatment starting from purely qualitative axioms does not yet exist in the literature, and the axioms given below represent such an effort. It is apparent that they are not the most general axioms possible, but they do provide a simple and hopefully rather elegant qualitative base.

From a formal standpoint, the basic structures to which the axioms apply are quadruples $\langle X, \mathscr{F}, \mathscr{S}, \geqslant \rangle$, where X is a non-empty set, $\mathscr{F}$ is an algebra of subsets of X, that is, $\mathscr{F}$ is a non-empty family of subsets of X and is closed under union and complementation, $\mathscr{S}$ is a similar algebra of sets, intuitively the events that are used for standard measurements, and I shall refer to the events in $\mathscr{S}$ as *standard* events S, T, etc. The relation $\geqslant$ is the familiar ordering relation on $\mathscr{F}$. I use familiar abbreviations for equivalence and strict ordering in terms of the weak ordering relation. (A weak ordering is transitive and strongly connected, i.e. for any events A and B, either $A \geqslant B$ or $B \geqslant A$.)

Definition 2. A structure $\mathscr{X} = \langle X, \mathscr{F}, \mathscr{S}, \geqslant \rangle$ is a finite approximate measurement structure for beliefs if and only if X is a non-empty set, $\mathscr{F}$ and $\mathscr{S}$ are algebras of sets on X, and the following axioms are satisfied for every A, B and C in $\mathscr{F}$ and every S and T in $\mathscr{S}$:

Axiom 1. The relation $\geqslant$ is a weak ordering of $\mathscr{F}$;

Axiom 2. If $A \cap C = \emptyset$ and $B \cap C = \emptyset$ then $A \geqslant B$ if and only if $A \cup C \geqslant B \cup C$;

Axiom 3. $A \geqslant \emptyset$;

Axiom 4. $X > \emptyset$;

Axiom 5. $\mathscr{S}$ is a finite subset of $\mathscr{F}$;

Axiom 6. If $S \neq \emptyset$ then $S > \emptyset$;

Axiom 7. If $S \geqslant T$ then there is a V in $\mathscr{S}$ such that $S \approx T \cup V$.

In comparing Axioms 3 and 6, note that A is an arbitrary element of the general algebra $\mathscr{F}$, but event S (referred to in Axiom 6) is an arbitrary element of the subalgebra $\mathscr{S}$. Also in Axiom 7, S and T are standard events in the subalgebra $\mathscr{S}$, not arbitrary events in the general algebra. Axioms 1–4 are just the familiar de Finetti axioms without any change. Because all the standard events (finite in number) are also events (Axiom 5), Axioms 1–4 hold for standard events as well as arbitrary events. Axiom 6 guarantees that every minimal element of the subalgebra $\mathscr{S}$ has positive qualitative probability. Technically a minimal element of $\mathscr{S}$ is any event A in $\mathscr{S}$ such that $A \neq \emptyset$, and it is not the case that there is a non-empty B in $\mathscr{S}$ such that B is a proper subset of A. A *minimal open interval* (S, S') of $\mathscr{S}$ is such that $S < S'$ and $S' - S$ is equivalent to a minimal element of $\mathscr{S}$. Axiom 7 is the main

structural axiom, which holds only for the subalgebra and not for the general algebra; it formulates an extremely simple solvability condition for standard events. It was stated in this form in Suppes (1969, p. 6) but in this earlier case for the general algebra $\mathscr{F}$.

In stating the representation and uniqueness theorem for structures satisfying Definition 3, in addition to an ordinary probability measure on the standard events, I shall use upper and lower probabilities to express the inexact measurement of arbitrary events. A good discussion of the quantitative properties one expects of such upper and lower probabilities is found in Good (1962). All of his properties are not needed here because he dealt with conditional probabilities. The following properties are fundamental, where $P_*(A)$ is the lower probability of an event A and $P^*(A)$ is the upper probability (for every A and B in $\mathscr{F}$):

I. $P_*(A) \geqslant 0$.
II. $P_*(X) = P^*(X) = 1$.
III. If $A \cap B = \varnothing$ then

$$P_*(A)+P_*(B) \leqslant P_*(A \cup B) \leqslant P_*(A)+P^*(B) \leqslant P^*(A \cup B) \leqslant P^*(A)+P^*(B).$$

Condition (I) corresponds to Good's Axiom D2 and (III) to his Axiom D3.

For standard events $P(S) = P_*(S) = P^*(S)$. For an arbitrary event A not equivalent in qualitative probability to a standard event, I think of its "true" probability as lying in the open interval $(P_*(A), P^*(A))$.

Originally I included as a fourth property

$$P_*(A)+P^*(\neg A) = 1,$$

where $\neg A$ is the complement of A, but Mario Zanotti pointed out to me that this property follows from (II) and (III) by the following argument:

$$1 = P_*(X) = P_*(A \cup \neg A) \leqslant P_*(A)+P^*(\neg A) \leqslant P^*(A \cup \neg A) = P^*(X) = 1.$$

A stronger property possessed by some upper and lower measures is this:

IV. $P_*(A \cup B)+P_*(A \cap B) \geqslant P_*(A)+P_*(B)$.

Good mentions that he suspected that this principle is independent of the others he introduces. (He actually states the dual form in terms of upper probabilities.) After the proof of Theorem 2, I give a counterexample to show that (IV) does not hold for every qualitative structure satisfying Definition 3.

In the fourth part of Theorem 2, I define a certain relation and state it is a semiorder with an implication from the semiorder relation holding to an inequality for upper and lower probabilities. Semiorders have been fairly widely discussed in the literature as a generalization of simple orders first introduced by Duncan Luce. I use here the axioms given by Scott and Suppes (1958). A structure $\langle U, R\rangle$ where U is a non-empty set and R is a binary relation on U is a *semiorder* if and only if for all $x, y, z, w \in U$:

Axiom 1. Not xRx;
Axiom 2. If xRy and zRw then either xRw or zRy;
Axiom 3. If xRy and yRz then either xRw or wRz.

Theorem 2. Let $\mathscr{X} = \langle X, \mathscr{F}, \mathscr{S}, \geqslant\rangle$ be a finite approximate measurement structure for beliefs. Then

(i) there exists a probability measure P on $\mathscr{S}$ such that for any two standard events S and T

$$S \geqslant T \quad \text{if and only if } P(S) \geqslant P(T),$$

(ii) the measure P is unique and assigns the same positive probability to each minimal event of $\mathscr{S}$,

(iii) if we define P_* and P^* as follows:

(a) for any event A in $\mathscr{F}$ equivalent to some standard event S,

$$P_*(A) = P^*(A) = P(S),$$

(b) for any A in $\mathscr{F}$ not equivalent to some standard event S, but lying in the minimal open interval (S, S') for standard events S and S'

$$P_*(A) = P(S) \quad \text{and} \quad P^*(A) = P(S'),$$

then P_* and P^* satisfy conditions (I)–(III) for upper and lower probabilities on $\mathscr{F}$, and

(c) if n is the number of minimal elements in $\mathscr{S}$ then for every A in $\mathscr{F}$

$$P^*(A) - P_*(A) \leqslant 1/n,$$

(iv) if we define for A and B in $\mathscr{F}$

$$A *> B \quad \text{if and only if } \exists\, S \text{ in } \mathscr{S} \text{ such that } A > S > B,$$

then $*>$ is a semiorder on $\mathscr{F}$, if $A *> B$ then $P_*(A) \geqslant P^*(B)$, and if $P_*(A) \geqslant P^*(B)$ then $A \geqslant B$.

Proof. Parts (i) and (ii) follow from the proof given in Suppes (1969, pp. 7–8) once it is observed that the subalgebra $\mathscr{S}$ is isomorphic to a finite algebra $\mathfrak{A}$ of sets with the minimal events of $\mathscr{S}$ corresponding to unit sets, i.e. atomic events of $\mathfrak{A}$.

As to part (iii), conditions (I) and (II) for upper and lower probabilities are verified immediately. To verify condition (III) it will be sufficient to assume that neither A nor B is equivalent to a standard event, for if either is, the argument given here is simplified, and if both are, (III) follows at once from properties of the standard measure P. So we may assume that A is in a minimal interval (S, S') and B in a minimal interval (T, T'), i.e. $S < A < S'$ and $T < B < T'$. Since by hypothesis of (III), $A \cap B = \emptyset$, $T \leqslant \neg S$ for if $T > \neg S$, we would have $A \cup B > S \cup \neg S$, which is impossible. Now it is easily checked that for standard events if $T \leqslant \neg S$ then $\exists\, T^*$ in $\mathscr{S}$ such that $T^* \approx T$ and $T^* \subseteq \neg S$. So we have

$$P_*(A) + P_*(B) \leqslant P(S) + P(T^*) = P(S \cup T^*) \leqslant P_*(A \cup B),$$

with the last inequality following from $S \cup T^* < A \cup B$, which is itself a direct consequence of $S < A$, $T^* < B$, $A \cap B = \emptyset$ and Axiom 2. For the next step, if $\exists T^{**}$ in $\mathscr{S}$ such that $T^{**} \approx T'$ and $T^{**} \subseteq \neg S'$, then $A \cup B < S' \cup T^{**}$ and let $A \cup B$ be in the minimal closed interval $[V, V']$, i.e. $V \leqslant A \cup B \leqslant V'$. Then it is easy to show that $V \leqslant S \cup T^{**}$, whence

$$P_*(A \cup B) = P(V) \leqslant P(S \cup T^{**}) = P(S) + P(T^{**}) = P_*(A) + P^*(B)$$

and since $S \cup T^* < A \cup B$, and $V \leqslant S \cup T^{**}$, either $A \cup B \leqslant S \cup T^{**}$ or $A \cup B \leqslant S' \cup T^{**}$. In either case

$$\begin{aligned} P_*(A) + P^*(B) &= P(S \cup T^{**}) \leqslant P^*(A \cup B) \leqslant P(S' \cup T^{**}) \\ &= P(S') + P(T^{**}) = P^*(A) + P^*(B). \end{aligned}$$

On the other hand, if there were no T^{**} such that $T^{**} \approx T'$ and $T^{**} \subseteq \neg S'$, then $T' > \neg S'$, so that $S \cup T^* = S \cup \neg S$, and consequently $A \cup B \approx S \cup T^*$, so that $A \geqslant S$ or $B \geqslant T^*$ contrary to hypothesis, which completes the proof of (III).

Proof of (c) of part (iii) follows at once from (ii) and the earlier parts of (iii). Proof of (iv) is also straightforward and will be omitted.

I turn now to some remarks about Theorem 2. The implications stated in part (iv) cannot be strengthened to equivalence. It is easy to give counterexamples to each of the following four equivalences:

$$A *> B \quad \text{if and only if } P_*(A) \geqslant P^*(B);$$

$$A \geqslant B \quad \text{if and only if } P_*(A) \geqslant P^*(B);$$

$$A > B \quad \text{if and only if } P_*(A) \geqslant P^*(B);$$

$$A > B \quad \text{if and only if } P_*(A) > P^*(B).$$

A counterexample to the strong condition (IV) for upper and lower probabilities is the following. Let the outcomes of X be the four possible outcomes of two flips of a coin, the first without bias and the second with some unknown bias favouring heads. Explicitly, let $X = \{hh, ht, th, tt\}$. Then the standard events are X, $\emptyset$, $\{hh, ht\}$ and $\{th, tt\}$, with $P(\{hh, ht\}) = P(\{th, tt\}) = \frac{1}{2}$. Let $A = \{ht, hh\}$ and $B = \{hh, tt\}$. Then it is easy to see that $P_*(A) = P_*(B) = P_*(A \cup B) = \frac{1}{2}$, but $P_*(A \cap B) = 0$, and thus (IV) does not hold.

In my opening remarks I mentioned the embarrassing problem of being asked for the next decimal of a subjective probability. Without claiming to have met all such problems, the results embodied in Theorem 2 show that the axioms of Definition 3 provide a basis for a better answer. If there are n minimal standard events, then the probabilities of the 2^n standard events are known exactly as rational numbers of the form m/n, with $0 \leqslant m \leqslant n$, and further questions about precision are mistaken. The upper and lower probabilities of all other events are defined in terms of the probabilities of the 2^n standard events, and so the upper and lower probabilities are also known exactly as rational numbers of the same form m/n.

One can object to knowing the probabilities of standard events exactly, but this is to raise another problem that I also think can be dealt with in a way that improves on the axioms of Definition 3, but these additional matters will have to be pursued on another occasion.

Finally, I note explicitly that there is no need in Definition 3 to require that the sample space X be finite. The only essential requirement is that the set $\mathscr{S}$ of standard events be finite. The algebra $\mathscr{F}$ could even have a cardinality greater than that of the continuum and thus the order relation $\geqslant$ on $\mathscr{F}$ might not be representable numerically, and yet the upper and lower probabilities for all events in $\mathscr{F}$ would exist and be defined as in the theorem.

4. Comparison with Geometry

I mentioned at the beginning that I wanted to compare the measurement of belief with the kind of classical measurement characteristic of geometry. We are all familiar with what we expect of geometry, namely, that sufficient postulates are laid down to lead to a unique representation of the Euclidean plane or Euclidean space in terms of numerical Cartesian co-ordinates. The theory leads to exact results, and the uniqueness of the measurements, that is, the numbers assigned to points, is determined up to the group of rigid motions. This is a seductive ideal and is often taken as the ideal we should aim at in the case of the measurement of belief.

My point is to express scepticism that this is the correct ideal and to conjecture that the situation is more like the prototypical situation of quantum mechanics. Any

time we measure a microscopic object by using macroscopic apparatus we disturb the state of the microscopic object and, according to the fundamental ideas of quantum mechanics, we cannot hope to improve the situation by using new methods of measurement that will lead to exact results of the classical sort for simultaneously measured conjugate variables. I do not mean to suggest that the exact theoretical ideas of quantum mechanics carry over in any way to the measurement of belief, but I think the general conceptual situation does. In fact, it seems to me that some of the recent empirical work of Tversky and his collaborators shows how sensitive the measurement of belief in the sense of subjective probability is to the particular method of measurement chosen. There is a general way of talking about this situation that is suggestive of a line of investigation, in terms of the theory of the measurement of belief, that has not yet been explored, but that may be promising for the future.

The basic idea is that it is a mistake to think of beliefs as being stored in some fixed and inert form in the memory of a person. When a question is asked about personal beliefs, one constructs a belief coded in a belief statement as a response to the question. As the kind of question varies, the construction varies, and the results vary. What I am saying about the construction of beliefs is similar to a view commonly held about memory, namely, that very little of memory represents an inert encoding. We are primarily constructing detailed memories by procedures that we do not at present understand, but that operate in a more subtle way on encoded data than simply by a direct retrieval of information. As many of you will recognize, such a conception of memory is classical and is especially associated with the early important work on memory by Sir Frederic Bartlett, especially in his book, *Remembering* (1932). A good recent overview of these matters, including an appraisal of the current status of Bartlett's ideas, is found in Cofer (1973).

Let me be clear about the basic point I want to make. After all, constructions are familiar in geometry and lead to exact results. A similar claim might be made about the constructive processes in memory in which we examine past experience in reaching comparative evaluations of belief. My point is, however, that the constructive processes in the case of belief are not of this kind, but are easily disturbed by slight variations in the situation in which the constructive processes are operating. This kind of view backs up the layman's view that it is ridiculous to seek exact measurements of belief; it can also be used to defend the expert opinion that it is unseemly to ask for the next decimal in a measurement of subjective probability. I do not at the present time have any good ideas of how to think about these constructive processes. My conjecture is that this is a move in the right direction, and that in making this move we should try to operate at an abstract level that will lead to specific results in explaining the felt uneasiness of any attempts to seek exact measurements of belief.

My one definite idea about such constructive processes is that mathematical models of learning provide a preliminary, simple schema. Modern rationalists of human thought sometimes seem to think that beliefs are changed simply by the use of Bayes's theorem, or at least in first approximation this is what happens empirically. And, ideally, this is what always should happen in the case of a rational man. There are many ingredients for considering this Bayesian idea a fantasy of reason, however, and I have on a previous occasion tried to state several of them (Suppes, 1966). Let me summarize the matter by saying that in many cases of change of belief it appears obvious that we cannot identify directly or indirectly the evidence on the basis of which the belief is changed, much less the relevant likelihoods or probabilities.

Simple learning models that work in first approximation, both for animals and humans, give some idea of how such constructive processes operate. It seems appropriate to say that the kind of changes that take place in learning can be regarded as examples of changes in belief. Thus if we study as a process of stimulus sampling and conditioning the acquisition of simple mathematical concepts by children, it is correct to say that during the course of learning, their beliefs about the concepts being taught change, as reflected in their responses. In making this remark, I am not suggesting for a moment that changes in belief are always reflected in responses, but rather that this is one way of getting evidence on changes of belief.

It is of course sometimes said that learning theories that postulate learning primarily on the basis of stimulus sampling and conditioning are too passive in nature, and that they do not consider adequately the conscious use of cognitive strategies by learners. I think that on occasion conscious strategies are used but, ordinarily, these strategies are not articulated, and when a learning theory based on stimulus sampling and conditioning is formulated in proper mathematical terms (see, for example, Estes, 1959; Suppes, 1969; Estes and Suppes, 1974), there is no commitment to whether the internal processes are constructive or passive in nature. The level of abstraction in handling the concept of stimulus is such that constructive processes could easily be assumed for handling the conditioning of stimulus patterns or, if you will, in more cognitive terms, the formation and storage of hypotheses.

To illustrate these ideas with a concrete but simple example I draw upon some earlier work reported in Suppes and Ginsberg (1963). The two-element model I consider may be conceptualized as follows. There are two stimulus features or patterns associated with each experimental situation. With equal probability exactly one of the two features is sampled on every trial. Let us call the features or elements σ and τ. When either element is unconditioned there is associated with it a guessing probability g_σ or g_τ, as the case may be, that the correct response will be made when that unconditioned stimulus is sampled. An assumption of particular importance to the present model is that the probability of the sampled stimulus element becoming conditioned is not necessarily the same when both elements are unconditioned as it is when the non-sampled element is already conditioned. We call the first probability a and the second b.

Under these assumptions, together with appropriate general independence of path assumptions as given, for example, in Suppes (1969), the basic learning process may be represented by the following four-state Markov process, where the four states (σ, τ), σ, τ and 0 represent the possible states of conditioning of the two stimulus elements.

	(σ,τ)	σ	τ	0
(σ,τ)	1	0	0	0
σ	$b/2$	$1-b/2$	0	0
τ	$b/2$	0	$1-b/2$	0
0	0	$a/2$	$a/2$	$1-a$

The model just described can be applied with reasonable success to data on children's learning simple mathematical concepts. A typical example would be the

experiment on geometrical forms of Stoll (1962) reported in Suppes and Ginsberg (1963). In this experiment the subjects were kindergarten children who were divided into two equal groups. For both groups the experiment required successive discrimination, with three possible responses permitted. One group discriminated between triangles, quadrilaterals and pentagons, and the other group discriminated between acute, right and obtuse angles. For all subjects a typical case of each was shown immediately above the appropriate response key.

I shall not go into the detailed analysis of data; those interested are referred to the references just given. From the standpoint of concern here, it is easy to see why passive stimulus sampling seems absurd. The stimulus displays varied from trial to trial. For example, the same acute angle was not displayed on each trial. Obviously the subjects had to go through the constructive process of approximately matching the salient features of the display (conceptualized by the model to be two in number) to obtain a decision on their presence or absence. It is certainly true that the kind of theory I have described does not provide adequate details of this processing. It does provide a coarse analysis that fits data remarkably well. This is a simple example, but hopefully illustrates my point. The subjects were too young to verbalize in any precise way what they had learned, but they were able to learn the constructive processes of identification, and their beliefs and knowledge were changed in the process. There is a good deal more I would like to say about how these internal constructive processes operate. Conceptually I currently think of them in terms of computer programs written in terms of a simple set of instructions involving perceptual as well as internal processes. The features σ and τ in the simple model described above are each represented internally by two elementary programs. In a recent publication I have tried to spell out this approach to learning for the case of children's acquisition of the standard algorithm of numerical addition, but it is not possible to enter into detail here (Suppes, 1973).

Final remark

When one examines the status of learning theory in relation to complex concepts, or the analysis from any other standpoint, including contemporary cognitive psychology, of the acquisition and holding of beliefs, it seems appropriate to be sceptical of our ever achieving a complete theory of such matters. The information we can obtain about an individual's beliefs will, in my judgment, always be schematic and partial in character. Even if the time comes when we shall be able to have what we feel is an adequate fundamental schema of the processes involved, it is doubtful that we shall be able to implement a complete quantitative study of an individual's beliefs.

To accept the necessary incompleteness of what we can analyse is, to my mind, no different from accepting the impossibility of complete meteorological predictions. It is hopeless and, probably in one sense, uninteresting to attempt to measure and predict exactly the motion of the leaves on a tree as a breeze goes by. Our beliefs, it seems to me, are rather like the leaves on a tree. They tremble and move under even a minor current of information. Surely we shall never predict in detail all of their subtle and evanescent changes.

ACKNOWLEDGEMENTS

Research in connection with this article has been partially supported by the U.S. National Science Foundation under Grant NSFGJ-443X. I am indebted to

Mario Zanotti for a number of useful comments on the ideas developed here, especially those dealing with upper and lower probability. Referees' comments on the first draft and suggestions for improvements of the exposition were also helpful.

References

BARTLETT, F. C. (1932). *Remembering*. New York: Macmillan.

COFER, C. N. (1973). Constructive processes in memory. *Amer. Scientist*, **61**, 537–543.

DE FINETTI, B. (1931). Sul significato della probabilità. *Fund. Math.*, **17**, 298–329.

—— (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. Poincaré*, **7**, 1–68. 1964. English translation in *Studies in Subjective Probability* (H. E. Kyburg, Jr., and H. E. Smokler, eds). New York: Wiley.

DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, **38**, 325–340.

Estes, W. K. (1959). Component and pattern models with Markovian interpretations. In *Studies in Mathematical Learning Theory* (R. R. Bush and W. K. Estes, eds). Stanford: Stanford University Press.

ESTES, W. K. and SUPPES, P. (1974). Foundations of stimulus sampling theory. In *Contemporary Developments in Mathematical Psychology* (D. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes, eds). San Francisco: Freeman.

GOOD, I. J. (1962). Subjective probability as the measure of a non-measurable set. In *Logic, Methodology and Philosophy of Science: Proceedings of the* 1960 *International Congress* (E. Nagel, P. Suppes and A. Tarski, eds). Stanford: Stanford University Press.

KOOPMAN, B. O. (1940a). The axioms and algebra of intuitive probability. *Ann. Math.*, **41**, 269–292.

—— (1940b). The bases of probability. *Bull. Amer. Math. Soc.*, **46**, 763–774.

KRAFT, C. H., PRATT, J. W. and SEIDENBERG, A. (1959). Intuitive probability on finite sets. *Ann. Math. Statist.*, **30**, 408–419.

RAMSEY, F. P. (1931). *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul.

SAVAGE, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

SCOTT, D. (1964). Measurement structures and linear inequalities. *J. Math. Psychol.*, **1**, 233–247.

SCOTT, D. and SUPPES, P. (1958). Foundational aspects of theories of measurement. *J. Symbol. Logic*, **23**, 113–128.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision. *J. R. Statist. Soc.* B, **23**, 1–25.

STOLL, E. A. (1962). Geometrical concept formation in kindergarten children. Ph.D. Thesis, Stanford University.

SUPPES, P. (1956). The role of subjective probability and utility in decision-making. In *Proc. 3rd Berkeley Symp. Math. Statist. and Prob.*, 1954–1955 (J. Neyman, ed.), Vol. 5. Berkeley: University of California Press.

—— Probabilistic inference and the concept of total evidence. In *Aspects of Inductive Logic* (J. Hintikka and P. Suppes, eds). Amsterdam: North-Holland.

—— Stimulus-response theory of finite automata. *J. Math. Psychol.*, **6**, 327–355.

—— Facts and fantasies of education. In *Changing Education: Alternatives from Educational Research* (M. C. Wittrock, ed.). Englewood Cliffs, N.J.: Prentice-Hall.

SUPPES, P. and GINSBERG, R. (1963). A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychol. Rev.*, **70**, 139–161.

Discussion on the Papers by Professor Tversky and by Professor Suppes

Professor CEDRIC A. B. SMITH (Galton Laboratory): Statisticians and psychologists have come together here to learn from one another. Both are interested in "probability", so one would think they would have much in common. However, many different ideas can be expressed by a single word. It seems worth while considering to what extent their ideas and motives do in fact run parallel, and how much they diverge.

A statistician's aim, as statisticians work at present, is to obtain plausible conclusions from data subject to random fluctuations. He necessarily has to study probability, since

conclusions from such data can never be certain. Now when we speak of probability we usually mean an attitude of mind. Peter looks at the sky and thinks it likely that it will rain; Paul may expect a sunny day. Such attitudes are personal ones, differing from one man to another. To a psychologist, it is of interest to study how such attitudes come about, and how strong they are, and why different people think differently. But to a statistician, such variability from person to person is a nuisance. He wants his conclusions to be generally acceptable; if Peter agrees with him, but Paul does not, this reflects badly both on his usefulness and his credibility. Thus, in a way, the motives of psychologists and statisticians are opposed. But, possibly because of this, interchange of thoughts between them is quite desirable and possibly productive. We can learn a lot from people who think in different terms.

One point is obvious; the importance in certain circumstances of judging probability correctly. When we cross a road, we must have in mind some idea of the chance of being hit by the next car coming along. If we misjudge this chance, the consequences can be fatal. Not all judgments are so crucial, but mistakes can lead to inconvenience and lost opportunities. Now at least some probabilities are ideally exactly (or nearly enough so) measurable; there is a 50 per cent chance of a coin falling heads, a 25 per cent chance of picking a heart at random from a pack of cards. The experiments of Tversky and others have shown that even such exact probabilities are not always judged correctly. This is unfortunate, but perhaps not too surprising. We all agree that a house or a telegraph pole has a precise measurable height. But we may not be very expert at judging it simply by looking at the house. It is much more accurate to use a measuring tape and a theodolite. The human brain is not intended as a refined measuring tape, nor does it seem be be best viewed as a mathematician performing the operations of the calculus of probabilities. What is rather more alarming is that Tversky's experiments seem to indicate a difficulty in combining prior probability and experimental information. If this is true, it is a serious defect. However, it may be that the kind of experimental situation psychologists use is somewhat artificial, in the sense that it does not closely relate to the kind of situation which would be of importance to primitive man in the course of evolution, and where accurate judgment would have been vital.

With chances which are not susceptible of exact measurement, the position is rather different. The statistician will be interested in trying to make a "best possible" estimate of probability, a "best possible" decision. Here he is aided by the theory due to Ramsey, de Finetti and Savage, which shows essentially that a self-consistent decision-maker would necessarily have a numerical scale of probabilities and preferences. Professor Suppes seems to object to this on the grounds that the procedure used by Savage for defining such a scale is unnatural. However, the absolute scale of temperature is defined in principle in terms of reversible heat-engines. This does not mean that in practice one carries such a heat-engine around everywhere to measure temperature. An ordinary thermometer will do, but only because it has been justified in principle by theoretical arguments using heat-engines. There seems in the normal course of events no particular advantage in being inconsistent, so as a statistician one may look on Savage's self-consistent decision-maker as an ideal to be (usually) aimed at; it is a guide to thought and action. However, it is less obvious that the attitudes of real individuals can be put on a numerical scale at all. Professor Tversky asks his subjects about "90 per cent probabilities"; how, I wonder, do they understand this phrase? To what extent does it represent a clear and well-defined idea? We tend to think of all well-ordered science as dealing with numerical-valued quantities. But I believe that such a simple quantity as "hardness" is still defined as an order relation; diamond will scratch glass, glass will scratch wood, wood will scratch jelly. Probability as an attitude is less well defined and certainly less easily measurable than hardness. Does it even necessarily have an ordering? It is plausible that if A is judged more probable than B, and B more probable than C, then A will be judged more probable than C. But is there any kind of rigorous justification for this plausibility? Is it always true? What does it mean, anyway, to say that "A is more probable than B"?

Both Professor Tversky and Professor Suppes are guilty of trying to get precise results from undefined terms. Professor Tversky asks subjects when their "subjective probability is 0·9", without apparently explaining what that might mean; Professor Suppes looks for a "rational" scheme of axioms, without defining rationality. Unfortunately, definitions are exceedingly elusive. Efforts based on such a yielding foundation must be regarded as heroic, but suspect. Let us respect the heroism.

I have much pleasure in proposing a vote of thanks.

Professor R. J. AUDLEY (University College London): The two papers we have listened to this afternoon complement one another very nicely. Professor Suppes has examined some influential views concerning the development of a normative theory of decision-making in uncertain situations. By critically examining the axioms of theories of this kind he is led to the view that for some circumstances at least they ascribe characteristics to the ideal decision-maker that not only go beyond what we can expect of real men but also go beyond what it is proper to ask even of a perfect problem-solver. In discussing this, he has offered ways of being less demanding on a decision-maker and has suggested, if I understand him correctly, that there is a need, even in developing normative theories, to find some way of taking into account the nature of the constructive processes which underlie the belief statements produced by real men.

Professor Tversky, on the other hand, has reported some results from a very ingenious programme of research into the way in which real men handle the concepts associated with risk and uncertainty. He in turn is led to join Suppes in questioning whether existing normative theories have really captured the essence of stating what is the meaning of the term "rational" in the context of decisions; and he too is led to raise the matter of the constructive processes underlying belief.

Even by now it may have become apparent that I have little knowledge of the problems of axiomatizing rational decision-making in situations of uncertainty. The decent thing for me to do, as a psychologist, should therefore be to confine my comments to Tversky's empirical investigations. However, it seems to me that the important feature of this afternoon's session has been to point out what to a layman appear to be possible weaknesses in existing formulations of normative theories of decision-making. I therefore rashly intend to raise some general, and possibly naive, questions about the way in which research on ideal men and real men can or should be brought together, with the hope of stimulating some further discussion of these issues by the two speakers and others.

As I see it, there are before us three main enterprises: first, developing satisfactory normative theories; second, developing satisfactory descriptive theories of human decision-making—it is about these two tasks that we have heard most this afternoon—and, third, the essentially educational problem of making the decision-making of men conform more closely to acceptable normative principles. This is implicit in Tversky's paper and is touched on in his discussion. However, neither speaker has dealt with it very explicitly, and yet it may be in this third, educational enterprise that the clues to bringing together the other two may be found.

It may or may not be true that one can develop a normative theory without regard to man's own capacities as a decision-maker. It is not clear to me, for example, that Suppes's criticisms of existing normative theories extend to the possibility of their principles being embodied in an ideal decision-maker, such as might be partly represented in a computer program. However, even if the principles of an ideal system may be developed without regard to the properties of real men, it is likely on the basis of what has been said today that the existing normative formulations do not lend themselves to easy assimilation by real men. If there is any chance of educating men to rationality, then we have to take account of their present capacities, limitations and intuitions. For this reason, therefore, it might be desirable for the time being to have the normative theories couched in terms which provide contacts with the properties of real men. It may, at first sight, be reasonable to insist that the principle effort should go into educating people to be good Bayesians, but

it would also seem to follow from this position that one should take into account the prior condition of those one wishes to educate.

The concept of subjective probability may be a key one in this respect. Savage, as I take it, has a preference for indirect ways of assessing this—that is via choices between alternative courses of actions—which might also be called a behavioural definition of subjective probability. This preference for the behavioural approach is largely on the grounds that direct estimates of the probabilities may not be properly employed by the individual when he comes to make a decision. This is reminiscent of the problem of measuring attitudes, when verbal statements by an individual may be contradicted by his actions. Suppes appears to go along with this in some way since his example of concept learning in children is essentially a behavioural analysis. What the children are said to know is based on their responses in the particular experiment. On this issue Tversky has raised the important point that one is unlikely to judge a subjective probability on the basis of observing one's own actions—although an equivalent viewpoint has sometimes been advanced in the study of attitudes—and I take it that he favours a cognitive interpretation of subjective probability. But he has also drawn attention to some of the difficulties that exist for men in so storing information about their experience that they can appreciate the statistical properties of the world and so come up with sound probability estimates. But it seems likely that stating principles of decision-making in terms of behavioural probabilities is going to offer even fewer possibilities for the individual to improve his comprehension or use of statistical material. In fact, there remain some fascinating psychological problems to be solved concerning the relation between subjective probabilities (or expectations) as these are reflected in behaviour—e.g. as measured in terms of the speed with which responses are made to different events—and conscious estimates of subjective probabilities. It is not a foregone conclusion that the conscious estimates are more accurate than expectancies which appear to underlie behaviour.

For these reasons there would seem to be an advantage in encouraging real men to make explicit judgments of the likelihood of events. The problem of educating for rationality then becomes one of encouraging them to make their decisions and their estimates of likelihood compatible; or, if you like, to make their behavioural and cognitive subjective probabilities converge. This is rather in the spirit of attempting to combat prejudice by making individuals aware of the contradictions between behaviour and expressed attitudes which are constrained by more reflective consideration of the common lot of men. Incidentally, I am not sure that Tversky is right to propose that we should ask a man to make his beliefs and judgments consistent with all he knows. Elsewhere, he has pointed out the dangers of using consistency of this kind as a basis for making judgments. Humans are only too willing to explain events so that they appear to be consistent with other particular events. In statistical matters the larger class of possibilities, including events that might have occurred, needs also to be taken into account. As Tversky has shown us today, many, if not all, of us have difficulties in incorporating valid statistical concepts in our thought. In conclusion, therefore, I would like to suggest that the indications are that, whilst we are still trying to educate ourselves as rational men, the emphasis of both the normative and the descriptive theorists should be upon the cognitive rather than the behavioural definition of subjective probability. I should add that I agree with Suppes that the evidence about cognition will mostly come from behavioural data; but as in the case of Tversky's experiments the questions asked of the subject should be aimed at exposing the processes underlying their judgments of likelihoods. There is of course already a large body of research on this topic, to which both of the speakers and Ward Edwards and his colleagues have made important contributions.

It can be gathered from what I have said that I am very unclear as to the appropriate terms in which a normative theory might eventually be couched and the relationship between a normative theory and a descriptive theory, and I would appreciate being enlightened on this point. For the time being, however, I believe myself to be supporting the kinds of links between the decisions of real and ideal men that Suppes and Tversky

have implied in their papers, and in my view these links are to be justified and looked for on educational grounds.

Most of my remarks are born of ignorance and I was delighted that Professor Tversky demonstrated that a non-informative statement may have the effect of reducing the effect of prior probabilities. At the very least, I therefore hope that I have cleared the debate of prejudice.

I take great pleasure in seconding the vote of thanks to the speakers.

The vote of thanks was carried by acclamation.

Professor J. M. DICKEY (University College London): An important use of "standard" theories is in the clarifying of muddled personal uncertainties by thinking about potential actions. The "structural" axioms serve as tools to help clarify uncertainty as well as to measure it. I am sorry that Professor Suppes finds it "very unlikely and psychologically very unrealistic to believe that thinking about finite sequences of flips of a fair coin will be of any help in making a rational decision on the part of the patient". It is an unfortunate disadvantage that men do not easily think in quantitative terms, that a patient and physician do not usually see how their subjective beliefs are motivated by fear and by wishful thinking, that they may not have awareness of the ways their preferences depend on accidental or "structural" features of the decision-problem environment. If they would obtain such awareness on each occasion by the *conceptualization* of various environments differing by patently inessential structures, would rational men not spontaneously lose some of their dependence on such differences? Would they not tend to form beliefs which are applicable to potential actions in the whole variety of imagined choice situations—beliefs which are hence worthy of being called beliefs. Does man not have the ability to imagine alternative situations differing by purely "structural" features, and thereby to clarify his thinking, or must he settle for "special axioms of rationality for special situations"? What about the goals of *know thyself* and *wholeness of mind*?

Surely the first author realizes that in decision practices it is not necessary to determine uncertainties to an infinite number of decimal digits. Much wisdom follows from mere orders of magnitude.

But assessed personal probabilities need to refer to clarified personal uncertainty in a real-world problem. One sees in the second author's paper that the presence of an imagined teaching parent leads a would-be probability assessor to deliver an immediate logical probability value to satisfy the imagined teacher's immediate point.

The "anchor" effect implies the need to assess in several ways from many starting points in order to gain sufficient self-awareness for the clarification of one's uncertainties.

I end my discussion with two technical points. If people do regard the sequence HTHTTH to be more likely than the sequence HHHTTT for a fair coin, could this be merely a confused carry-over of their more important regard that the fair sampling model itself is more likely posterior to the former data than posterior to the latter data?

My second point is that weather-probability assessors have made use of the Brier and Allen (1951) score,

$$\text{SCORE} = n^{-1}\sum_{i=1}^{n}(p_i - d_i)^2,$$

where $d_i \in \{1, 0\}$ indicates the occurrence or not of an event predicted with probability p_i on instance i. If the range (0, 1) for p_i is partitioned into subintervals I, one will have a decomposition of the sum of squares,

$$\begin{aligned}\text{SCORE} &= n^{-1}\sum_{I}\sum_{p_i \in I}(p_i - d_i)^2 \\ &\doteq \sum_{I}(n_I/n)\{(\bar{p}_I - s_I/n_I)^2 + (s_I/n_I)(1 - s_I/n_I)\},\end{aligned}$$

where

$$n_I = \sum_{p_i \in I} 1, \quad s_I = \sum_{p_i \in I} d_i,$$

and where one approximates each $p_i \in I$ by some value $\bar{p}_I$ (Sanders, 1958; Dickey and Walrath, 1971). The first component of the sum of squares measures the "calibration" referred to in the second author's paper. The second component measures the tendency toward "all occur" or "none occur" among the instances with assessed probabilities in each subinterval. (De Finetti is said to have used SCORE much earlier.)

Dr A. D. LOVIE (University of Liverpool): My major comment concerns the relationship between psychology and the normative systems of statistics and decision-making. We know from what Professor Tversky has said and also from other published work that people are rather poor at mimicking the prescriptions of either traditional or Bayesian statistics. I want to give you three further examples of these: if subjects are asked to say which of two sequences of numbers has the highest mean, then the range and values of the numbers have effects which are not meaningful in statistical terms (Salway, 1971). Secondly, when subjects are asked to give estimates of subjective probability, then the scale values of the units in which the subjects express their estimates again affect these estimates, and again this is not in any way predictable from classical probability theory (Salway, 1971). (In this connection it is interesting to note that Professor Tversky gave probabilities and other numbers which seemed to involve fives and tens, which is a pretty well-known phenomenon in this kind of work.) Finally, when subjects are asked to provide estimates of the variance of a sample drawn from a normally distributed population, then the estimate of that variance is dependent on the value of the mean of the sample, unlike the parent population, where the two parameters are independent (Beach and Scopp, 1968). And of course there is a vast amount of Bayesian work which has demonstrated how bad most untrained people are at updating hypotheses in the light of new evidence (Slovic and Lichtenstein, 1971). I suggest, therefore, that people are not merely bad at playing the role of intuitive statisticians and decision-makers but their behaviour is very nearly orthogonal to that suggested by statistical theory and practice. And yet it seems to me that it is absolutely vital for human beings to be able to participate fully in the running of projects and systems that are most suitably and economically characterized in terms of statistical and decision-making procedures.

If we can assume, on the one hand, that a subject's behaviour is somewhat orthogonal to the suggested prescriptions of statistics and decision-making and yet, on the other, that it is necessary that people should follow these prescriptions in many situations, then I have two suggestions to make. One is fairly general and pretty obvious; it is that psychologists should still pursue the study of people in numerical environments but that their reference to normative structures should be a little more circumspect; it would be helpful, for example, for such structures and processes to be recast in cognitive terms through a kind of task-analysis of their functional requirements. The second is a bit more radical; it is that the study of intuitive statistics and decision-making be taken from the grasp of the academic, laboratory-oriented psychologist and given to the human factors specialist, partly because he is more interested than most psychologists in seeing that man–system interfaces operate properly, but also—and I think this is more important—because such a specialist is much more situation-centred than academic psychologists. This means that he may then take into account the unique and specific aspects of a particular situation—something which is rarely pursued in academic decision-making studies. One other advantage of such a switch would be that he might also search for explanations in psychological terms, again something often discussed (cf. Slovic and Lichtenstein, 1968) but rarely carried out.

My personal line of attack on decision-making and intuitive statistics is to think of people as constrained decision-makers. This is not a very original idea; it is one that the

Japanese psychologist Toda has been pushing for a long time, and also Rapoport, who has applied such ideas to multi-stage decision-making studies (cf. Toda, 1962; Rapoport, 1973; Shuford, 1964). This approach assumes that the constraints are cognitive—that is, processing ones, memorial ones, perceptual ones, etc.—and then, for methodological reasons, selectively and appropriately aids the decision-maker. This leads me to suggest that the adoption by people of the three heuristics suggested by Professor Tversky is the result of faulty and inadequate cognitive processing mechanisms which force people to degrade or re-jig the environment in such a way that it may be more readily handled by such inadequate information-processing mechanisms.

I have one brief comment on Professor Suppes's paper: the measurement of inexact beliefs that he presents might perhaps provide a representation of the noise or fuzz in a subject's system of beliefs which is, we know, present over and above the effects of relatively stable factors such as probability of occurrence of the events. This might then provide the inferential basis for testing the various composition rules in polynomial conjoint measurement models of belief in a way analogous to the role of the error term in the linear models of analysis of variance (cf. Krantz *et al.*, 1971).

Dr L. D. Phillips (Brunel University): I do not have a lecture to give, only a question to ask. Since the successful application of decision theory often depends on the decision analyst being able to extract from his client a subjective probability distribution that is reasonably unbiased, could Professor Tversky say more about his suggestion that procedures could be developed to minimize bias, other than the method he talked about?

Professor D. V. Lindley (University College London): It is a pleasure to be at a meeting of our Society devoted entirely to problems of subjective probability, especially when the two speakers handle them so well. I do think that it is of the utmost importance that we discuss the Savage axioms, and their meaning for persons trying to assess uncertainty, because the whole theory of Bayesian statistics and decision theory rests on them.

One point about the axioms that the speakers do not bring out is that they refer to a large world, in Savage's language. The coherent individual is supposed to assess his probabilities and utilities for everything. Of course, taken literally, this is absurd; but it does not invalidate the theory any more than the failure of the claim to predict the whole future of the universe, given the position and velocities of particles now, invalidates Newton's theory. We deal with approximations, with small worlds. There are two lessons we can learn from this.

Firstly, my large world, and I am sure yours too, embraces the concept of a random variable uniformly distributed in a unit square. With no more difficulty than that of making a standard of measurement, such a standard of uncertainty can be constructed. Its existence will do for Savage's Axiom 6 (as described by Professor Suppes) and many of the difficulties he mentions disappear. Contrary to what he says, I have found this idea useful in managerial problems in medicine.

Secondly, we should learn not to take our worlds too small. Professor Tversky often appears to do this when he asks for a few isolated uncertainty assessments. It is good advice, when thinking about one probability that is needed, to introduce others, assess these as well, and then see if the judgments cohere. Typically they will not, and revision to make them so is called for. Since coherence is the basic idea, one would expect it to play an important role in decision-making: it is not always given an opportunity to do so.

There is one comment that I would like to address to psychologists present. Why do you spend your time studying how people make decisions, when we know how they *should* make decisions? Would it not be better to devote your energies to teaching them the principles of maximum expected utility? You are like a lot of people who ignore arithmetic and go around seeing what subjects think 2×2 is. Having got an average of 4·17 you solemnly announce a learning model with $2 \times 2 = 4{\cdot}17$. A bit more drilling in arithmetic in the *laissez-faire* schoolrooms of today would not come amiss.

An example is provided by the one quoted by Professor Tversky on p. 154. Let R_1 be the event of a red ball from the 50–50 urn, and R_2 the same for the urn of unknown constitution. With $\bar{R}_i = G_i$ there are four events $R_1 R_2$, $R_1 G_2$, $G_1 R_2$, $G_1 G_2$ and let $(c, c, 0, 0)$ etc. denote a gamble that yields a prize c if either of the first two events occurs. If $(c, 0, c, 0) \sim (0, c, 0, c)$ (for he might think that a red ball was more likely than green from the second urn, the concept of "unknown" being ill defined), we have

$$(c, c, 0, 0) \sim (0\,0, c, c) > (c, 0, c, 0) \sim (0, c, 0, c).$$

But, by the sure-thing principle, from the first and third $(c, 0) > (0, c)$ whereas from the second and fourth $(0, c) > (c, 0)$ both pairs referring to $R_1 G_2$ and $G_1 R_2$ only. This contradiction exposes the fallacy, and if the subject does not admit it he could just as well give me money. Of course, the discoveries of Tversky could enormously help this instruction, but instruction, and not model-building, should be the psychologist's primary aim.

There is one remark in the paper which might have been put in deliberately in order to give a Bayesian a debating point. On p. 152 he says: "the tendency to rely on the individuating information, with insufficient regard for prior probability, has been observed in . . . individuals who had extensive training in statistics". Of course, most statisticians have been trained to do just this in the name of objectivity—would you expect otherwise? I repeat, teach them proper statistics.

Professor V. R. Cane (University of Manchester): I was most interested to hear the discussion of the Bayesian approach given by Professor Suppes, in terms of axiomatic requirements, and by Professor Tversky in terms of experimental verification. I should like to discuss the use of Bayes's theorem from the point of view of model-building.

My aim is to start with a crude model which might represent the process of animal learning or of naive inference and to refine it by introducing rules which a person showing rational behaviour must obey, in the hope that this will lead to Bayes's theorem. I assume that a subject is shown the outcome of a series of Bernoulli trials—for example, he might observe a succession of draws from an urn, with replacement—and is asked to estimate the probability of success. Suppose that he can entertain some set of hypotheses, which may be discrete, about the value of this probability; call a typical hypothesis p. It is assumed that he attaches an initial weight $A(p)$ to this hypothesis and that, as the experiment proceeds, he adds further weights $b_n(p)$ $(w_n(p))$ if a success (failure) occurs at trial n $(n = 1, 2, 3, \ldots)$; if at any stage he is asked to estimate the probability of success he will either give that value of p which has the greatest weight or, if there is a threshold to discrimination, this value of p together with those others whose weights are not noticeably different from the greatest weight.

This is the crude model; we refine it by assuming that p can take any value in the range $0 \leqslant p \leqslant 1$ and adding the following rules:

(1) (independence) omit n, i.e. put $b_n(p) = b(p)$, $w_n(p) = w(p)$;
(2) (symmetry) put $w(p) = b(1-p)$;
(3) (conclusiveness of negative data) put $b(0) = -\infty$;
(4) (inconclusiveness of positive data) take $b(p)$ as bounded, for $0 < p \leqslant 1$;
(5) (consistency) make the maximum weight occur at the true value p_0 as the number of trials, N, tends to infinity.

By the weak law of large numbers, (5) implies that $p_0\, b(p) + (1-p_0)\, b(1-p)$ must be maximum for $p = p_0$, and this must be true for all p_0. Hence we must have

$$pb'(p) = (1-p)\, b'(1-p).$$

Using (3), (4) and the condition for a maximum value we have that any solution of the form $pb'(p) = \sum_0^\infty A_r\{p(1-p)\}^r$ is acceptable provided $A_0 > 0$ and the power series is positive in $0 \leqslant p \leqslant 1$. Thus $b(p)$ has the form

$$A_0 \log p - A_1(1-p)^2 + \ldots.$$

The logarithmic form of Bayes's theorem requires the solution $b(p) = \log p$ and this is said to correspond to the behaviour of the rational man. However, of the rules proposed, the first four cover all that can be logically derived from the experimental situation and the fifth provides a sort of natural selection principle—the survival of those who guess right—so that we can say that the use of any suitable function $b(p)$ denotes a sensible man even if not a rational one. We may then ask what additional rules are required to make sensible behaviour rational.

Professor A. BIRNBAUM (New York University): My comments are not directly relevant to the main problems discussed explicitly in the very interesting papers presented tonight. But they are relevant to important applications often regarded as within the scope of decision theories. They are also relevant to some comments of previous discussants, including Professor Lindley's challenge to the psychologists present, to teach people how they *should* make decisions, instead of just studying how people do make decisions. Implicit in that challenge is another: Why don't you psychologists teach each other, and yourselves, how you "should" analyse your statistical research data (that is, by use of Bayesian methods instead of standard methods)?

My comments concern typical "standard" statistical thinking and practice concerned with scientific research data. The Neyman–Pearson theory is generally regarded as underlying most such applications (including least-squares and maximum-likelihood methods when used to determine confidence regions and tests). Now the Neyman–Pearson theory (at least its mathematical part) is a special case of Wald's non-Bayesian decision theory; and (if we follow Savage's principal theoretical argument) Wald's theory is inadequate except as it can be assimilated into Bayesian theory. It seems to follow that reasonable statistical thinking and practice must be not only decision-theoretic but also Bayesian, instead of non-Bayesian as is currently the case predominantly in all the sciences.

This may be regarded as an interesting case of cultural or scientific lag among scientists including psychologists. It might be interesting to study this lag from the standpoints of the history and sociology of science. For example one might anticipate and prepare to observe systematically a general shift from a non-Bayesian to a Bayesian model for data-analysis among scientists and applied statisticians.

A different analysis of the general nature of current "standard" statistical thought and practice (including incidentally my own) seems more accurate. We can distinguish between a *literal* interpretation of the decision concept and an *elliptical* (or indirect, or abstract) interpretation. Such a distinction can be traced in the writings of Neyman and Pearson, and throughout the standard theoretical, applied, and expository literature. The distinction is often only implicit, but it is fundamental.

In the case of two simple hypotheses we have the familiar schema shown below.

Hypotheses	H_1	H_2
Possible decisions or conclusions	Reject H_1	Reject H_2
Error probabilities	$\alpha = \text{prob}\,[\text{reject } H_1 \mid H_1]$	$\beta = \text{prob}\,[\text{reject } H_2 \mid H_2]$

The Neyman–Pearson theory is characterized *mathematically* by (a) its explicit formulation of problems in such schemas and (b) its systematic treatment of problems of joint minimization of such error probabilities (e.g. in the sense of admissibility).

The *extra-mathematical* interpretation of the decision "reject H_2" is *literal*, for example, in the paradigmatic application of decision theory which appears in the writings of Neyman–Pearson, Wald and others: When a manufacturer places a batch of electric lamps on the market after testing a random sample, no significant questions can be raised

concerning *whether*, and *in what sense*, what he did should be called a decision and an action. In particular, to indicate the *meaning* of the decision "batch placed on market" (which can correspond to "H_2 rejected"), it is simply irrelevant and wrong to refer to the schema above or its elements; it is only the concrete literal meaning of "batch placed on market" which applies.

The elliptical interpretation is represented when a geneticist publishes his judgment that two genetic factors seem to lie on the same chromosome, with his evidence expressed in terms of a standard method derived in the Neyman–Pearson theory, with each important kind of possible misjudgment having a probability less than 5 per cent. Even if the geneticist uses a standard terminology such as "reject H_2, the hypothesis of no linkage", neither he nor his readers understand that he is making a decision, or even reaching a conclusion, in any literal and unqualified sense. Rather, the decision-like term "reject" expresses here an interpretation of the statistical-experimental evidence, as giving appreciable but limited support to one of the alternative statistical hypotheses. This interpretation of the experimental results refers in principle *not* just to the single term "reject H_2" in a schema like that above, but to the *complete* schema, even when this is only implicit.

It can be shown that a simple basic premise of Wald's theory, and of Savage's argument, is satisfied by the literal interpretation but not by the elliptical one. And it can be argued that only the elliptical interpretation cogently relates the decision concept to typical standard practice. This argument can be illustrated by simple formal examples and by examples drawn from genetic research. This argument implies that Savage's argument for Bayesian theory has no *direct* relevance for typical "standard" statistical thinking. (No claim is implied here that typical standard practice can be given a precise general theoretical justification.)

A paper giving details will be submitted shortly to this Society for possible discussion and publication.

Professor MAX HAMILTON (University of Leeds): I would like to make a couple of comments on Professor Tversky's paper.

As he states his first question, the probability that an object A belongs to class B, is very similar to the problem of making a diagnosis. He emphasizes that prior probabilities or base rates do not affect judgments of subjective probability. This may be true with relatively naive subjects, but it can certainly be influenced by training. There has been work done in Leeds by De Dombal and his colleagues on the problems of computer diagnosis of acute disease occurring in the abdominal cavity, and they have shown that errors of diagnosis (which have been worked out on a Bayesian model by the computer) are related to ignorance of prior probabilities, but can be corrected by giving the necessary information to the surgeon. So it is possible to do this.

His point that even trained statisticians do not take sufficiently into account the problem of prior probabilities simply illustrates the general problem concerning the transfer of training. Statisticians are trained to do statistics, not to make subjective judgments.

As for the question of the method of judgment by adjustment, presumably this is determined by the information available to the judge, the person making the decision, but this is subject to certain limitations. Again the work of De Dombal has shown that the accuracy of diagnosis improves with increasing information—that is to say, answers to questions. But he has shown that beyond about ten questions as a maximum the accuracy falls owing to the limits on the utilization of information.

I think these two points must be taken into account.

The following contributions were received in writing, after the meeting:

Dr H. G. BEVANS (St James's Hospital, University of Leeds): I wish here to raise, as with Professor Tversky's paper, certain objections to the narrow use of the word "belief"

for some expectation about material, objective, events. In common language belief means much more: If one asks an acquaintance what he believes, then—if he overcomes his surprise and warranted suspicions—he is likely not to produce a series of assertions about material events, but statements about his own characteristic ways of reacting to experience. It seems to me that the concept of subjective probability cannot logically be applied in the same way to beliefs in this wider sense. We may however ask what pressure of challenging facts will result in the abandonment of a belief, and here we should find developments in Personal Construct Theory useful.

Regarding the development of axiomatics for lower and upper probabilities I wonder if this is really only a half-way stage towards the second-order probability discussed in Tversky's paper, i.e. allocating a probability distribution over the whole range in place of a specific value. This probability distribution thereby recognizes that there may be no such entity as a fixed value within the distribution, only changes brought about by changing circumstances.

The learning model represented by the four-state Markov process I found difficult to follow on account of the brevity of exposition. As I think Professor Suppes would agree, statistical models may provide clues to learning processes but those processes are not likely to be explicable in statistical terms alone.

I appreciated Tversky's review of bias in the assessment of subjective probabilities and the interesting experiments he describes that reveal systematic sources of such bias. In my experience his adverse findings over the statistical intuitions of research psychologists are supported. The problem very often—certainly in clinical psychology—is that a major effort may be required to amass a sample even as low as thirty, involving research time spread over a year.

On another point Tversky says of judgments that they "... must be compatible with the entire web of beliefs held by the individual and not only consistent among themselves." This may not seem restrictive if one regards *values* as a form of belief, but are we correct in viewing them thus? An individual's set of values mean (for me, at least) a set of predispositions to act, to adopt behavioural strategies, mental and perceptual sets, which will affect the probability of his experiencing particular kinds of event. Values form a web of associations with beliefs and any internal inconsistency will tend to generate mental stress. Thus there should be scope for applying concepts of subjective probability not to beliefs about natural, objective, events, but to attitudinal and behavioural sets. With this application the individual is more recognizably someone trying to cope in a changing world, and less like an intellectual spider waiting for a trembling thread to signal incoming events. Then the important question becomes:

What change in external circumstances will induce the individual to change his strategy, his preference for attending to one kind of event rather than another?

I would be interested to hear how Tversky reacts to these observations, and how he would propose pursuing this alternative area of applying the concept of subjective probability.

Mr. D. W. Bunn (The London Graduate School of Business Studies): At a time when the Bayesian decision theoretic approach is developing considerable momentum, it is particularly welcome that both Professors Suppes and Tversky are, in their different ways, expressing scepticism on the validity of Savage's interpretation of subjective probability. Savage effectively derives subjective probability as a parameter of an individual's outward choice behaviour and imputes this, through structural axioms, to be a measure of the individual's degree of belief in the associated future contingent propositions. There is no reason to consider such an indirect evaluation of belief to be equivalent to one of the more direct methods of assessing uncertainty such as the use of standard devices, psychometric ranking or hypothetical samples. Indeed, insofar as they are based upon different sets of axioms, they must be considered logically inequivalent.

Of course, Savage's approach provides a basis for consistent decision-making on the part of the single individual. However, in the case where it is more appropriate for a different individual to assess the outcome's utility, the use of another individual's "decision parameter" (subjective probability) is clearly dangerous. A direct assessment of the "truthfulness" of the proposition is more appropriate in such cases. For example, a research and development manager may be best able to assess the future success of a particular research project, but the utility to the organization is more appropriately assessed at a hierarchically higher level.

Self-consistency in an assessment method, as Professor Tversky remarks, is too weak a criterion. Any "sensibly" constructed assessment method should give self-consistent results, with those methods that the subject finds conceptually more difficult taking a larger number of iterations of feedback to achieve the desired consistency level. The emphasis should rather be on the type of subjective probability derived insofar as it is consistent with the overall situation in which it is functional.

Professor V. P. Godambe (University of Waterloo): I would like to congratulate Professor Suppes for his paper which provides a lucid discussion of some of the intricate issues concerning Bayesian logic. We are assured of his extremely realistic attitude by his emphasis on "relentless programs of empirical investigations" and also by his remark "One could argue that we need special axioms of rationality for special situations . . .". I wonder if for some special situations the condition of "closure under submodels" would not be too restrictive. Anyway I find Professor Suppess's emphasis on distinguishing structural axioms from rationality axioms very valuable. Even though he made this distinction quite explicit as early as 1956, unfortunately Bayesians as well as other statisticians apparently did not take enough notice of it. A clear understanding of this distinction possibly could have helped avoid some confusion found in some previous discussions/controversies concerning some related topics. In this connection I shall be grateful for the following clarification. If a coin is being tossed a number of times one can think of a (behaviouristic) betting situation where one wants to bet on the number of "heads" that would turn up. In arriving at his "bets" he may (or may not) make *assumptions* of independence or exchangeability, etc. Am I right in saying that these assumptions are existential or structural in character? Further "inference" about the "number of heads" can have simple behaviourial or betting interpretation. But similar interpretation for inference about independence or exchangeability would seem too far fetched—unless in some extraordinary situation some simple alternative assumptions are available. It is nice to know that Professor Suppes also finds an "air of paradox" in the Bayesian attitude towards "randomization": Accept randomization to enlarge the decision space but reject experimental randomization. In this connection I would like to ask (perhaps at the risk of going outside the scope of this paper) Professor Suppes what his attitude is towards "experimental randomization". It seems to me that underlying the works of Fisher, Neyman, Pitman, Welch, Yates and others during the thirties on experimental randomization was a motivation (though not made explicit then): Protect the probability statements (inferences) concerning unknown parameters made with the *assumptions* of normality, independence etc., in terms of the frequency distribution generated by the randomization, should the assumptions go wrong. A detailed discussion of this is given in a paper by Thompson and myself (1973). Another relevant reference is Godambe and Thompson (1971).

The authors replied separately, in writing, as follows:

Professor Amos Tversky: My response to the many interesting comments offered by the participants in the discussion centres around two issues. First, the practical implications of the observed biases in judgment under uncertainty, and second the normative and educational implications of these findings.

The realization that our precious intuitive judgments are subject to common and severe biases is a sobering experience. While some people believe that these biases reflect serious deficiencies in our intuitive inductive reasoning—to be counteracted by proper education—others prefer to attribute the observed biases to the unfamiliar, artificial or contrived nature of the judgment task. Indeed, Professor Smith has suggested that the presence of systematic errors in the intuitive assessment of probability are no more surprising than the presence of systematic errors in the intuitive estimation of height, for example, because people are simply not equipped to handle these tasks very effectively. The presence of biases in judgment under uncertainty may be no more surprising than the presence of perceptual biases; however, it is considerably more serious. We have several relatively inexpensive methods for measuring the height of a building or the distance between two points, so we do not have to rely on intuitive estimates of distance or height. But we do not have generally accepted models according to which the probability of a critical event (e.g. the guilt of a defendant or the occurrence of a war) can be computed. Consequently, we must rely on the intuitive judgments of informed individuals and we cannot bypass the problem of biases by the introduction of a physical measurement device. Furthermore, the fact that we often experience difficulties in expressing our belief in the form of probability statements does not necessarily reflect on the meaningfulness of the task. Rather, it indicates that we are not accustomed to acknowledge and express uncertainty in an explicit fashion.

It has often been argued that the biases observed in the psychological laboratory are less likely to occur in situations of great importance in the real world because in such situations people are apt to be more critical and to utilize their knowledge more effectively. Much as I would like to adopt this optimistic position, I know of no evidence that supports it. Analyses of economic, military and political decision-making reveal major errors in the processing and evaluation of evidence—even in situations of the utmost importance. The following example is a case in point.

Several years ago, psychologists investigated an experimental paradigm, called probability learning, where a subject had to predict on each trial which of two lights will turn on. In a typical experiment, the light on the right is lit on 2/3 of the trials and the light on the left is lit on the remaining 1/3 of the trials. The subject is instructed to make as many correct predictions as he can and he is rewarded for his correct predictions. Contrary to the optimal policy of predicting the more frequent light on every trial, people usually probability-match, that is, predict the right light on 2/3 of the trials and the left light on 1/3 of the trials. Moreover, many people maintain (erroneously) that expected gain is maximized by matching the probability with which the lights are turned on. This result has been viewed by some as an example of a relatively simple probabilistic problem in which the optimal solution is non-transparent to naive subjects. Others, however, have maintained that the observed non-optimality is due to the artificial nature of the situation and that it will surely disappear in a realistic context where the payoffs are substantial. The following observation, reported by J. L. Bower of the Harvard Business School, shows that this is not the case.

Fighter pilots in the Pacific during World War Two encountered situations requiring incendiary shells about 1/3 of the time and armour-piercing shells about 2/3 of the time. Since there was no general procedure for predicting on every mission which type of shells would be required, the optimal policy was clearly to use armour-piercing shells on every mission. It was observed, however, that when left to their own devices, pilots armed themselves with incendiary and armour-piercing shells in the proportion of 1 to 2. Thus, the experienced fighter pilots acted much like the naive subjects in the psychological laboratory, even though their own lives were at stake. The point of this example is not to argue that laboratory findings can be readily generalized to the real world. The point is that serious errors of judgment that stem from the lack of proper statistical intuition which are observed in the laboratory under very simple conditions, may very well exist in much more complicated situations where the consequences are considerably more severe. The belief that the

severity of the consequences alone is sufficient to avoid serious errors of judgment is simply unfounded.

Let me turn now to discuss briefly the interplay between descriptive and normative theories of subjective probability and their contribution to the improvement of human judgment. Professor Audley has put the key questions very succinctly. First, how do people make judgments and decisions in the face of uncertainty? Second, how should they make them? Third, how can we teach them to make better judgments and better decisions? Our research project constitutes a modest attempt to answer the first question. I agree wholeheartedly with Professor Lindley that our main goal, as scientists and teachers, should be to help people make better judgments.

There are various ways in which we can contribute towards this goal. First, we could teach people elements of probability theory and inductive logic not as formal schemes but rather as intellectual tools that are applicable to the evaluation of evidence in everyday life. Second, we could introduce courses on probabilistic thinking in both high school and university. Thirdly, we could develop better procedures for the elicitation and the critique of subjective probabilities, and for the analysis of complex decisions—along the lines suggested by the question of Dr Phillips. I do believe that better understanding of the manner in which people evaluate probabilities and make actual decisions could be of great help in the development of better teaching programmes and more effective judgmental technology. Furthermore, I believe that a deeper understanding of the psychology of decision is essential to the proper application of the normative theory—whose interpretation is often a great deal more problematic than some of us are willing to acknowledge in public.

To illustrate, consider again the example of a person who would rather bet £1,000 on a red ball drawn at random from an urn containing 50 red and 50 green balls than on a red ball drawn from an urn containing 100 red and green balls in unknown proportions. Professor Lindley pointed out that such behaviour violates the axioms of rationality. This is certainly true whenever the person in question is concerned only with the monetary payoff. This, however, need not be the case. The person may regard a loss in the 50–50 urn to be less aversive than a loss in the urn of the unknown composition, presumably because he tends to attribute a loss in the former case to bad luck and loss in the latter case to bad decision. In this case, however, the preference for the first urn becomes totally unobjectionable from the standpoint of the normative theory. But is it really unobjectionable? Are we really justified in treating the losses in the two situations differently? Normative decision theorists are quite eager to tell us how we should act, but they are very reluctant to tell us anything about how we should feel. A comprehensive theory of rationality should deal with the legitimacy of utility assignments and not only with the legitimacy of preferences. We need a coherence theory for utility, like the one developed for subjective probability. No such theory has been proposed so far. It is my hope and belief that closer contact and deeper understanding between psychologists and statisticians will help us develop better descriptive theories, more comprehensive normative theories, and better ways for teaching people to cope with the uncertainty that surrounds them.

Professor PATRICK SUPPES: The general issues raised in discussion are numerous, and fully adequate answers would require more space than is available. If I am not able to do justice to the comments, I shall at least try to exhibit the virtue of brevity.

Concerning the comments of Professor Smith, I would like to make two remarks. In his comparison of Savage's results with the absolute scale of temperature, he does not take into account the need for a theory of approximation that we have in all ordinary cases of physical measure. The closest correspondence to axioms on probability are axioms on measurement as, for example, measurements of mass or temperature, and not the theory of the heat-engine. If we consider such matters of physical measurement as the most direct analogue of measurement of probability, then in the physical case as well as in the probabilistic case we need a theory of approximation and a theory of error. The axioms I

have given for upper and lower probabilities can also be taken as a theory for the measurement of mass or the measurement of distance and are meant to provide an explicit analysis of the role of a standard scale on an equal-arm balance or a standard set of measuring rods. From my standpoint it is incongruous to appreciate the necessity for a theory of approximate measurement in the case of physical quantities and yet to seek axioms that yield exact measurements in the case of partial beliefs.

Concerning Professor Smith's last remark about trying to get "precise results from undefined terms", I believe this is simply a misunderstanding of the axiomatic method of investigation. At least since the time of Hilbert it has been clear that we must separate the formal character of the undefined primitive terms of the theory from their intended intuitive interpretation. The intuitive interpretation of the concepts I have introduced for approximate measurement of qualitative probabilities is clear, but from a formal standpoint the primitive terms are not defined, as in the case of geometry or in many other branches of pure mathematics.

Concerning the three main enterprises mentioned by Professor Audley, the axioms for measuring belief that I have given contribute to the development of a more satisfactory normative theory—one that is more attuned to the capacities and limitations of ordinary men. As I stressed in my paper and have stressed in other publications, it is important that not only descriptive theories but also that normative theories be realistic. There is not yet an adequate body of discussion of ways in which normative theories should be realistic. I assume that the kind of approximation of belief I have introduced is one we might well expect. Another feature I have considered only briefly and inexplicitly is the problem of the normative mechanisms for changing belief. As I have indicated, it is my view that Bayes's theorem covers only certain cases when the evidence is completely explicit, as in many cases it is not. The statement of acceptable normative principles here seems more difficult and is a subject that as yet has hardly been investigated. The thrust of these remarks should make it evident that I am also sceptical about educating people simply to be Bayesians. Although they should understand Bayesian principles, I would also argue for the inadequacy of Bayesian principles to provide a full set of normative principles of rational action, especially in terms of the problem just mentioned of always perceiving the explicit new evidence, and of being able to conceptualize it and its likelihood under alternative hypotheses.

Concerning Professor Dickey's remarks, I would agree that we need to be able to use structural axioms, but the point of my analysis is to make explicit that some axioms are structural and some are axioms of pure rationality. Much of the earlier literature has not been clear on this point. I shall not try to comment on Professor Dickey's particular example on scoring procedures, but I want to mention the current work in this direction based on the earlier work of de Finetti (1965) and Shuford *et al.* (1966).

Concerning Dr Lovie's remarks, it seems to me that his excellent examples of the difficulties human beings have with probability concepts reinforce any need for an approximate rather than an exact theory of the measurement of belief. It would be my hope that the kind of standard events I have introduced in my theory of inexact measurement of beliefs might in a modest way eventually play the role that the use of standard weights plays in the measurement of mass, or the use of standard measuring rods in the case of the measurement of length. The point would be to try to make everyone familiar with the properties of the standard events, and then their judgments about more complex probabilistic phenomena would be referred to their more intimate and detailed knowledge of the restricted set of standard events.

Concerning Professor Lindley's admonitions that psychologists should teach Bayesian decision-making rather than investigate the fallibilities of most men's practices, I would reply that it is the obligation of the psychologist *qua* scientist to teach the truth, not to preach it. I am all in favour of teaching sound methods of decision-making, but the psychological task of understanding the subtle and complex problems of how real choices are made remains still largely unexplored territory.

Concerning Dr Bevans's queries about the narrow use of the concept of belief in the theory of subjective probability, I would agree that many aspects of the logic and psychology of belief are not covered by present formulations of the theory, including the one presented in my own paper. A specific subtle example is provided by the discussions in philosophy of the logic of propositional attitudes. In the present case, a propositional attitude is a statement of belief that in general will be non-truth functional in character, e.g. from the assertion: Jones believes that p, and from knowledge of the truth or falsity of p, it is not possible to infer the truth or falsity of the statement about Jones's belief. From the standpoint of the theory of beliefs as developed in the context of subjective probability, the most glaring difficulty that comes from the logic of propositional attitudes is the intensional character of belief statements. Thus, we may say: Jones believes that the author of *Middlemarch* is a woman, and from the identity, George Eliot = the author of *Middlemarch*, we cannot infer that Jones believes that George Eliot is the author of *Middlemarch*. For example, Jones may have seen a list of famous novels written by women and may have remembered *Middlemarch* being on the list, but have forgotten the author. These intensional matters have not been seriously discussed in any of the traditional literature on subjective probability beginning with either Ramsey or de Finetti. What I have had to say about propositional attitudes is one particular way of interpreting Dr Bevans's statements about beliefs being characteristic of "ways of reacting to experience". As to the question of whether upper and lower probabilities are a halfway house on the road to second-order probabilities, I do not think this is the case, because I feel that the introduction of second-order probabilities introduces an overly elaborate theory for the phenomena at hand. However, it is too early to tell whether I am correct in this judgment. Concerning learning models, I certainly do not think that purely statistical models will provide an adequate theory of learning; on the other hand, I think that statistical considerations or stochastic considerations are integral to any learning theories likely to be developed in the near future and that have any hope of standing up under extensive experimental test.

Concerning Professor Godambe's questions about the assumptions of independence or exchangeability being structural assumptions in the sense defined in my paper, I would certainly agree that they are. In another publication, I have, for example, made qualitative assumptions of independence and exchangeability critical structural assumptions (Suppes, 1973). Concerning experimental randomization, I agree with Professor Godambe that the motivation to be found for this in the works of Fisher and others is the protection of the theoretical assumptions made about the probability distribution of various quantities. Approached in this manner, randomization becomes almost like linearization of a problem, something that is done for the sake of computational convenience and mathematical simplicity, not for any deep-running conceptual reason. I think the same kind of case can be made for the problems of bias in sampling as an argument for randomization. With effort we can make a good case for avoiding randomization, but it is the most efficient and the simplest way of protecting ourself from bias in experimental design. In previous writings, I have emphasized the elimination of bias as an important argument for experimental randomization, but the theoretical argument Godambe makes explicit is, I think, more powerful and important.

References in the Discussion

Beach, L. R. and Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organ. Behav. and Hum. Perf.*, **3**, 109–123.

Brier, G. W. and Allen, R. A. (1951). Verification of weather forecasts. In *Compendium of Meteorology* (T. F. Malone, ed.), pp. 841–848. Boston, Mass.: American Meteorological Society.

De Finetti, B. (1965). Methods of discriminating levels of partial knowledge concerning a test item. *Brit. J. Math. Statist. Psychol.*, **13**, 87–123.

DICKEY, J. M. and WALRATH, JUDY (1971). Computers and Bayesian statistical inference in the analysis of clinical data. Res. Rep. No. 56, Dept of Statistics, State University of New York at Buffalo.

GODAMBE, V. P. and THOMPSON, MARY E. (1971) Bayes, fiducial and frequency aspects of statistical inference in regression analysis and survey-sampling. *J. R. Statist. Soc.* B, **33**, 361–390.

—— (1973). Philosophy of survey-sampling practice. Paper delivered at the London (Ontario) meetings (to be published).

KRANTZ, D. H., LUCE, R. D., SUPPES, P. and TVERSKY, A. (1971). *Foundations of Measurement*, Vol. 1. New York: Academic Press.

RAPOPORT, A. (1973). Research paradigms for studying dynamic decision behaviour. Paper delivered at the Fourth Research Conference on Subjective Probability, Utility and Decision-making (to be published).

SALWAY, J. (1971). Unpublished paper, with the Department of Psychology, University of Liverpool.

SANDERS, F. (1958). The evaluation of subjective probability forecasts. Scientific Report No. 5, Contract No. AFCRC-TN-58-465, M.I.T., Cambridge, Mass.

SHUFORD, E. H. (1964). Some Bayesian learning processes. In *Human Judgments and Optimality* (M. S. Shelly and G. L. Bryan, eds), pp. 127–152. New York: Wiley.

SHUFORD, E. H., ALBERT, A. and MASSENGILL, H. (1966). Admissible probability measurement procedures. *Psychometrika*, **31**, 125–145.

SLOVIC, P. and LICHTENSTEIN, S. C. (1968). The relative importance of probabilities and payoffs in risk taking. *J. Exper. Psychol. Mono. Suppl.*, **78**, No. 3, Part 2.

SUPPES, P. (1973). New Foundations of objective probability: axioms for propensities. In *Logic, Methodology and Philosophy of Science* (P. Suppes *et al.*, eds). Amsterdam: North-Holland.

TODA, M. (1962). The design of a fungus-eater: a model of human behaviour in an unsophisticated environment. *Behav. Sci.*, **7**, 164–183.