The Annals of Statistics
1976, Vol. 4, No. 6, 1051–1064

COHERENT PREFERENCES[1]

BY ROBERT J. BUEHLER

University of Minnesota

De Finetti has defined coherent previsions and coherent probabilities, and others have described concepts of coherent actions or coherent decisions. Here we consider a related concept of coherent preferences. Willingness to accept one side of a bet is an example of a preference. A set of preferences is called incoherent if reversal of some subset yields a uniform increase in utility, as with a sure win for a collection of bets. In both probability and statistical models (where preferences are conditional on data) separating hyperplane theorems show that coherence implies existence of a probability measure from which the preferences could have been inferred. Relationships to confidence intervals and to decision theory are indicated. No single definition of coherence is given which covers all cases of interest. The various cases distinguish between probability and statistical models and between finite and infinite spaces. No satisfactory theory is given for continuous statistical models.

1. Introduction.

In this section we will indicate the relationship of the present paper to earlier theories of coherence.

1.1. *Coherence in the sense of de Finetti.* In de Finetti's (1974) notation X denotes a "random quantity." (The term "random variable" would wrongly imply satisfaction of the Kolmogorov axioms.) The *prevision* of X, denoted by $P(X)$ or $\bar{x}$, is a value such that the subject prefers a constant reward $\bar{x} + a$ to a random reward X for any $a > 0$, and prefers X to $\bar{x} - b$ for any $b > 0$. Thus a prevision is essentially a subjective expectation (whose definition however does not depend on probability). Previsions $\bar{x}_1, \cdots, \bar{x}_n$ of random quantities $X_1, \cdots, X_n$ are called *coherent* if there exist no constants $c_1, \cdots, c_n$ such that $Y = \sum c_i(X_i - \bar{x}_i)$ is uniformly negative, that is sup $Y < 0$, where the supremum is taken over all possible outcomes. Here we can regard c_i as a stake, $c_i(X_i - \bar{x}_i)$ as the payoff of a gamble on the random X_i, and sup $Y < 0$ as a sure loss. (Since c_i can be positive or negative, the condition inf $Y > 0$ is of course equivalent.) De Finetti establishes an equivalent condition: For given $k_1 > 0, \cdots, k_n > 0$, there do not exist $x_1^*, \cdots, x_n^*$ such that $\sup(L^* - L) < 0$ where $L = \sum [(X_i - \bar{x}_i)/k_i]^2$, $L^* = \sum [(X_i - x_i^*)/k_i]^2$.

For example suppose there are two possible outcomes and three random quantities with $X_1 = (1, 4)$, $X_2 = (2, 3)$, $X_3 = (5, 1)$. Then it can be shown that $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ are coherent if and only if $(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (1 + 3p, 2 + p, 5 - 4p)$ for some

Received August 15, 1975; revised January 1976.

[1] Supported by National Science Foundation Grant GP-42674.

AMS 1970 *subject classifications.* 60A05, 62A15, 90A10.

Key words and phrases. Coherence, subjective probability, utility, separating hyperplane, finite additivity, conditional inference.

$0 \leq p \leq 1$, that is the vector of previsions lies on the line segment joining (1, 2, 5) and (4, 3, 1).

When X takes only values 0, 1, then X is called a "random event," the prevision of X becomes the probability of X (with X denoting interchangeably a set and its indicator function), and coherent previsions become coherent probabilities. Manipulative rules become consequences of coherence rather than axioms as in Kolmogorov's theory. It is an easy exercise to prove that to be coherent, the probability of the union of two (or finitely many) disjoint events must be the sum of the separate probabilities.

Earlier but less complete accounts of this theory are given in de Finetti (1937; 1949; 1964; 1972, Section 5.9). Similar theories are given by Shimony (1955) and Kemeny (1955).

The present paper shifts the focus from the previsions, defined in terms of preferences, to preferences themselves. A set of previsions will be coherent in de Finetti's sense if and only if the preferences they imply are coherent as defined below. On the other hand it does not seem to be possible to define coherent preferences in terms of coherent previsions.

1.2. *Coherence in statistical models.* By a statistical model we mean one having a parameter space and observations whose distribution depends on the parameter.

Cornfield (1969) refers to coherence of probabilities assigned to parameter values θ after observing data x, where coherence means nonexistence of stakes giving negative expectation for every θ. Cornfield notes that this is analogous to de Finetti's sure loss, but fails to observe that it is in fact weaker since the loss is only guaranteed in the long run, not on every individual trial. It is shown by Cornfield and independently by Freedman and Purves (1969) that coherence of posterior "probabilities" assigned to all possible outcomes implies their agreement with values calculated by Bayes' theorem for some prior distribution. Similar results of Quiring (1972) pertain to cases where probabilities are assigned only to some subsets of θ values. Infinite models are treated by Quiring (1972), Dawid and Stone (1972, 1973) and Pierce (1973). In Section 6 below we give results analogous to those of Cornfield, Freedman and Purves, showing that preferences (rather than probabilities) conditional on data x are coherent if and only if they correspond to posterior preferences calculated according to some prior distribution, not necessarily unique.

Lindley (1971, page 6) speaks of coherent decision making or coherent actions:

> A decision maker whose actions agree with these axioms has variously been described as *rational*, *consistent*, or *coherent*. We shall use the last term because it effectively captures the idea that the basic principle behind the axioms is that our judgements should fit together or cohere.

Later in the same section (pages 13–16) Lindley gives examples of decision procedures which he calls incoherent.

2. A simple example. Let X denote the event that a horse called Xantippe wins a certain race. In de Finetti's theory every subject can attach a unique subjective value $\bar{x}$ for the prevision or probability $P(X)$ whether or not he knows anything about horse racing in general or Xantippe in particular. Suppose we attempt a less extreme theory (as Smith (1961) and Dempster (1968) have) in which partial knowledge is represented by upper and lower probabilities. Suppose our subject, Peter, believes $P(X)$ lies in the range $\frac{1}{4} < P(X) < \frac{3}{4}$, but doesn't know where. Let (α, β) denote a bet which pays α if X occurs and β otherwise. Peter would presumably accept bets $(4, -1)$, $(-1, 4)$ but would reject bets $(2, -1)$, $(-1, 2)$, since the former have positive expectation for all $\frac{1}{4} < P(X) < \frac{3}{4}$ but the latter do not. From our present point of view Peter would not be faulted for any of the individual choices mentioned above. But if the bets $(2, -1)$, $(-1, 2)$ were offered simultaneously we would say his preferences for $(0, 0)$ (that is, no bet) over $(2, -1)$ and for $(0, 0)$ over $(-1, 2)$ are incoherent because by reversing both choices he has a sure win of $+1$ no matter who wins the race.

The above example illustrates the key idea of "preference reversal (PR) incoherence:" a set of preferences will be called PR-incoherent if there exists a subset whose simultaneous reversal guarantees a higher utility. Typical results establish that PR-coherence is equivalent to an ordering calculated from some probability measure.

3. Relationships between bets and preferences. To further clarify the relationship between this paper and certain earlier work it is helpful to understand how bets and preferences are related. Let (α, β) denote a prospect which pays $a > 0$ if event X occurs and $\beta < 0$ otherwise, and let c be a positive scalar. A preference $(c\alpha, c\beta) \succeq (0, 0)$ corresponds to willingness of our subject, Peter, to bet *on* X with odds corresponding to probability $P(X) = \beta/(\alpha + \beta)$. Willingness to accept either side of this bet corresponds to preferences $(c\alpha, c\beta) \succeq (0, 0)$ for positive or negative c, or to indifference between (α, β) and $(0, 0)$.

For three rather than two possible events, say X_1, X_2, X_3, exclusive and exhaustive, a preference might be expressed as $(\alpha_1, \alpha_2, \alpha_3) \succeq (0, 0, 0)$. This preference is equivalent to willingness to accept a "compound bet" which pays α_i if X_i occurs, but it is not equivalent to willingness to accept any combination of "simple bets" (bets on or against X_1, X_2, X_3 separately).

This can be seen geometrically. A preference $(\alpha_1, \alpha_2, \alpha_3) \succeq 0$ is consistent with probabilities $p_i = P(X_i)$ satisfying $\sum \alpha_i p_i \geq 0$, in other words with probability vectors lying on or to one side of an arbitrary straight line in the simplex $p_i \geq 0$, $\sum p_i = 1$. For simple bets the boundary line is parallel to one side of the simplex.

A collection of preferences will be coherent in the sense of Section 5 iff the intersection of the sets to which the probability vectors are restricted is not empty.

The results in Section 6 below would not follow from those of Cornfield (1969), Freedman and Purves (1969), or Quiring (1972) for two reasons: the earlier work (i) relates only to simple bets, and (ii) assumes Peter's willingness to bet either for or against.

4. Mathematical preliminaries. Our main tools are familiar "separating hyperplane" theorems of linear algebra and their extensions. This type of argument has often been used on similar problems [1, page 119; 6; 12, page 90; 16; 19; 23; 25; 26].

DEFINITION 1. For any vector $w = (w_1, \cdots, w_r)'$ we will write

$w \geqq 0$ if w is nonnegative, that is $w_i \geq 0$ all i,
$w > 0$ if w is positive, that is $w_i > 0$ all i,
$w \geq 0$ if w is semipositive, that is $w \geqq 0$ and $w \neq 0$.

The first lemma is given for example by Gale (1960), Theorem 2.10. The bracketed part is a variation used only incidentally in the present paper.

LEMMA 1. *Let M be an m-by-n matrix. Exactly one of the following alternatives holds:* (i) *there exists an n-by-1 vector $v \geq 0$ such that $Mv > 0$ [$Mv \geq 0$]*, or (ii) *there exists an m-by-1 vector $w \geq 0$ [$w > 0$] such that $-w'M \geqq 0$.*

In applications of Lemma 1, the columns of M correspond to de Finetti's random quantities (we may also call them utility vectors or prospects); w is a probability vector in R^m corresponding to a sample space of m outcomes; v gives a weighted combination of the n columns of M. Let $s = 1, \cdots, m$ label the rows and $t = 1, \cdots, n$ the columns. In generalizations either s or t take infinitely many values. Lemma 2 is a variant of the lemma of the same number in Pierce (1973). Our proof is more elementary in not requiring the theory of duality of L_p spaces.

LEMMA 2. *Let (T, B, λ) be a σ-finite measure space and let L_1 and L_∞ denote respectively λ-integrable and bounded functions on T. If $M_s(t) \in L_1$ for $s = 1, \cdots, m$, then exactly one of the following alternatives holds:* (i) *there exists $v(t) \in L_\infty$, $v(t) \geq 0$ for all t, such that*

$$\int v(t) M_s(t)\, d\lambda > 0 \qquad \text{for all} \quad s = 1, \cdots, m\,, \tag{4.1}$$

or (ii) *there exists a vector $w = (w_1, \cdots, w_m)' \geq 0$ such that $\sum w_s M_s(t) \leq 0$* a.e. (λ).

PROOF. We define a convex cone in m-space by $C = \{(a_1, \cdots, a_m) \mid a_s = \int v(t) M_s(t)\, d\lambda,\ v(t) \geq 0,\ v(t) \in L_\infty\}$. Let Q denote the positive orthant of vectors ≥ 0. If (i) holds, then $C \cap Q \neq \emptyset$. Assume (i) is false so that $C \cap Q = \emptyset$. Then there exists a separating hyperplane, whose normal direction we call w, such that $w'q \geq 0$ for all $q \in Q$ (implying $w \in Q$) and $w'c \leq 0$ for all $c \in C$. The last inequality is equivalent to

$$\int v(t) \sum w_s M_s(t)\, d\lambda \leq 0 \qquad \text{for all} \quad v(t) \in L_\infty\,, \qquad v(t) \geq 0\,. \tag{4.2}$$

Define $S_w^\varepsilon = \{t \mid \sum w_s M_s(t) > \varepsilon\}$, $S_w^0 = \lim_{\varepsilon \to 0} S_w^\varepsilon$, and $v_\varepsilon(t) =$ indicator function of S_w^ε. Then

$$0 \geq \int v_\varepsilon(t) \sum w_s M_s(t)\, d\lambda \geq \varepsilon\lambda(S_w^\varepsilon)\,, \tag{4.3}$$

so that $\lambda(S_w^\varepsilon) = 0$ and $\lambda(S_w^0) = \lim \lambda(S_w^\varepsilon) = 0$. Thus when (i) is false (ii) is true, and it is straightforward to show that (i) and (ii) cannot both be true.

Pierce's (1973) Lemma 2 is not-(ii)$'$ implies (i)$'$, where (i)$'$ is obtained from (i) by deleting $v(t) \geq 0$, and (ii)$'$ is obtained from (ii) by changing ≤ 0 a.e. (λ) to $= 0$ a.e. (λ). The equivalent result that not-(i)$'$ implies (ii)$'$ can be obtained from Lemma 2 above since nonexistence of v of either sign satisfying (4.1) implies nonexistence of either $v \geq 0$ or $v \leq 0$ satisfying (4.1), which implies both ≤ 0 and ≥ 0 a.e. (λ) in (ii).

In Lemma 3 both s and t range over arbitrary spaces. Lemma 3 is a restatement in the present notation of Theorem 1$'$ of Heath and Sudderth (1973). (Theorem 1$'$ is a strengthening of Theorem 1 of Heath and Sudderth (1972) in which "$\sum_{i=1}^n c_i f_{t_i}(s) > 0$ for all $s \in S$" is replaced by "$\inf_{s\in S} \sum_{i=1}^n c_i f_{t_i}(s) > 0$" and "or both" is replaced by "but not both." The proof is essentially unchanged.)

LEMMA 3. *Let S and T be sets and let $\{M_t(s) : t \in T\}$ be a family of bounded, real-valued functions defined on S. Exactly one of the following alternatives holds:* (i) *there exist $t_1, \cdots, t_n \in T$ and $v_1 \geq 0, \cdots, v_n \geq 0$ such that*

$$\inf_{s\in S} \sum_{j=1}^n v_j M_{t_j}(s) > 0\,, \tag{4.4}$$

or (ii) *there exists a finitely additive probability w on S such that*

$$E_w M_t = \int M_t(s)\, dw(s) \leq 0 \qquad \textit{for all} \quad t \in T\,. \tag{4.5}$$

In applications w plays the role of a subjective probability. In connection with Lemma 3 we note that de Finetti (1974), page 119, has clearly expressed a preference for finitely additive probabilities, as they seem to arise more naturally in subjective theory than do countable additive probabilities. For an axiomatic system implying existence of a unique finitely additive probability measure see Fishburn (1969). Lemma 3 differs in that w need not be unique.

5. Probability models. We begin with a space S of points s to be thought of as a space of simple (or elementary) outcomes, postponing the assumption of any probability measure until needed in Definition 5.

DEFINITION 2. A prospect $g(s)$ is a bounded real valued function defined on S. It is to be considered as reward of $g(s)$ given to our subject, Peter, when outcome s *is* observed.

It will be seen that as with de Finetti's random quantities, and as with losses in decision theory, the values of g are handled linearly like utilities. As usual the reader has a choice of viewpoints: (i) utilities exist and g is measured in utility units; (ii) g is measured in dollars and the subject's utility function is linear (or nearly so over the range of interest).

DEFINITION 3. $g' \succeq g$ means Peter prefers g' to g or is indifferent. More precisely "$\succeq$" is a preference relation on some specified set with the usual conventions including transitivity and definitions of $\preceq$, $\succ$ and $\prec$.

5.1. *Finite models.*

DEFINITION 4. Let $g_t' \succeq g_t$ be a set of preferences for $t = 1, \cdots, n$. For a finite space $S = \{1, \cdots, m\}$ this set is called preference-reversal-(or PR-) incoherent if there exist $v_1 \geq 0, \cdots, v_n \geq 0$ such that

$$\sum_{t=1}^{n} v_t(g_t'(s) - g_t(s)) < 0 \qquad \text{for} \quad s = 1, \cdots, m\,. \tag{5.1}$$

Here and in the sequel preferences which are not PR-incoherent are called PR-coherent.

It is clear that the t values for which $v_t \neq 0$ single out a subset of preferences such that a weighted combination violates the preferences uniformly in the outcome s. The weights v_t can variously be regarded as stakes for bets or (when normed) as probabilities (v_t is the probability that Peter is granted his choice of g_t' over g_t). The term "preference reversal" is intended to suggest that the subject can do uniformly better by reversing preferences on the subset where $v_t \neq 0$.

DEFINITION 5. A set of preferences $g_t' \succeq g_t$, $t = 1, \cdots, n$, is called w-coherent if there exists a probability vector $w_1, \cdots, w_m$ such that

$$\sum_{s=1}^{m} w_s g_t'(s) \geq \sum_{s=1}^{m} w_s g_t(s) \qquad \text{for} \quad t = 1, \cdots, n\,. \tag{5.2}$$

THEOREM 1. *Preferences are* PR-*coherent if and only if they are* w-*coherent for some* w.

PROOF. Apply Lemma 1 with $M_{st} = g_t(s) - g_t'(s)$.

Since Definitions 4 and 5 are equivalent we can drop the prefixes PR and w and speak simply of coherent preferences.

Alternatives to the formulation using Definition 4 (which seems to us forceful and enlightening) are certain axiomatic treatments, as we now indicate.

Axiom L_2. $g'(s) \geq g(s)$ for all s implies $g' \succeq g$.
Axiom L_3. $g' \succeq g$ implies $g' + h \succeq g + h$ for all h.
Axiom L_4. $g' \succeq g$ implies $vg' \succeq vg$ for any constant $v \geq 0$.
Axiom L_5. $g' \succeq g$ and $h' \succeq h$ imply $g' + h' \succeq g + h$.

In this approach we think of initially being given a finite (or larger) set of preferences, and then of extending by the axioms to some larger set.

COROLLARY 1. *Preferences are* w-*coherent* iff *they satisfy* L_2, L_4, L_5.

PROOF. Clearly w-coherence implies all four axioms. Assume w-incoherence. By Theorem 1 this implies existence of $v_1 \geq 0, \cdots, v_n \geq 0$, such that (5.1) holds. But the summation in (5.1) represents a compounding of L_4 and L_5, and the inequality itself represents a violation of L_2. Thus w-incoherence implies L_2, L_4, L_5 cannot all hold.

Theorem 4.3.1 of Blackwell and Girshick (1954) establishes that for a preference relation defined on all m-dimensional vectors, L_2, L_3 imply w-coherence (an axiom L_1 is also assumed, but is not used in their proof). Since L_2, L_5 clearly imply L_3, we see that L_2, L_4, L_5 imply L_2, L_3. The converse is true but less evident, and in this sense the Blackwell–Girshick result is stronger than Corollary 1.

EXAMPLE 1. Let $g_t' - g_t = h_t$, $h_1 = (1, -2, 0)$, $h_2 = (-2, 1, 0)$. Then $-h_1 - h_2 = (1, 1, 0)$ so that a reversal of preferences improves Peter's lot for $s = 1, 2$ and leaves it unchanged for $s = 3$. The improvement is not uniform in s, and the preferences are in fact coherent. A unique w gives a consistent ordering: $w = (0, 0, 1)'$. An alternative theory is possible in which the above h_1 and h_2 might be called weakly incoherent, the negation, say strict coherence, would correspond to the existence of $w > 0$ (i.e., $w_s > 0$, all s) giving a consistent ordering. The bracketed version of Lemma 1 is then relevant. The distinction is essentially the same as that between "fair" and "strictly fair" in the sense of Shimony (1955) and Kemeny (1955), and between "strict" and "weak" coherence of Quiring (1972, Chapter II). The philosophical choice between the two criteria is clearly linked to one's attitude toward the acceptability of subjective probabilities which equal zero. De Finetti (1972, page 91; 1974, page 87) favors the analog of the (weak) coherence of our Definitions 4 and 5, and gives his reasoning in detail (1974, Section 3.11). See also Kyburg and Smokler (1964), page 11.

5.2. *Infinite models.* Suppose $s = 1, \cdots, m$, $t \in T$, where (T, B, λ) is a measure space. Then $g_t' \geq g_t$ represents an infinity of preferences involving a finite number of alternatives. Assume that $h_t(s) = g_t'(s) - g_t(s)$ is λ integrable (but not necessarily bounded). PR-incoherence can be defined as the existence of a bounded λ-measurable function $v(t) \geq 0$ such that $\int v(t)h_t(s)\,d\lambda < 0$ for all s, and w-coherence can be defined as existence of a probability vector w such that $\sum w_s h_t(s) \geq 0$ a.e. (λ). Lemma 2 shows that PR-coherence is equivalent to w-coherence for some w.

Finally, we may let s and t range over arbitrary spaces S and T. Define preferences $g_t' \geq g_t$, $t \in T$, to be PR-incoherent if there exist $t_1, \cdots, t_n \in T$, $v_1 \geq 0, \cdots, v_n \geq 0$, such that

$$(5.3) \qquad \sup_{s \in S} \textstyle\sum_{i=1}^{n} v_i h_{t_i}(s) < 0\,, \qquad \text{where} \quad h_t(s) = g_t'(s) - g_t(s)$$

and define them to be w-coherent if there exists a finitely additive probability measure w such that $\int h_t(s)\,dw \geq 0$ for all $t \in T$. Then Lemma 3 implies equivalence of PR-coherence and w-coherence for some w.

EXAMPLE 2. If $w = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots)$, $h_1 = (-1, +3, 0, \cdots)$, $h_2 = (0, -1, +3, 0, \cdots)$, etc., then $Eh_j > 0$, so that each h is strictly preferred to 0. Nevertheless $h_1 + 4h_2 + 4^2h_3 + \cdots$ is strictly negative. Thus an infinite combination of fair bets can be unfair (at least when stakes are unbounded), so that any theory of coherence seems to need a condition like the boundedness of stakes in Lemma 2 or the finiteness of number of stakes in Lemma 3.

EXAMPLE 3. Let $S = T = \{1, 2, \cdots\}$, $h_1(1) = 1$, $h_t(1) = 0$ for $t = 2, 3, \cdots$. No matter what the values of $h_t(s)$ for $s \geq 2$, the preferences $h_t \succeq 0$ are coherent because no positive combination gives a negative reward when $s = 1$. A probability vector consistent with these preferences is $(1, 0, 0, \cdots)$.

EXAMPLE 4. $h_1 = (-1, 2, 0, \cdots)$, $h_2 = (0, -1, 2, 0, \cdots)$, etc. For a combination $c_1 h_1 + c_2 h_2 + \cdots$ to be strictly negative we would need $c_1 > 0$, $-2c_1 + c_2 > 0$, $-2c_2 + c_3 > 0$, etc. No finite number of strictly positive c_j will suffice so that the preferences $h_t \succeq 0$ are PR-coherent. Any probability vector with $p_{j+1} \geq \frac{1}{2} p_j$ gives preferences which agree.

EXAMPLE 5. $h_1 = (-2, 1, 0, \cdots)$, $h_2 = (0, -2, 1, 0, \cdots)$, etc. The same argument shows the preferences $h_t \succeq 0$ are coherent. Here, however, the probability measure must attach zero probability to each individual outcome and so be only finitely, not countably, additive.

6. Statistical models. In the statistical model the set S of outcomes s is replaced by a set Θ of parameter values θ. A new ingredient is the sample space $\mathscr{X}$ having points x. In the present section we restrict $\mathscr{X}$ to be countable. The function $p(x; \theta)$ denotes the likelihood, that is a probability law: $P(X = x \mid \theta) = p(x; \theta)$; $\sum_x p(x; \theta) = 1$ for each θ. A prospect now is a reward to Peter of $g(\theta)$ when θ is the true parameter value. Preferences are expressed conditional on observed data x, and we write $h(x; \theta)$ for $g'(\theta) - g(\theta)$ when $g' \succeq g$ given x. A set of preferences then indexed by $t \in T$ defines a set of functions $h_t(x, \theta)$ depending on x and θ according to

$$h_t(x, \theta) = g_t'(\theta) - g_t(\theta) \quad \text{when} \quad x = x_t \qquad (6.1)$$
$$= 0 \quad \text{otherwise.}$$

For fixed θ the expectation of $h_t(x, \theta)$ with respect to x is

$$E_\theta h_t(x, \theta) = h_t(x_t, \theta) p(x_t, \theta) = f_t(\theta), \quad \text{say.} \qquad (6.2)$$

6.1. *Finite T and Θ.* Suppose $\theta = 1, \cdots, m$, $t = 1, \cdots, n$.

DEFINITION 6. A set of preferences $g_t' \succeq g_t$ given $x = x_t$, $t = 1, \cdots, n$, is called PR-incoherent if there exist $v_1 \geq 0, \cdots, v_n \geq 0$ such that for h_t defined by (6.1)

$$E_\theta \textstyle\sum_t v_t h_t(x, \theta) = \sum_t v_t f_t(\theta) < 0 \quad \text{for} \quad \theta = 1, \cdots, m. \qquad (6.3)$$

The interpretation is that preference reversal, with suitable weights, would increase the *expected* reward to Peter for every parameter value.

For prior probability vector $w = (w_1, \cdots, w_m)'$, denote expectation with respect to w by

$$E_w \varphi(\theta) = \textstyle\sum_{\theta=1}^m w_\theta \varphi(\theta). \qquad (6.4)$$

Then the marginal distribution of x is

$$p_w(x) = E_w p(x, \theta) = \textstyle\sum w_\theta p(x, \theta). \qquad (6.5)$$

If we define

$$S_{pw} = \{x \mid p_w(x) = 0\}\,, \tag{6.6}$$

then the posterior density is

$$\begin{aligned} w(\theta \mid x) &= w_\theta\, p(x, \theta)/p_w(x) \qquad \text{if} \quad x \notin S_{pw} \\ &= \text{undefined} \qquad\qquad\quad \text{if} \quad x \in S_{pw}\,. \end{aligned} \tag{6.7}$$

If $x_t \notin S_{pw}$ then the posterior expectation of $h_t(x, \theta)$ given $x = x_t$ is

$$\begin{aligned} E_{w|x_t} h_t(x, \theta) &= (p_w(x_t))^{-1} \textstyle\sum_\theta w_\theta\, p(x_t, \theta) h_t(x_t, \theta) \\ &= (p_w(x_t))^{-1} E_w f_t(\theta)\,. \end{aligned} \tag{6.8}$$

If $x_t \in S_{pw}$ then in a strict sense according to (6.7) conditional expectation is undefined for $x = x_t$. But from the point of view of the subject, Peter, whose prior distribution is w_θ, x_t has probability zero of occurring, so that preferences when $x = x_t$ are irrelevant to him. In keeping with the choice of the "weak coherence" option mentioned at the end of Section 5.1, it seems appropriate to adopt the convention

$$E_{w|x_t} h_t(x, \theta) = 0 \qquad \text{if} \quad x_t \in S_{pw} \tag{6.9}$$

for use with the following definition.

DEFINITION 7. In the statistical model a set of preferences $\{g_t' \succeq g_t$ given $x = x_t$, $t = 1, \cdots, n\}$ is called w-coherent if there exists a (prior) probability vector w such that

$$E_{w|x_t} h_t(x, \theta) \geq 0 \qquad\qquad \text{for} \quad t = 1, \cdots, n\,. \tag{6.10}$$

THEOREM 2. *A set of preferences $\{g_t' \succeq g_t$ given $x = x_t$, $t = 1, \cdots, n\}$, is* PR-*coherent if and only if it is w-coherent for some w.*

PROOF. Assume w-coherence for some w. For Case 1, $x_t \notin S_{pw}$, we see that (6.8) and (6.10) give $E_w f_t(\theta) \geq 0$. For Case 2, $x_t \in S_{pw}$, we have $\sum_\theta w_\theta p(x_t, \theta) = 0$, which implies $w_\theta p(x_t, \theta) = 0$ for all θ. It follows that $E_w f_t = 0$. Thus for either case $E_w f_t(\theta) \geq 0$, and Lemma 1 applies with $M_{st} = -f_t(s)$. Contrariwise, PR-coherence, or nonexistence of the v vector, implies existence of w such that $E_w f_t(\theta) \geq 0$ for all t. In Case 1 we appeal to (6.8) and in Case 2 to the convention (6.9) to deduce (6.10).

If all x_t were the same, say $x_t = x_0$ for all t, then one could alternatively appeal to Theorem 1 to argue that PR-coherence implies existence of a conditional probability, given x_0, consistent with the preferences. Dividing this by the likelihood $p(x_0, \theta)$ and normalizing gives prior measure known to exist by Theorem 2. We emphasize that in Theorem 2 we may have multiple preferences for some x values while for others there may be none at all.

The case studied by Cornfield (1969) and Freedman and Purves (1969), wherein Peter states odds for every subset of θ values given every x, corresponds to preferences $h_t(x_t, \theta) \succeq 0$ and $-h_t(x_t, \theta) \succeq 0$ where for each t, $h_t(x_t, \theta)$ takes only two values (whose ratio is determined by the odds) as a function of θ. The above authors show that this large set of preferences makes the prior measure unique.

6.2. *Some cases where T and Θ are not both finite.* In extending the model of Section 6.1 we continue to restrict $\mathscr{X}$ to be a countable space to avoid difficulties in defining conditional distributions.

Lemma 2 is relevant to the case where Θ remains finite but T does not. Rather than take (T, B, λ) to be an arbitrary measure space for simplicity we take $T = \{1, 2, \cdots\}$ and $\lambda =$ counting measure. Then Lemma 2 asserts that either there exist $w_1 \geq 0, \cdots, w_m \geq 0$ such that

$$\sum_{s=1}^{m} w_s M_s(t) \leq 0 \qquad \text{for } t = 1, 2, \cdots, \tag{6.11}$$

or there exist $v_1, v_2, \cdots$, such that $0 \leq v_t < R < \infty$ for all t, and

$$\sum_{t=1}^{\infty} v_t M_s(t) > 0 \qquad \text{for } s = 1, \cdots, m. \tag{6.12}$$

Definitions 6 and 7 can be extended by taking $n = \infty$ (with the restriction $v_t < R < \infty$), and Theorem 2 continues to hold by the same argument.

Finally Lemma 3 applies to the case where the space Θ is arbitrary. Let w be a finitely additive measure on Θ and let E_w denote expectation with respect to w. We define the posterior measure on Θ by

$$w(d\theta \mid x) = (p(x, \theta)/p_w(x))w(d\theta) \qquad \text{when } p_w(x) \equiv E_w p(x, \theta) \neq 0, \tag{6.13}$$

and denote posterior expectation by $E_{w|x}$. Definition 7 is changed to allow w to be finitely additive and (6.10) holds for all t in an arbitrary set T. Definition 6 of PR-incoherence is altered to read: there exist $t_1, \cdots, t_n \in T$, $v_1 \geq 0, \cdots, v_n \geq 0$ such that

$$\sup_{\theta \in \Theta} \sum_{i=1}^{n} v_i f_{t_i}(\theta) < 0 \tag{6.14}$$

(the definition (6.2) of $f_t(\theta)$ is still used). Lemma 3 then shows that PR-coherence is equivalent to w-coherence for some finitely additive w.

EXAMPLE 6. $\Theta = \{0, \pm 1, \pm 2, \cdots\}$, $p(x; \theta) = \frac{1}{2}$ if $x - \theta = \pm 1$ and $= 0$ otherwise. For a given x_0 the structural probability distribution has mass $\frac{1}{2}$ at each of the two points $x_0 \pm 1$ (Fraser, 1971). Therefore when x_0 is observed the structural probabilist is indifferent between $h_{x_0} = (\cdots 0, -1, 0, +1, 0, \cdots)$ (where the middle zero is in the x_0 position) and the zero vector. The betting scheme proposed by Buehler (1971) corresponds to the infinite linear combination $H_1 = h_1 + h_2 + \cdots$ for which $E_\theta H_1 = -\frac{1}{2}$ for $\theta = 0$ or 1 and $E_\theta H_1 = 0$ otherwise. The analog of the betting scheme proposed by Rubin (1971) is an infinite combination $H_2 = \sum_{-\infty}^{\infty} v_j h_j$ where by choosing $0 < v_{j-1} < v_j < 1$ for $j = 0, \pm 1, \cdots$, we have $E_\theta H_2 = v_{\theta-1} - v_{\theta+1} < 0$ for all θ. Both H_1 and H_2 fail to demonstrate incoherence in the sense of the previous paragraph for two reasons: (i) The linear combination is infinite rather than finite, and (ii) $\sup_\theta E_\theta H_1 = \sup_\theta E_\theta H_2 = 0$ rather than < 0. The set of structural probability preferences $\{h_x \succeq 0, -h_x \succeq 0, x = 0, \pm 1, \pm 2, \cdots\}$ are in fact coherent in the sense of the previous paragraph. This is, however, not a consequence of their derivation but rather of the degeneracy of the model. Any set of preferences $h_t \succeq 0$, is w-coherent, when w assigns probability zero to each θ value, by convention (6.9), because $p_w(x) = 0$ for all x. Examples in which the likelihood has infinite support would not necessarily have this degeneracy.

The difficulties with Example 6 indicate that we have yet to arrive at a suitable theory of coherence for statistical models having arbitrary parameter spaces.

6.3. *Replacing the probability law of x by preferences.* Let Θ and $\mathscr{X}$ be finite spaces. We now replace the probability law $p(x, \theta)$ by a set of preferences conditional on θ. For example $g_0'(x) \succeq g_0(x)$ could mean that given $\theta = \theta_0$ Peter prefers a reward of $g_0'(x)$ to $g_0(x)$. A given probability function $p(\cdot, \theta_0)$ would of course determine all such (conditional on θ_0) preferences. A set of conditional preferences could be incoherent, or could determine a unique conditional probability law, or could be consistent with a number of probability laws. The preference $g_0' \succeq g_0$ can be written $h \succeq 0$ where

$$(6.15) \qquad \begin{aligned} h = h(x, \theta) &= g_0'(x) - g_0(x) \qquad &\text{if} \quad \theta = \theta_0 \\ &= 0 &\text{otherwise.} \end{aligned}$$

Certain beliefs concerning the conditional probability law of x given θ can be represented by a set of preferences $h_t \succeq 0$, $t \in T_1$, where each $h_t = h_t(x, \theta)$ is nonzero only when $\theta = \theta_t$.

Similarly as in the previous sections Peter can have conditional preferences given particular x values, for example, $k_0'(\theta) \succeq k_0(\theta)$ given $x = x_0$. Putting

$$(6.16) \qquad \begin{aligned} h = h(x, \theta) &= k_0'(\theta) - k_0(\theta) \qquad &\text{if} \quad x = x_0 \\ &= 0 &\text{otherwise,} \end{aligned}$$

a set of preferences conditional of x is represented by $h_t \succeq 0$, $t \in T_2$, where each h_t is nonzero only when $x = x_t$.

DEFINITION 8. The conditional preferences $\{h_t \succeq 0, t \in T_1 \cup T_2\}$ are called PR-incoherent if there exist $v_1 \geq 0, \cdots, v_n \geq 0$, $t_1, \cdots, t_n \in T_1 \cup T_2$ such that

$$(6.17) \qquad \sum v_i h_{t_i}(x, \theta) < 0 \qquad \text{for all} \quad x, \theta .$$

We remark that if there exist (x_0, θ_0) such that $\theta_t \neq \theta_0$ for all $t \in T_1$, $x_t \neq x_0$ for all $t \in T_2$, then (6.17) fails for $(x, \theta) = (x_0, \theta_0)$ so that the preferences are necessarily PR-coherent.

DEFINITION 9. Sets T_1 and T_2 of preferences $h_t \succeq 0$ conditional on θ and x respectively are called π-p-coherent if there exists a prior measure $\pi(\theta)$ and a likelihood $p(x, \theta)$ such that

$$(6.18) \qquad \begin{aligned} &E_{x|\theta} h_t(x, \theta) \geq 0 \quad \text{for all} \quad t \in T_1 , \qquad \text{and} \\ &E_{\theta|x} h_t(x, \theta) \geq 0 \quad \text{for all} \quad t \in T_2 . \end{aligned}$$

Zero values of π and p are not precluded, and in (6.18) the following conventions are to be understood: when $\pi(\theta_0) = 0$ and $\theta_t = \theta_0$ then $E_{x|\theta} h_t(x, \theta) = 0$; when $\sum_\theta \pi(\theta)p(x_0, \theta) = 0$ and $x_t = x_0$ then $E_{\theta|x} h_t(x, \theta) = 0$.

THEOREM 3. *Conditional preferences are* PR-*coherent if and only if they are π-p-coherent for some prior $\pi(\theta)$ and some likelihood $p(x, \theta)$.*

PROOF. If the preferences are PR-coherent then Lemma 1 implies the existence of $w(x, \theta) \geq 0$ (not identically zero) such that

$$\sum_x \sum_\theta w(x, \theta) h_t(x, \theta) \geq 0 \qquad \text{for all} \quad t \in T_1 \cup T_2 . \tag{6.19}$$

We may assume w to be normalized to a probability measure on $\mathscr{X} \times \Theta$, and define $\pi(\theta) = \sum_x w(x, \theta)$, $p(x, \theta) = w(x, \theta)/\pi(\theta)$ when $\pi(\theta) \neq 0$, and $p(x, \theta)$ is arbitrary when $\pi(\theta) = 0$. For $t \in T_1$, $E_{x|\theta} h_t(x, \theta) = 0$ unless $\theta = \theta_t$. When $\theta = \theta_t$ and $\pi(\theta_t) = 0$, then $E_{x|\theta} h_t(x, \theta) = 0$ by convention. When $\theta = \theta_t$ and $\pi(\theta_t) \neq 0$ then

$$\begin{aligned} E_{x|\theta} h_t(x, \theta) &= (\pi(\theta_t))^{-1} \sum_x h_t(x, \theta) w(x, \theta) \\ &= (\pi(\theta_t))^{-1} \sum_x \sum_\theta h_t(x, \theta) w(x, \theta) \\ &\geq 0 , \quad \text{by (6.19)}. \end{aligned}$$

The proof for $t \in T_2$ is similar, as is the converse.

7. Some relationships to theory of confidence intervals. Cornfield (1969) stressed connections between coherence and confidence intervals, but the relationship was slightly strained because of the restriction to finite parameter spaces and the requirement that confidence levels be assigned to every point in the parameter space (rather than fixing a confidence level and choosing a confidence set).

In a typical continuous model let A denote a subset of $\mathscr{X} \times \Theta$, and let $A_x = \{\theta \mid (x, \theta) \in A\}$. Then A_x are confidence sets with confidence level γ if $P_\theta\{\theta \in A_x\} = \gamma$ for all θ. The confidence statement corresponds to an infinite set of pairs of conditional preferences:

$$\begin{aligned} &(1 - \gamma \ \text{ if } \ \theta \in A_x, \ \ -\gamma \ \text{ otherwise}) \succeq 0 , \qquad \text{and} \\ &(-(1 - \gamma) \ \text{ if } \ \theta \in A_x, \ \ \gamma \ \text{ otherwise}) \succeq 0 . \end{aligned} \tag{7.1}$$

(To represent "conservative" confidence sets, write $P_\theta\{\theta \in A_x\} \geq \gamma$ and delete the second line of (7.1).) It is by no means clear how best to extend the preceding theory. To restrict to finite sets of preferences as in Lemma 3 would seem inadequate because any finite set occurs with probability zero.

If we do allow reversal of a continuum of preferences, then we can relate to known conditional properties of confidence intervals. A subset C of $\mathscr{X}$ is a positively biased relevant subset in the sense of Buehler (1959) if for some $\varepsilon > 0$, $P_\theta(A \mid C) \geq \gamma + \varepsilon$ for all θ. If we choose the second preference of (7.1) and reverse it, but only in C (ignoring the complement C'), then the expected payoff after reversal is $(1 - \gamma)P_\theta(AC) - \gamma P_\theta(A'C) = [P_\theta(A \mid C) - \gamma]P_\theta(C) \geq \varepsilon P_\theta(C)$. If

$P_\theta(C) > \varepsilon' > 0$ for all θ, then we have (informally speaking) a very strong incoherence with a uniformly positive increase in the exptected payoff by preference reversal. A weaker form results if $P_\theta(C)$ is not bounded away from zero as in fact happens in the Student t example of Buehler and Feddersen (1963). More generally of course we might consider preference reversals incorporating suitable weight functions $v(x)$. There is no question that $v(x)$ must be restricted (recall Example 2), but it is not at all clear what restrictions are most suitable, or precisely how coherence should be defined.

8. Some relationships to decision theory. Let $L(\theta, d)$ be a loss function and δ be a decision function. Finite models will serve for the present discussion with risk function $R(\theta, \delta) = \sum_x L(\theta, \delta(x))p(x, \theta)$. Let $L^* = -L$, $R^* = -R$ so that the preference is for the larger L^* or R^*. The difference of risk functions is

$$R^*(\theta, \delta') - R^*(\theta, \delta) = \sum_x h_x(\theta)p(x, \theta) \tag{8.1}$$

where

$$h_x(\theta) = L^*(\theta, \delta'(x)) - L^*(\theta, \delta(x)) \,. \tag{8.2}$$

A preference $h_{x_t} \geq 0$ corresponds to a preference for δ' over δ conditional on $x = x_t$. This type of conditional preference does not arise in non-Bayesian decision theory where only average like (8.1) over the sample space are considered.

Suppose that δ_w is a Bayes solution corresponding to prior measure w, and δ is any other decision function. Then for each $x \in \mathscr{X}$

$$\sum_\theta \{L^*(\theta, \delta_w(x)) - L^*(\theta, \delta(x))\}p(x, \theta)w(\theta) \geq 0 \,. \tag{8.3}$$

This set of inequalities tells us that the set of preferences $\{h_x \geq 0, x \in \mathscr{X}\}$, where h_x is defined by (8.2) with $\delta' = \delta_w$, is w-coherent. By Theorem 2 we see that if δ_w is Bayes, the conditional preferences for δ_w over any other δ are PR-coherent.

Next suppose that δ' dominates δ, that is, $R^*(\theta, \delta') \geq R^*(\theta, \delta)$ for all θ, or equivalently $\sum_x h_x(\theta)p(x, \theta) \geq 0$ for all θ. It does not follow that the preferences $\{h_x \geq 0, x \in \mathscr{X}\}$ are coherent. To see this let $x = 1, 2$, $\theta = 1, 2$, $f_x(\theta) = h_x(\theta)p(x, \theta)$. We could arrange $f_1(1) = f_1(2) = -1$, $f_2(1) = f_2(2) = 2$. The evident incoherence suggests defining $\delta''(1) = \delta(1)$, $\delta''(2) = \delta'(2)$, thereby obtaining a δ'' which dominates δ'.

A referee has called attention to similarity between the expectation in (6.2) and the quantities $p_{ij}^{(\delta)}$ defined by Lindley (1953). These $p_{ij}^{(\delta)}$ represent the probability of making decision i using decision function δ when $\theta = j$. It is assumed that decision i is the only correct one when $\theta = i$. Let $L(\theta, d)$ be the corresponding zero-one loss function and define $f_x^\delta(\theta) = L(\theta, \delta(x))p(x, \theta)$. We then find for $i \neq j$, $p_{ij}^{(\delta)} = \sum_{x \in A} f_x^\delta(j)$, where $A = \{x \mid \delta(x) = i\}$, but we have been unable to use this relationship to establish further links with Lindley's theory.

9. Acknowledgment. We wish to thank the referees for several helpful suggestions.

REFERENCES

[1] BLACKWELL, DAVID and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[2] BUEHLER, ROBERT J. (1959). Some validity criteria for statistical inferences. *Ann. Math. Statist.* **30** 845–863.

[3] BUEHLER, R. J. (1971). Measuring information and uncertainty. In *Foundations of Statistical Inference* (Godambe and Sprott, eds.) 330–341.

[4] BUEHLER, R. J. and FEDDERSEN, A. P. (1963). Note on a conditional property of Student's *t*. *Ann. Math. Statist.* **34** 1098–1100.

[5] CORNFIELD, J. (1969). The Bayesian outlook and its applications. *Biometrics* **25** 617–658.

[6] DAWID, A. P. and STONE, M. (1972). Expectation consistency of inverse probability distributions. *Biometrika* **59** 486–489.

[7] DAWID, A. P. and STONE, M. (1973). Expectation consistency and generalized Bayes inference. *Ann. Statist.* **1** 478–485.

[8] DE FINETTI, BRUNO (1937). La Prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré Sect. B* **7** 1–68.

[9] DE FINETTI, BRUNO (1949). Sull' impostazione assiomatrica del calcolo delle probabilità. *An. Treistini dell' Università di Trieste* **19** 29–81.

[10] DE FINETTI, BRUNO (1964). Foresight: Its logical laws, its subjective sources. In *Studies in Subjective Probability* (Kyburg and Smokler, eds.) 93–158.

[11] DE FINETTI, BRUNO (1972). *Probability, Induction and Statistics*. Wiley, New York.

[12] DE FINETTI, BRUNO (1974). *Theory of Probability*, **1**. Wiley, New York.

[13] DEMPSTER, A. P. (1968). A generalization of Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **30** 205–247.

[14] FISHBURN, PETER C. (1969). A general theory of subjective probabilities and expected utilities. *Ann. Math. Statist.* **40** 1419–1429.

[15] FRASER, D. A. S. (1971). Events, information processing, and the structured model. In *Foundations of Statistical Inference*. (Godambe and Sprott, eds.) 32–55.

[16] FREEDMAN, DAVID A. and PURVES, ROGER A. (1969). Bayes' method for bookies. *Ann. Math. Statist.* **40** 1177–1186.

[17] GALE, DAVID (1960). *The Theory of Linear Economic Models*. McGraw-Hill, New York.

[18] GODAMBE, V. P. and SPROTT, D. A. (1971). *Foundations of Statistical Inference*. Holt, Rinehart and Winston, Toronto.

[19] HEATH, DAVID C. and SUDDERTH, WILLIAM D. (1972). On a theorem of de Finetti, oddsmaking and game theory. *Ann. Math. Statist.* **43** 2072–2077.

[20] HEATH, DAVID C. and SUDDERTH, WILLIAM D. (1973). Private communication.

[21] KEMENY, JOHN G. (1955). Fair bets and inductive probabilities. *J. Symbolic Logic* **20** 263–273.

[22] KYBURG, H. E., JR., and SMOKLER, H. E. (1964). *Studies in Subjective Probability*. Wiley, New York.

[23] LINDLEY, D. V. (1953). Statistical inference. *J. Roy. Statist. Soc. Ser. B* **15** 30–76.

[24] LINDLEY, D. V. (1971). *Bayesian Statistics, a Review*. SIAM Publications, Philadelphia.

[25] PIERCE, DONALD A. (1973). On some difficulties in a frequency theory of inference. *Ann. Statist.* **1** 241–250.

[26] QUIRING, JOHN N. (1972). Coherent odds and Bayesian odds. Ph.D. thesis and Technical Report 179, Univ. of Minnesota School of Statistics, Minneapolis.

[27] RUBIN, H. (1971). Discussion of Buehler (1971), in *Foundations of Statistical Inference* (Godambe and Sprott, eds.) 337–341.

[28] SHIMONY, ABNER (1955). Coherence and the axioms of confirmation. *J. Symbolic Logic* **20** 1–28.

[29] SMITH, C. A. B. (1961). Consistency in statistical inference and decision. *J. Roy. Statist. Soc. Ser. B* **23** 1–25.

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455