

Qualitative Theory of Subjective Probability

Patrick Suppes
Stanford University

Because we want to use probability concepts to talk about everything from the chance of drawing four aces in a hand of bridge to the probability of rain tomorrow or the probability distribution of position in a quantum-mechanical experiment, it is hardly surprising that no simple categorical theory of probability can be found. The subjective theory of probability accepts this diversity of applications, and, in fact, utilizes it to argue that the many ways in which information must be processed to obtain a probability distribution do not admit of categorical codification. Consequently, two reasonable men in approximately the same circumstances can hold differing beliefs about the probability of an event as yet unobserved. For example, according to the subjective theory of probability, two meteorologists can be presented with the same weather map and the same history of observations of basic meteorological variables such as surface temperature, air pressure, humidity, upper air pressures, wind, etc., and yet still differ in the numerical probability they assign to the forecast of rain tomorrow morning. I hasten to add, however, that the term “subjective” can be misleading. It is not part of the subjective theory of probability to countenance every possible diversity in the assignment of subjective probabilities. It is a proper and important part of subjective theory to analyze, e.g., how classical relative-frequency data are to be incorporated into proper estimates of subjective probability. Such data are obviously

important in any systematic estimate of probabilities, as we can see from examination of the scientific literature in which probabilities play a serious part. It is also obvious that the principles of symmetry naturally applied in the classical definition of probability play an important role in the estimation of subjective probabilities whenever they are applicable.

Bayes' theorem provides an example of the sort of strong constraints to be placed on any subjective theory. The prior probability distributions selected by different investigators can differ widely without violating the subjective theory; but if these investigators agree on the method of obtaining further evidence, and if common observations are available to them, then these commonly accepted observations will often force their beliefs to converge.

2.1 DE FINETTI'S QUALITATIVE AXIOMS

Let us turn now to a more systematic discussion of the major aspects of the subjective theory. For a more detailed treatment of many questions the reader is referred to the historically important article of de Finetti (1937), which has been translated in Kyburg and Smokler (1964), and also to de Finetti's treatise (1974; 1975). The 1937 article of de Finetti's is one of the important pieces of work in the foundations of probability in this century. Probably the most influential work on these matters since 1950 is the book by Savage (1954). Savage extends de Finetti's ideas by paying greater attention to the behavioral aspects of decisions, but this extension cannot be examined in any detail in this chapter.

Perhaps the best way to begin a systematic analysis of the subjective theory is by a consideration of de Finetti's axioms for qualitative probability. The spirit of these axioms is to place restraints on qualitative judgments of probability which will be sufficient to prove a standard representation theorem, i.e. to guarantee the existence of a numerical probability measure in the standard sense. From this standpoint the axioms may be regarded as a contribution to the theory of measurement with particular reference to comparative judgments of probability. The central question for such a set of axioms is how complicated must be the condition on the qualitative relation *more probable than* in order to obtain a numerical probability measure over events.

The intuitive idea of using a comparative qualitative relation is that individuals can realistically be expected to make such judgments in a direct way, as they cannot when the comparison is required to be quantitative. On most occasions I can say unequivocally whether I think it is more likely to rain or not in the next four hours at Stanford, but I cannot in the same direct way make a judgment of how much more likely it is not to rain than rain. Generalizing this example, it is a natural move on the subjectivist's part to next ask what formal properties a qualitative comparative relation must have in order to be represented by a standard probability measure. (Later we shall review

some of the experimental literature on whether people's qualitative judgments do have the requisite properties.)

We begin with the concept of a qualitative probability structure, the axioms for which are very similar formally to those for a finitely additive probability space. The set-theoretical realizations of the theory are triples $(\Omega, \mathcal{F}, \succeq)$ where Ω is a nonempty set, \mathcal{F} is a family of subsets of Ω , and the relation \succeq is a binary relation on \mathcal{F} . We follow here the discussion given in Luce & Suppes (1965).

Definition 1 A structure $\Omega = (\Omega, \mathcal{F}, \succeq)$ is a *qualitative probability structure* if the following axioms are satisfied for all A, B , and C in \mathcal{F} :

- S1. \mathcal{F} is an algebra of sets on Ω ;
- S2. If $A \succeq B$ and $B \succeq C$, then $A \succeq C$;
- S3. $A \succeq B$ or $B \succeq A$;
- S4. If $A \cap C = \emptyset$ and $B \cap C = \emptyset$, then $A \succeq B$ if and only if $A \cup C \succeq B \cup C$;
- S5. $A \succeq \emptyset$;
- S6. Not $\emptyset \succeq \Omega$.

The first axiom on \mathcal{F} is the same as the first axiom of finitely additive probability spaces. Axioms S2 and S3 just assert that \succeq is a weak ordering of the events in \mathcal{F} . Axiom S4 formulates in qualitative terms the important and essential principle of additivity of mutually exclusive events. Axiom S5 says that any event is (weakly) more probable than the impossible event, and Axiom S6 that the certain event is strictly more probable than the impossible event. Defining the strict relation $>$ in the customary fashion:

$$A > B \text{ if and only if not } B \succeq A,$$

we may state the last axiom as: $\Omega > \emptyset$.

To give a somewhat deeper sense of the structure imposed by the axioms, we state some of the intuitively desirable and expected consequences of the axioms. It is convenient in the statement of some of the theorems to use the (weakly) less probable relation, defined in the usual manner.

$$A \preceq B \text{ if and only if } B \succeq A.$$

The first theorem says that \preceq is an extension of the subset relation.

Theorem 1 If $A \subseteq B$, then $A \preceq B$.

Proof. Suppose on the contrary, that not $A \preceq B$, i.e. that $A > B$. By hypothesis $A \subseteq B$, so there is a set C , disjoint from A such that $A \cup C = B$. Then, because $A \cup \emptyset \neq A$, we have at once

$$A \cup \emptyset = A > B = A \cup C,$$

whence by contraposition of Axiom S4, $\emptyset > C$, which contradicts Axiom S5.

Q.E.D.

Some other elementary properties follow.

Theorem 2

- (i) If $\emptyset < A$ and $A \cap B = \emptyset$, then $B < A \cup B$;
- (ii) if $A \geq B$, then $-B \geq -A$;
- (iii) if $A \geq B$ and $C \geq D$ and $A \cap C = \emptyset$, then $A \cup C \geq B \cup D$;
- (iv) if $A \cup B \geq C \cup D$ and $C \cap D = \emptyset$, then $A \geq C$ or $B \geq D$;
- (v) if $B \geq -B$ and $-C \geq C$, then $B \geq C$.

Because it is relatively easy to prove that a qualitative probability structure has many of the expected properties, as reflected in the preceding theorems, it is natural to ask the deeper question whether or not it has all of the properties necessary to guarantee the existence of a numerical probability measure P such that for any events A and B in \mathfrak{F}

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

If Ω is an infinite set, it is moderately easy to show that the axioms of Definition 1 are not strong enough to guarantee the existence of such a probability measure. General arguments from the logical theory of models in terms of infinite models of arbitrary cardinality suffice; a counterexample is given in Savage (1954, p. 41). De Finetti stressed the desirability of obtaining an answer in the finite case. Kraft, Pratt & Seidenberg (1959) showed that the answer is also negative when Ω is finite; in fact, they found a counterexample for a set Ω having five elements, and, thus, 32 subsets. The gist of their counterexample is the following construction. Let $\Omega = \{a, b, c, d, e\}$, and let ϕ be a measure (not a probability measure) such that

$$\begin{aligned} \phi(a) &= 4 - \epsilon \\ \phi(b) &= 1 - \epsilon \\ \phi(c) &= 2 \\ \phi(d) &= 3 - \epsilon \\ \phi(e) &= 6, \end{aligned}$$

and

$$0 < \epsilon < \frac{1}{3}.$$

Now order the 32 subsets of Ω according to this measure—the ordering being, of course, one that satisfies Definition 1. We then have the following strict inequalities in the ordering:

$$\begin{aligned} \{a\} &> \{b, d\} \text{ because } \phi(a) = 4 - \epsilon > 4 - 2\epsilon = \phi(b) + \phi(d) \\ \{c, d\} &> \{a, b\} \text{ because } \phi(c) + \phi(d) = 5 - \epsilon > 5 - 2\epsilon = \phi(a) + \phi(b) \\ \{b, e\} &> \{a, d\} \text{ because } \phi(b) + \phi(e) = 7 - \epsilon > 7 - 2\epsilon = \phi(a) + \phi(d) \end{aligned}$$

We see immediately also that any probability measure P that preserves these three inequalities implies that

$$\{c, e\} > \{a, b, d\},$$

as may be seen just by adding the three inequalities. In the case of ϕ

$$\phi(c) + \phi(e) = 8 > 8 - 3\varepsilon = \phi(a) + \phi(b) + \phi(d).$$

However, no set A different from $\{c, e\}$ and $\{a, b, d\}$ has the property that

$$\phi(\{c, e\}) \geq \phi(A) \geq \phi(\{a, b, d\}).$$

Thus we can modify the ordering induced by ϕ to the extent of setting

$$\{a, b, d\} > \{c, e\} \tag{II}$$

without changing any of the other inequalities. But no probability measure can preserve (II) as well as the three earlier inequalities, and so the modified ordering satisfies Definition 1 but cannot be represented by a probability measure.

Of course, it is apparent that by adding special structural assumptions to the axioms of Definition 1 it is possible to guarantee the existence of a probability measure satisfying (I). In the finite case, for example, we can demand that all the atomic events be equiprobable, although this is admittedly a very strong requirement to impose.

Fortunately, a simple general solution of the finite case has been found by Scott (1964). (Necessary and sufficient conditions for the existence of a probability measure in the finite case were formulated by Kraft, Pratt and Seidenberg, but their multiplicative conditions are difficult to understand. Scott's treatment represents a real gain in clarity and simplicity.) The central idea of Scott's *formulation* is to impose an algebraic condition on the indicator (or characteristic) functions of the events. Recall that the indicator function of a set is just the function that assigns the value 1 to elements of the set and the value 0 to all elements outside the set. For simplicity of notation, if A is a set we shall denote by A^i its indicator function. Thus if A is an event

$$A^i(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Scott's conditions are embodied in the following theorem, whose proof we do not give.

Theorem 3 (Scott's representation theorem). Let Ω be a finite set and \geq a binary relation on the subsets of Ω . Necessary and sufficient conditions that

there exists a probability measure P on Ω satisfying (I) are the following: for all subsets A and B of Ω ,

- (1) $A \geq B$ or $B \geq A$;
- (2) $A \geq 0$;
- (3) $\Omega > 0$;
- (4) for all subsets $A_0, \dots, A_n, B_0, \dots, B_n$ of Ω , if $A_i \geq B_i$ for $0 \leq i < n$, and for all x in Ω

$$A'_0(x) + \dots + A'_n(x) = B'_0(x) + \dots + B'_n(x),$$

then $A_n \leq B_n$.

To illustrate the force of Scott's condition (4), we may see how it implies transitivity. First, necessarily for any three indicator functions

$$A^i + B^i + C^i = B^i + C^i + A^i,$$

i.e. for all elements x

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

By hypothesis $A \geq B$ and $B \geq C$, whence by virtue of condition (4),

$$C \leq A,$$

and thus by definition, $A \geq C$, as desired. The algebraic equation of condition (4), just requires that any element of Ω , i.e. any atomic event, belong to exactly the same number of A_i and B_i , for $0 \leq i \leq n$. Obviously, this algebraic condition cannot be formulated in the simple set language of Definition 1 and thus represents quite a strong condition.

2.2 GENERAL QUALITATIVE AXIOMS

In the case that Ω is infinite, a number of strong structural conditions have been shown to be sufficient but not necessary. For example, de Finetti (1937) and independently Koopman (1940a, 1940b, 1941) use an axiom to the effect that there exist partitions of Ω into arbitrarily many events equivalent in probability. This axiom, together with those of Definition 1, is sufficient to prove the existence of a numerical probability measure. Related existential conditions are discussed in Savage (1954). A detailed review of these various conditions is to be found in Chapters 5 and 9 of Krantz et al. (1971). However, as is shown in Suppes & Zanotti (1976), by going slightly beyond the indicator functions, simple necessary and sufficient conditions can be given for both the finite and infinite case.

In the present case the move is from an algebra of events to the algebra \mathfrak{F}^* of extended indicator functions relative to \mathfrak{F} . The algebra \mathfrak{F}^* is just the

smallest semigroup (under function addition) containing the indicator functions of all events in \mathcal{F} . In other words, \mathcal{F}^* is the intersection of all sets with the property that if A is in \mathcal{F} then A' is in \mathcal{F}^* and if A^* and B^* are in \mathcal{F}^* , then $A^* + B^*$ is in \mathcal{F}^* ; It is easy to show that any function A^* in \mathcal{F}^* is an integer-valued function defined on Ω . It is the extension from indicator functions to integer-valued functions that justifies calling the elements of \mathcal{F}^* extended indicator functions.

The qualitative probability ordering must be extended from \mathcal{F} to \mathcal{F}^* , and the intuitive justification of this extension must be considered. Let A^* and B^* be two extended indicator functions in \mathcal{F}^* . Then, to have $A^* \succeq B^*$ is to have the expected value of A^* equal to or greater than the expected value of B^* . As should be clear, extended indicator functions are just random variables of a restricted sort. The qualitative comparison is now not one about the probable occurrences of events, but about the expected value of certain restricted random variables. The indicator functions themselves form, of course, a still more restricted class of random variables, but qualitative comparison of their expected values is conceptually identical to qualitative comparison of the probable occurrences of events.

There is more than one way to think about the qualitative comparisons of the expected value of extended indicator functions, and so it is useful to consider several examples.

- (1) Suppose Smith is considering two locations to fly to for a weekend vacation. Let A_j be the event of sunny weather at location j and B_j be the event of warm weather at location j . The qualitative comparison Smith is interested in is the expected value of $A_1^i + B_1^i$ versus the expected value of $A_2^i + B_2^i$. It is natural to insist that the utility of the outcomes has been too simplified by the sums $A_j^i + B_j^i$. The proper response is that the expected values of the two functions are being compared as a matter of belief, not value or utility. Thus it would seem quite natural to bet that the expected value of $A_1^i + B_1^i$ will be greater than that of $A_2^i + B_2^i$, no matter how one feels about the relative desirability of sunny versus warm weather. Put another way, within the context of decision theory, extended indicator functions are being used to construct the subjective probability measure, not the measure of utility.

Note that if Smith prefers the country ($j = 1$) to the city ($j = 2$) when it is warm and sunny, then even if

$$A_1^i + B_1^i \approx A_2^i + B_2^i$$

in belief, his choice of country or city could vary depending on the degree of belief or expectation: with high expectation go to the country; with low expectation go to the city.

- (2) Consider a particular population of n individuals, numbered $1, \dots, n$. Let A_j be the event of individual j going to Hawaii for a vacation this year, and let B_j be the event of individual j going to Acapulco. Then define

$$A^* = \sum_{i=1}^n A_j^i \text{ and } B^* = \sum_{i=1}^n B_j^i.$$

Obviously A^* and B^* are extended indicator functions—we have left implicit the underlying set Ω . It is meaningful and quite natural to qualitatively compare the expected values of A^* and B^* . Presumably such comparisons are in fact of definite significance to travel agents, airlines, and the like.

We believe that such qualitative comparisons of expected value are natural in many other contexts as well. What the representation theorem below shows is that very simple necessary and sufficient conditions on the qualitative comparison of extended indicator functions guarantee existence of a strictly agreeing, finitely additive measure in the sense of (I), whether the set Ω of possible outcomes is finite or infinite.

The axioms are embodied in the definition of a qualitative algebra of extended indicator functions. Several points of notation need to be noted. First, Ω^i and \emptyset^i are the indicator or characteristic functions of the set Ω of possible outcomes and the empty set \emptyset , respectively. Second, the notation nA^* for a function in \mathcal{F}^* is just the standard notation for the (functional) sum of A^* with itself n times. Third, the same notation is used for the ordering relation on \mathcal{F} and \mathcal{F}^* , because the one on \mathcal{F}^* is an extension of the one on \mathcal{F} : for A and B in \mathcal{F} ,

$$A \geq B \text{ iff } A^i \geq B^i.$$

Finally, the strict ordering relation $>$ is defined in the usual way: $A^* > B^*$ iff $A^* \geq B^*$ and not $B^* \geq A^*$.

Definition 2 Let Ω be a nonempty set, let \mathcal{F} be an algebra of sets on Ω , and let \geq be a binary relation on \mathcal{F}^* , the algebra of extended indicator functions relative to \mathcal{F} . Then the qualitative algebra $(\Omega, \mathcal{F}, \geq)$ is *qualitatively satisfactory* if and only if the following axioms are satisfied for every A^*, B^* , and C^* in \mathcal{F}^* :

- Axiom 1** The relation \geq is a weak ordering of \mathcal{F}^* ;
Axiom 2 $\Omega^i > \emptyset^i$;
Axiom 3 $A^* \geq \emptyset^i$;
Axiom 4 $A^* \geq B^*$ iff $A^* + C^* \geq B^* + C^*$;

Axiom 5 If $A^* \succ B^*$ then for every C^* and D^* in \mathcal{F}^* there is a positive integer n such that

$$nA^* + C^* \succeq nB^* + D^*.$$

These axioms should seem familiar from the literature on qualitative probability. Note that Axiom 4 is the additivity axiom that closely resembles de Finetti's additivity axiom for events: If $A \cap C = B \cap C = \emptyset$, then $A \succeq B$ iff $A \cup C \succeq B \cup C$. As we move from events to extended indicator functions, functional addition replaces union of sets. What is formally of importance about this move is seen already in the exact formulation of Axiom 4. The additivity of the extended indicator functions is unconditional—there is no restriction corresponding to $A \cap C = B \cap C = \emptyset$. The absence of this restriction has far-reaching formal consequences in permitting us to apply without any real modification the general theory of extensive measurement. Axiom 5 has, in fact, the exact form of the Archimedean axiom used in Krantz et al. (1971, p. 73) in giving necessary and sufficient conditions for extensive measurement.

Theorem 4 Let Ω be a nonempty set, let \mathcal{F} be an algebra of sets on Ω and let \succeq be a binary relation on \mathcal{F} . Then a necessary and sufficient condition that there exists a strictly agreeing probability measure on \mathcal{F} is that there be an extension of \succeq from \mathcal{F} to \mathcal{F}^* such that the qualitative algebra of extended indicator functions $(\Omega, \mathcal{F}^*, \succeq)$ is qualitatively satisfactory. Moreover, if $(\Omega, \mathcal{F}^*, \succeq)$ is qualitatively satisfactory, then there is a unique strictly agreeing expectation function on \mathcal{F}^* and this expectation function generates a unique strictly agreeing probability measure on \mathcal{F} .

Proof. The main tool used in the proof is from the theory of extensive measurement: necessary and sufficient conditions for existence of a numerical representation, as given in Krantz *et al.* (1971, pp. 73–74). In particular, let A be a nonempty set, \succeq a binary relation on A , and \circ a binary operation closed on A . Then there exists a numerical function ϕ on A unique up to a positive similarity transformation (i.e. multiplication by a positive real number) such that for a and b in A

- (i) $a \succeq b$ iff $\phi(a) \geq \phi(b)$,
- (ii) $\phi(a \circ b) = \phi(a) + \phi(b)$

if and only if the following four axioms are satisfied for all a, b, c , and d in A :

- E1. The relation \succeq is a weak ordering of A ;
- E2. $a \circ (b \circ c) \approx (a \circ b) \circ c$, where \approx is the equivalence relation defined in terms of \succeq ;
- E3. $a \succeq b$ iff $a \circ c \succeq b \circ c$ iff $c \circ a \succeq c \circ b$;

E4. If $a > b$ then for any c and d in A there is a positive integer n such that $na \circ c \geq nb \circ d$, where na is defined inductively.

It is easy to check that qualitatively satisfactory algebras of extended indicator functions as defined above satisfy these four axioms for extensive measurement structures. First, we note that functional addition is closed on \mathcal{F}^* . Second, Axiom 1 is identical to E1. Extensive Axiom E2 follows immediately from the associative property of numerical functional addition: for any A^* , B^* , and C^* in \mathcal{F}^*

$$A^* + (B^* + C^*) = (A^* + B^*) + C^*$$

and so we have not just equivalence but identity. Axiom E3 follows from Axiom 4 and the fact that numerical functional addition is commutative. Finally, E4 follows from the essentially identical Axiom 5.

Thus, for any qualitatively satisfactory algebra $(\Omega, \mathcal{F}^*, \geq)$ we can infer there is a numerical function ϕ on Ω such that for A^* and B^* in \mathcal{F}^* .

- (i) $A^* \geq B^*$ iff $\phi(A^*) \geq \phi(B^*)$,
- (ii) $\phi(A^* + B^*) = \phi(A^*) + \phi(B^*)$,

and ϕ is unique up to a positive similarity transformation.

Second, since for every A^* in \mathcal{F}^*

$$A^* + \emptyset' = A^*$$

we have at once that from (ii)

$$\phi(\emptyset') = 0.$$

Since $\Omega^i > \emptyset'$ by Axiom 2, we can choose

$$\phi(\Omega^i) = 1.$$

And thus have a standard (unique) expectation function E for extended indicator functions:

- (i) $E(\emptyset') = 0$
- (ii) $E(\Omega^i) = 1$
- (iii) $E(A^* + B^*) = E(A^*) + E(B^*)$.

But such an expectation function for \mathcal{F}^* defines a unique probability measure P on \mathcal{F} when it is restricted to the indicator functions in \mathcal{F}^* , i.e. for A in \mathcal{F} , we define

$$P(A) = E(A').$$

Thus the axioms are sufficient, but it is also obvious that the only axioms, Axioms 2 and 3, that go beyond those for extensive structures are also necessary for a probabilistic representation. From the character of extended

indicator functions, it is also clear that for each probability measure there is a unique extension of the qualitative ordering from \mathcal{F} to \mathcal{F}^* .

Q.E.D.

The proof just given, even more than the statement of the theorem itself, shows what subset of random variables defined on a probability space suffices to determine the probability measure in a natural way. Our procedure has been to axiomatize in qualitative fashion the expectation of the extended indicator functions. There was no need to consider all random variables, and, on the other hand, the more restricted set of indicator functions raises the same axiomatic difficulties confronting the algebra of events.

2.3 QUALITATIVE CONDITIONAL PROBABILITY

One of the more troublesome aspects of the qualitative theory of conditional probability is that $A|B$ is not an object—in particular it is not a new event composed somehow from events A and B . Thus the qualitative theory rests on a quaternary relation $A|B \succeq C|D$, which is read: event A given event B is at least as probable as event C given event D . There have been a number of attempts to axiomatize this quaternary relation (Koopman, 1940a, 1940b; Aczél, 1961, 1966, p. 319; Luce, 1968; Domotor, 1969; Krantz et al., 1971; and Suppes, 1973). The only one of these axiomatizations to address the problem of giving necessary and sufficient conditions is the work of Domotor, which approaches the subject in the finite case in a style similar to that of Scott (1964).

By using indicator functions or, more generally, extended indicator functions, the difficulty of $A|B$ not being an object is eliminated, for $A^i|B$ is just the indicator function of the set A restricted to the set B , i.e. $A^i|B$ is a partial function whose domain is B . In similar fashion if X is an extended indicator function, $X|A$ is that function restricted to the set A . The use of such partial functions requires care in formulating the algebra of functions in which we are interested, for functional addition $X|A + Y|B$ will not be well defined when $A \neq B$ but $A \cap B \neq \emptyset$. Thus, to be completely explicit we begin with a nonempty set Ω , the probability space, and an algebra \mathcal{F} of events, i.e. subsets of Ω , with it understood that \mathcal{F} is closed under union and complementation. Next we extend this algebra to the algebra \mathcal{F}^* of extended indicator functions, i.e. the smallest semigroup (under function addition) containing the indicator functions of all events in \mathcal{F} . This latter algebra is now extended to include as well all partial functions on Ω that are extended indicator functions restricted to an event in \mathcal{F} . We call this algebra of partial extended indicator functions

\mathcal{RF}^* , or, if complete explicitness is needed, $\mathcal{RF}^*(\Omega)$. From this definition it is clear that if $X|A$ and $Y|B$ are in \mathcal{RF}^* , then

- (i) $A = B$, $X|A + Y|B$ is in \mathcal{RF}^* .
- (ii) If $A \cap B = \emptyset$, $X|A \cup Y|B$ is in \mathcal{RF}^* .

In the more general setting of decision theory or expected utility theory there has been considerable discussion of the intuitive ability of a person to directly compare his preferences or expectations of two decision functions with different domains of restriction. Without reviewing this literature, we do want to state that we find no intuitive general difficulty in making such comparisons. Individual cases may present problems, but not necessarily because of different domains of definition. In fact, we believe comparisons of expectations under different conditions is a familiar aspect of ordinary experience. In the present setting the qualitative comparison of restricted expectations may be thought of as dealing only with beliefs and not utilities. The fundamental ordering relation is a weak ordering \succeq of \mathcal{RF}^* with strict order $>$ and equivalence \approx defined in the standard way.

Following Suppes & Zanotti (1982), we give axioms that are strong enough to prove that the probability measure constructed is unique when it is required to cover expectation of random variables. It is worth saying something more about this problem of uniqueness. The earlier papers mentioned have all concentrated on the existence of a probability distribution, but from the standpoint of a satisfactory theory it seems obvious for many different reasons that one wants a unique distribution. For example, if we go beyond properties of order and have uniqueness only up to a convex polyhedron of distributions, as is the case with Scott's axioms for finite probability spaces, we are not able to deal with a composite hypothesis in a natural way, because the addition of the probabilities is not meaningful.

Definition 3 Let Ω be a nonempty set, let $\mathcal{RF}^*(\Omega)$ be an algebra of partial extended indicator functions, and let \succeq be a binary relation on \mathcal{RF}^* . Then the structure $(\Omega, \mathcal{RF}^*, \succeq)$ is a *partial qualitative expectation structure* if and only if the following axioms are satisfied for every X and Y in \mathcal{F}^* and every A, B and C in \mathcal{F} with $A, B > \emptyset$:

Axiom 1 The relation \succeq is a weak ordering of \mathcal{RF}^* ;

Axiom 2 $\Omega^i > \emptyset^i$;

Axiom 3 $\Omega^i|A \succeq C^i|B \succeq \emptyset^i|A$;

Axiom 4 If $X_1|A \succeq Y_1|B$ and $X_2|A \succeq Y_2|B$ then

$$X_1|A + X_2|A \succeq Y_1|B + Y_2|B;$$

Axiom 5 If $X_1|A \leq Y_1|B$ and $X_1|A + X_2|A \succeq Y_1|B + Y_2|B$ then

$$X_2|A \succeq Y_2|B;$$

Axiom 6 If $A \subseteq B$ then

$$X|A \geq Y|A \text{ iff } X \cdot A^i|B \geq Y \cdot A^i|B;$$

Axiom 7 (Archimedean). If $X|A > Y|B$ then for every Z in \mathcal{F}^* there is a positive integer n such that

$$nX|A \geq nY|B + Z|B.$$

The axioms are simple in character and their relation to the axioms of Definition 2 is apparent. The first three axioms are very similar. Axiom 4, the axiom of addition, must be relativized to the restricted set. Notice that we have a different restriction on the two sides of the inequality. The really new axiom is Axiom 6. In terms of events and numerical probability, this axiom corresponds to the following: If $A \subseteq B$, then

$$P(C|A) \geq P(D|A) \text{ iff } P(C \cap A|B) \geq P(D \cap A|B).$$

Note that in the axiom itself, function multiplication replaces intersection of events. (Closure of \mathcal{F}^* under function multiplication is easily proved.) This axiom does not seem to have previously been used in the literature. Axiom 7 is the familiar and necessary Archimedean axiom.

We now state the main theorem. In the theorem we refer to a strictly agreeing expectation function on $\mathcal{RF}^*(\Omega)$. From standard probability theory and conditional expected utility theory, it is evident that the properties of this expectation should be the following for $A, B > \emptyset$:

- (1) $E(X|A) \geq E(Y|B)$ iff $X|A \geq Y|B$,
- (2) $E(X|A + Y|A) = E(X|A) + E(Y|A)$,
- (3) $E(X \cdot A^i|B) = E(X|A)E(A^i|B)$ if $A \subseteq B$,
- (4) $E(\emptyset^i|A) = 0$ and $E(\Omega^i|A) = 1$.

Using primarily (3), it is then easy to prove the following property, which occurs in the earlier axiomatic literature mentioned above:

$$E(X|A \cup Y|B) = E(X|A)E(A^i|A \cup B) + E(Y|B)E(B^i|A \cup B),$$

for $A \cap B = \emptyset$.

Theorem 5 Let Ω be a nonempty set, let \mathcal{F} be an algebra of sets on Ω , and let \geq be a binary relation on $\mathcal{F} \times \mathcal{F}$. Then a necessary and sufficient condition that there is a strictly agreeing conditional probability measure on $\mathcal{F} \times \mathcal{F}$ is that there is an extension \geq^* of \geq from $\mathcal{F} \times \mathcal{F}$ to $\mathcal{RF}^*(\Omega)$ such that the structure $(\Omega, \mathcal{RF}^*(\Omega), \geq^*)$ is a partial qualitative expectation structure. Moreover, if $(\Omega, \mathcal{RF}^*(\Omega), \geq^*)$ is a partial qualitative expectation structure, then there is a unique strictly agreeing expectation function on $\mathcal{RF}^*(\Omega)$ and this expectation generates a unique strictly agreeing conditional probability measure on $\mathcal{F} \times \mathcal{F}$.

The proof is given in Suppes & Zanotti (1982).

2.4 GENERAL ISSUES

I now want to turn to a number of general issues that arise in evaluating the correctness or usefulness of the subjective view of probability.

Use of symmetries

A natural first question is to ask how subjective theory utilizes the symmetries that are such a natural part of the classical, Laplacian definition of probability. If we think in terms of Bayes' theorem the answer seems apparent. The symmetries that we all accept naturally in discussing games of chance are incorporated immediately in the subjective theory as prior probabilities. Thus, for example, if I wish to determine the probability of getting an ace in the next round of cards dealt face up in a hand of stud poker, given the information that one ace is already showing on the board, I use as prior probabilities the natural principles of symmetry for games of chance, which are a part of the classical definition. Of course, if I wish to introduce refined corrections I could do so, particularly corrections arising from the fact that in ordinary shuffling, the permutation groups introduced are groups of relatively small finite order, and, therefore, there is information carry-over from one hand to another. These second-order refinements with respect to shuffling are simply an indication of the kind of natural corrections that arise in the subjective theory and that would be hard to deal with in principle within the framework of the classical definition of probability. On the other hand, I emphasize that the principles of symmetry used in the classical definition are a natural part of the prior probabilities of an experienced card player. The extent to which these symmetries are compelling is a point I shall return to later.

Use of relative frequencies

It should also be clear that the proper place for the use of relative-frequency data in the subjective theory is in the computation of posterior probabilities. It is clear what is required in order to get convergence of opinion between observers whose initial opinions differ. The observers must agree on the method of sampling, and, of course, they must also agree on the observations that result from this sampling. Under very weak restrictions, no matter how much their initial opinions differ, they can be brought arbitrarily close to convergence on the basis of a sufficient number of sampled observations. The obvious requirement is that the individual observations be approximately independent. If, for example, the observations are strongly dependent, then many observations will count for no more than a single observation.

Reflection upon the conditions under which convergence of beliefs will take place also throws light on the many situations in which no such convergence occurs. The radically differing opinions of men about religion, economics, and

politics are excellent examples of areas in which there is a lack of convergence; no doubt a main source of this divergence is the lack of agreement on what is to count as evidence.

Problem of the forcing character of information

As already indicated, it is an important aspect of the subjective theory to emphasize that equally reasonable men may hold rather different views about the probability of the same event. But the ordinary use of the word "rational" seems to go beyond what is envisaged in the subjective theory of probability. Let us consider one or two examples of how this difference in usage may be expressed.

The first kind of example deals with the nonverbal processing of information by different individuals. One man is consistently more successful than another in predicting tomorrow's weather. At least before the advent of powerful mathematical methods of predicting weather, which are now just beginning to be a serious forecasting instrument, it was the common observation of experienced meteorologists that there was a great difference in the ability of meteorologists with similar training and background and with the same set of observations in front of them to predict successfully tomorrow's weather in a given part of the world. As far as I can see, in terms of the standard subjective theory as expressed, for example, by de Finetti, there is no very clear way of stating that on a single occasion the better predictor is in some sense more rational in his processing of information than the other man; yet in common usage we would be very inclined to say this. It is a stock episode in novels, and a common experience in real life for many people, to denigrate the intelligence or rationality of individuals who continue to hold naïve beliefs about other people's behavior in the face of much contrary, even though perhaps subtle, evidence.

But successive predictions can be studied like any other empirical phenomena, and there is a large literature on evaluating the performance of forecasters, an important practical topic in many arenas of experience. Examination of quantitative methods of evaluation of subjective, as well as objective, forecasts lies outside the scope of this chapter. The *Journal of Forecasting* is entirely devoted to the subject. See also, for example, Makridakis *et al.* (1984) and Dawid (1986).

Contrary to the tenor of many of de Finetti's remarks, it seems fair to say that the subjective theory of probability provides necessary but not sufficient conditions of rationality.

Bayesian probabilities and the problem of concept formation

An important point revolving around the notion of *mistaken* belief is involved in the analysis of how information is processed. In common usage, a belief is

often said to be mistaken or irrational when later information shows the belief to be false. According to the subjective theory of probability, and much sensible common usage in addition, this view of mistaken beliefs is itself a mistake. A belief is not shown to be mistaken on the basis of subsequent evidence not available at the time the belief was held. Proper changes in belief are reflected in the change from a prior to a posterior probability on the basis of new information. The important point for subjective theory is that the overall probability measure does not itself change, but rather we pass from a prior to a posterior conditional probability. Applications of Bayes' theorem illustrate this well enough. The following quotation from de Finetti (1937/1964, page 146) illustrates this point beautifully.

Whatever be the influence of observation on predictions of the future, it never implies and never signifies that we *correct* the primitive evaluation of the probability $P(E_{n+1})$ after it has been *disproved* by experience and substitute for it another $P^*(E_{n+1})$ which conforms to that experience and is therefore probably *closer to the real probability*; on the contrary, it manifests itself solely in the sense that when experience teaches us the result A on the first n trials, our judgment will be expressed by the probability $P(E_{n+1})$ no longer, but by the probability $P(E_{n+1} | A)$, i.e., that which our initial opinion would already attribute to the event E_{n+1} considered as conditioned on the outcome A . Nothing of this initial opinion is repudiated or corrected; it is not the function P which has been modified (replaced by another P^*), but rather the argument E_{n+1} which has been replaced by $E_{n+1} | A$, and this is just to remain faithful to our original opinion (as manifested in the choice of the function P) and coherent in our judgment that our predictions vary when a change takes place in the known circumstances.

In spite of the appeal of what de Finetti says, there seems to be a wide class of cases in which the principles he affirms have dubious application. I have in mind all those cases in which a genuinely new concept is brought to bear on a subject. I do not mean necessarily the creation of a new scientific concept, but rather any situation in which an individual suddenly becomes aware of a concept that he was not previously using in his analysis of the data in front of him.

Suppose an organism has the sensory capability to recognize at least 100 features, but does not know how to combine the features to form the concepts being forced upon it by experience. Assuming the features have only two values (presence or absence), even with this drastic simplification it does not make sense from a computational standpoint to suppose the organism has a prior distribution that is positive for each of the 2^{100} possible patterns that might be nature's choice.

I am not entirely certain what subjectivists like de Finetti would say about this kind of example. I cannot recall reading anywhere a systematic discussion of concept formation, or even identification, by one of the main proponents

of subjective probability. It is my own view that no adequate account of concept formation can be given within the framework of subjective probability, and that additional more complicated and detailed learning processes in organisms must be assumed in order to provide an adequate account. This is not to denigrate the theory of subjective probability, but to be realistic about its limitations.

Problem of unknown probabilities

Another feature of the subjective theory of probability that is in conflict with common usage of probability notions is the view that there are no unknown probabilities. If someone asks me what is the probability of rain in the Fiji Islands tomorrow, my natural inclination is to say, "I don't know," rather than to try to give a probability estimate. If another person asks me what I think the probability is that Stanford University will have an enrollment of at least 50 000 students 500 years from now, I am naturally inclined simply to say, "I haven't the faintest idea what the probability or likelihood of this event is." De Finetti insists on the point that a person always has an opinion, and, therefore, a probability estimate about such matters, but it seems to me that there is no inherent necessity of such a view. It is easy to see one source of it. The requirement that one always have a probability estimate of any event, no matter how poor one's information about the circumstances in which that event might occur may be, arises from a direct extension of two-valued logic. Any statement is either true or false, and, correspondingly, any statement or event must always have a definite probability for each person interrogated. From a formal standpoint it would seem awkward to have a logic consisting of any real number between 0 and 1, together with the quite disparate value, "I don't know."

A little later we shall examine the view that one can always elicit a subjective probability for events about which the individual has very little background information by asking what sort of bet he will make concerning the event. Without anticipating that discussion, I still would like to insist that it does not really seem to be a proper part of the subjective theory to require an assignment of a probability to every imaginable event. In the same spirit with which we reply to a question about the truth of a statement by saying that we simply don't know, we may also reply in the same fashion to a request for an estimate of a probability. This remark naturally leads to the next problem I want to consider.

Inexact probability estimates

There are many reasons, some of which were just mentioned, for being skeptical of one's own or other people's ability to make sensible exact probability

estimates of possible events about which little is known. A retreat, but within the general subjective framework, is to give upper and lower probability estimates. So, in what we might think of as a state of nearly total ignorance about the possible occurrence of an event E , we assign upper probability $P^*(E) = 1$, and lower probability $P_*(E) = 0$. Note that the positions of the stars distinguishes in a natural way upper from lower probabilities.

Development of this idea is pursued in some detail in Chapter 4 by Glenn Shafer. There is, however, one important concept I want to mention here. The use of upper and lower probabilities developed by Dempster and Shafer, for example, assumes a supporting probability, i.e. the upper and lower probabilities assigned to an algebra of events are consistent with the existence of at least one probability measure P such that for every event E in the algebra

$$P_*(E) \leq P(E) \leq P^*(E).$$

But it is easy to envisage realistic incoherent inexact probabilities, incoherent in the sense that they are not compatible with the existence of a probability measure. As a simple hypothetical example, consider the person whose partial beliefs about the economy of his country ten years from now are expressed in part by the following correlations. Let u = high unemployment, p = at least moderate prosperity, and d = at least fairly high deficit. Let these three events be represented by random variables U , P and D respectively, with value $+1$ for occurrence and -1 otherwise, let the subjective expectations of all three be 0 , and let the subjective correlations satisfy the three inequalities $\rho(U, P) < -0.5$, $\rho(D, P) < -0.5$ and $\rho(U, D) < 0.0$. I think subjective correlations of this sort are not unlikely for many triples of events. But then there can be no upper and lower probability of the kind envisaged by Dempster and Shafer to express these beliefs, for there is no possible joint probability distribution of the three random variables satisfying the expectations and the correlation inequalities—here I am assuming, for simplicity, exact subjective probabilities for the marginal distribution of each of the three pairs of random variables.

On the other hand, there can be a nonmonotonic upper probability compatible with any pairwise distribution satisfying the correlation inequalities. Such an upper probability P^* satisfies for any two events A and B such that $A \cap B \neq \emptyset$

$$P^*(A \cup B) \leq P^*(A) + P^*(B),$$

with, of course,

$$P^*(\Omega) = 1 \quad \text{and} \quad P^*(\emptyset) = 0.$$

But the necessary nonmonotonicity means there are events A and B such that $A \subseteq B$, but

$$P^*(B) < P^*(A).$$

which is not possible for any probability measure.

It seems desirable to extend the theory of subjective probability to situations in which it is natural to start with an incoherent or inexact prior because of lack of knowledge—or the opposite problem of too much—with accompanying computational problems. (For a natural application of such nonmonotonic upper probabilities to physics, see Suppes and Zanotti, 1991.)

Decisions and the measurement of subjective probability

It is commonplace to remark that a man's actions or decisions, and not his words, are the true mark of his beliefs. As a reflection of this commonly accepted principle, there has been considerable discussion of how one may measure subjective probabilities on the basis of decisions actually made. This is a complex subject, and I shall not attempt to give it a fully detailed treatment.

The classical response in terms of subjective probability is that we may find out the subjective probabilities a man truly holds by asking him to place wagers. For example, if he thinks the probability of rain tomorrow is really $\frac{1}{2}$, then he will be willing to place an even-money bet on this occurrence. If he thinks that the probability of snow tomorrow has a subjective probability of 0.1, then he will bet against snow at odds of 1 : 9. It is also clear how this same procedure may be used to test precise statements. For example, if a man says the probability of rain tomorrow is at least $\frac{1}{2}$, then presumably he will accept any bet that provides odds at least this favorable to him.

Unfortunately, there is a central difficulty with this method of measuring subjective probability. Almost all people will change the odds at which they will accept a bet if the amount of money varies. For example, the man who will accept an even-money bet on the probability of rain tomorrow with the understanding that he wins the bet if in fact it does rain, will not accept an even-money bet if the amount of money involved moves from a dollar on each side to a hundred dollars on each side. Many people who will casually bet a dollar will not in the same circumstances and at the same odds be willing to bet a hundred dollars, and certainly not a thousand dollars. The man who will accept an even-money bet on its raining tomorrow will perhaps be willing to accept odds of two to one in his favor only if the wager is of the order of a hundred dollars, while he will accept still more favorable odds for a bet involving a larger sum of money. What then are we to say is his true estimate of the probability of rain tomorrow if we use this method of wagers to make the bet?

In spite of this criticism, there have been a number of interesting empirical studies of the measurement of subjective probability using the simple scheme we have just described. A review of the older literature is to be found in Luce & Suppes (1965). An excellent review of recent literature on risk-sensitive models, together with new results, are to be found in Luce & Fishburn (1991).

ACKNOWLEDGMENT

Dennis Lindley suggested several clarifying improvements in an earlier draft.

REFERENCES

- Aczél, J. (1961) Über die Begründung der Additions- und Multiplikationsformeln von bedingten Warscheinlichkeiten. *Magyar Tud. Akad. Mat. Kutató Int Közl.* 6, 110–22.
- Aczél, J. (1966) *Lectures on Functional Equations and Their Applications*. Academic Press, New York.
- Dawid, A.P. (1986) Probability forecasting. In S. Kotz, N.L. Johnson, and C.B. Read (eds.), *Encyclopedia of Statistical Sciences*, Vol. 7. Wiley, New York, pages 210–18.
- deFinetti, B. (1937) La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré*, 7, 1–68. Translated into English in H.E. Kyburg, Jr., & H.E. Smokler (eds.), *Studies in Subjective Probability*, 1964, pages 93–158.
- de Finetti, B. (1974) *Theory of Probability*, Vol. 1. Wiley, New York.
- de Finetti, B. (1975) *Theory of Probability*, Vol. 2. Wiley, New York.
- Domotor, Z. (1969) Probabilistic Relational Structures and Their Applications. Technical Report No. 144, Stanford University, Institute for Mathematical Studies in the Social Sciences.
- Koopman, B.O. (1940a) The axioms and algebra of intuitive probability. *Annals of Math.* 42, 269–92.
- Koopman, B.O. (1940b) The bases of probability. *Bulletin of the American Mathematical Society*. 46, 763–74. Reprinted in H.E. Kyburg, Jr., & H.E. Smokler (eds.), *Studies in Subjective Probability*. Wiley, New York, pages 159–72.
- Koopman, B.O. (1941) Intuitive probability and sequences. *Ann. of Math.* 42, 169–87.
- Kraft, C.H., Pratt, J.W., & Seidenberg, A. (1959) Intuitive probability on finite sets. *Ann. Math. Statist.* 30, 408–19.
- Kranz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971) *Foundations of Measurement*, Vol. I. Academic Press, New York.
- Kyburg, H.E. Jr. & Smokler, H.E. (eds) (1964) *Studies in Subjective Probability*, Wiley, New York.
- Luce, D. (1968) On the numerical representation of qualitative conditional probability. *Ann. Math. Statist.* 39, 481–91.
- Luce, D. & Fishburn, P. (1991) Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty*. 4, 29–59.

- Luce, R.D. & Suppes, P. (1965) Preference, utility and subjective probability. In R.D. Luce, R.R. Bush and E. Galanter (eds.), *Handbook of Mathematical Psychology*. Vol. 3, Wiley, New York, pages 249–410.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1984). *The Forecasting Accuracy of Major Time Series Methods*. Wiley, Chichester.
- Savage, L.J. (1954) *The Foundations of statistics*, Wiley, New York.
- Scottt, D. (1964) Measurement models and linear inequalities. *Journal of Mathematical Psychology*, 1, 233–47.
- Suppes, P. (1973) New foundations of objective probability: axioms of propensities. In P. Suppes, L. Henkin, G.C. Moisil, & A. Joja (eds.), *Logic, Methodology and Philosophy of Science IV*, Proceedings of the 4th International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971. Amsterdam: North Holland, pages 515–29.
- Suppes, P. & Zanotti, M. (1976) Necessary and sufficient conditions for existence of a unique measure strictly agreeing with a qualitative probability ordering. *Journal of Philosophical Logic*, 5, 431–8.
- Suppes, P. & Zanotti, M. (1982) Necessary and sufficient qualitative axioms for conditional probability. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*. 60, 163–9.
- Suppes, P. & Zanotti, M. (1991) Existence of hidden variables having only upper probabilities. *Foundations of Physics*. 12, 1479–99.