The Annals of Statistics
1987, Vol. 15, No. 2, 845–864

DE FINETTI'S COHERENCE AND STATISTICAL INFERENCE[1]

By Eugenio Regazzini

Università degli Studi di Milano

Dedicated to the memory of Bruno de Finetti

The main concerns of this paper are the definition of coherent inference (dF-coherent inference) in conformity with a theory of conditional probabilities derived from de Finetti's coherence principle, and a critical comparison of such a definition with one proposed by Heath, Lane and Sudderth.

0. Introduction. Recent articles by Heath and Sudderth (1978), Lane (1981) and Lane and Sudderth (1983, 1984) represent a meaningful attempt to analyze statistical inference in the light of a condition of coherence. The starting point of these works is an extension to infinite spaces of a result of Freedman and Purves (1969) which, in turn, relates to de Finetti's classical coherence condition (*dF-coherence*) well known in the theory of subjective probabilities. On the other hand, it is apparent that the above-mentioned extension does not exactly agree with the "strengthening" of the condition of coherence proposed by de Finetti (1937), Chapter 1 and (1970), Section 16 of the Appendix, in order to define the probability of a conditional event; see also Lane and Sudderth (1985).

The main objectives of the present paper are to analyze the original formulation of de Finetti's condition with respect to its impact on inferential problems and to compare it with a related proposal of Heath, Lane and Sudderth. Section 1 includes a concise summary of the notion of coherence which, accepting de Finetti's suggestion, suitably defines conditional probabilities and conditional previsions. Section 2, after showing that such a condition substantially agrees with the meaning of inference in a Bayesian framework, gives the definition of dF-coherent posterior (Definition 2.1); furthermore, it analyzes the structure of coherent priors and likelihoods (Theorems 2.1 and 2.2) and provides necessary and sufficient conditions in order that a posterior be dF-coherent (Theorem 2.3). Section 3 briefly deals with dF-coherence of the procedures which Bayesian statisticians usually employ in order to construct posteriors. Section 4, after showing that the definition of Heath, Lane and Sudderth amounts to demanding that the involved probabilities be conglomerative with respect to (w.r.t.) some distinguished partitions, introduces a slight modification of that definition in order to frame it into a dF-coherence condition; the resulting notion is denoted by *H-coherence*. Theorem 4.1 may be used to understand the connections between our approach to H-coherence and the original one of Heath, Lane and

Received November 1984; revised September 1986.

[1]Research partially supported by M.P.I. (40% 1983, National Group for Stochastic Models) and by G.N.A.F.A. of C.N.R.

AMS 1980 *subject classifications*. Primary 62A15; secondary 60A05.

Key words and phrases. Conglomerative prevision, dF-coherent conditional prevision, dF-coherent posterior, H-coherent posterior, improper prior, marginalization paradox, probability charge, Stone's strong inconsistency.

Sudderth. Section 5 provides some arguments aimed at showing that dF-coherence is all that theory can prescribe in order to define a "well-behaved" inference. These arguments attempt to meet some of the objections that might be raised with regard to the adequacy of the definition of dF-coherent inference.

When citing references of books or articles originally written in Italian and subsequently translated into English, the year cited is that of the Italian original, whereas the page or section citation refers to the English translation. The recent book by K. P. S. and M. Bhaskara-Rao, on finitely additive set functions, will be referenced throughout by the abbreviation BR(1983).

1. Preliminaries. The present section aims at pointing out some developments of the concept of coherent prevision on a class of conditional bounded random quantities (r.q.'s); de Finetti (1970) provides the basic definitions and results. We will denote the sure event by Ω, the impossible event by $\varnothing$, the negation of an event A by $\bar{A}$ and adopt de Finetti's suggestion that the same symbol be used both for an event and also for its indicator. Whenever Ω is thought of as a set of elementary events, the event $H \neq \varnothing$ as a subset of Ω and the r.q. X as a function from Ω to $\mathbb{R}$, then "the r.q. X conditional on H", $X|H$, coincides with the restriction of X to H. Since $X = X|\Omega$, the theory of conditional previsions includes that of unconditional previsions as a particular case.

According to de Finetti's betting scheme, a person (bookie) who wants to summarize his degree of belief in the different values of a r.q. $X|H$ by a real number, $P(X|H)$, is supposed to be obliged to accept any bet on $X|H$ with gain $S\{P(X|H) - X\}|H$, S being an arbitrary (positive or negative) amount of money; this is the same as saying that the bet is called off if H does not happen. In order to define a prevision on an arbitrary class $\mathscr{X}$ of conditional bounded r.q.'s we will follow Lehman (1955) and Williams (1975). Let P be a map from $\mathscr{X}$ to $\mathbb{R}$ chosen by the bookie on the understanding that, after making the choice, the bookie is committed to accepting any *finite* combination of bets on elements of $\mathscr{X}$ with stakes (positive or negative) arbitrarily chosen by a gambler. Consequently, if the gambler decides to bet on $X_1|H_1, \dots, X_n|H_n \in \mathscr{X}$ with stakes $s_1, \dots, s_n \in \mathbb{R}$, the gain of the bookie is given by

$$\sum_1^n s_K\{P(X_K|H_K) - X_K\}H_K|H_0, \qquad H_0 = \bigcup_1^n H_K.$$

In other words, the values of the gain coincide with the realizations of $\sum_1^n s_K\{P(X_K|H_K) - X_K\}H_K$, which are compatible with the hypothesis that at least one H_K occurs; otherwise, the betting system under consideration will remain without effect. In this framework, the chosen P is said to be coherent if it does not allow betting systems with uniformly negative gains. This leads us to state:

DEFINITION 1.1. Let $\mathscr{X}$ be an arbitrary class of conditional bounded r.q.'s and P be a map from $\mathscr{X}$ to $\mathbb{R}$. Then P is said to be a *prevision* on $\mathscr{X}$ if, for every $\{X_1|H_1, \dots, X_n|H_n\} \subset \mathscr{X}$ and $s_1, \dots s_n \in \mathbb{R}$, $n = 1, 2, \dots,$ the corresponding random gain $G = \sum_1^n s_K\{P(X_K|H_K) - X_K\}H_K$ satisfies the inequalities

$$\text{(1.1)} \qquad \inf G|H_0 \le 0 \le \sup G|H_0,$$

where $\inf G|H_0$ and $\sup G|H_0$ are the infimum and the supremum of the values that $G|H_0$ assumes. If $\mathscr{X}$ includes events only, then a function P satisfying (1.1) is said to be a *probability*.

Being unacquainted with Lehman's and Williams's papers, the present author [cf. Regazzini (1983)] gave a seemingly different definition of conditional coherence, according to which the gain from a finite system of bets is considered conditional on a general event B, so that $X_i|H_i$ and $P(X_i|H_i)$ are replaced by $X_i|H_i \cap B$ and $P(X_i|H_i \cap B)$, respectively. Clearly, B has to be chosen in such a way that $X_i|H_i \cap B$ can belong to $\mathscr{X}$ for every $X_i|H_i$ included in the chosen betting system. The resulting coherence condition coincides with the one expressed by Definition 1.1 according to

THEOREM 1.1. *P is a prevision on $\mathscr{X}$ if, and only if, for all $\{X_1|H_1,\dots,X_n|H_n\} \subset \mathscr{X}$, $s_1,\dots,s_n \in \mathbb{R}$, and B's such that $\{X_1|H_1 \cap B,\dots, X_n|H_n \cap B\} \subset \mathscr{X}$, one obtains*

$$\inf G_B|B \le 0 \le \sup G_B|B,$$

where $G_B = \sum_1^n s_K\{P(X_K|H_K \cap B) - X_K\}H_K$.

Proof of Theorem 1.1 as well as those of the other results of the present section are omitted [see Regazzini (1985), where a condition of coherence, not involving betting systems and equivalent to the ones given above, is provided]. It is easy to prove that, whenever $\mathscr{X}$ includes unconditional r.q.'s only, Definition 1.1 reduces to the usual condition of coherence [cf. de Finetti (1970), Sections 3.3 and 3.4]. Such a condition does not suffice, by itself, in order to extend the rules of the calculus of probability to the case of conditional r.q.'s; in fact, if $\bigcup_1^n H_i \ne \Omega$, then G vanishes when $\bigcup_1^n H_i$ does not happen and, consequently, the usual condition of coherence may hold without further restrictions on P. Hence, Definition 1.1 states an actual strengthening of the classical coherence principle. Here are some consequences of that definition.

THEOREM 1.2. *Let P be a prevision on $\mathscr{X}$. Then*

(i) $X_1|H,\ X_2|H,\ (X_1 + X_2)|H \in \mathscr{X} \Rightarrow P(X_1 + X_2|H) = P(X_1|H) + P(X_2|H)$;

(ii) $X|H \in \mathscr{X},\ a \in \mathbb{R},\ aX|H \in \mathscr{X} \Rightarrow P(aX|H) = aP(X|H)$;

(iii) $X|H \in \mathscr{X} \Rightarrow \inf X|H \le P(X|H) \le \sup X|H$;

(iv) $X \cdot H_1|H_2,\ H_1|H_2,\ X|H_1 \cap H_2 \in \mathscr{X} \Rightarrow P(X \cdot H_1|H_2) = P(X|H_1 \cap H_2) \cdot P(H_1|H_2)$.

If $\mathscr{X}$ is a class of conditional events and P is a probability on $\mathscr{X}$, then

(v) $A|H \in \mathscr{X} \Rightarrow P(A|H) \geq 0$;

(vi) $A|H$, $B|H$, $A \cup B|H \in \mathscr{X}$ *and* $A \cap B = \varnothing \Rightarrow P(A \cup B|H) = P(A|H) + P(B|H)$;

(vii) $A|H \in \mathscr{X}$ *and* $H \subset A \Rightarrow P(A|H) = 1$;

(viii) $A \cap H_1|H_2$, $H_1|H_2$, $A|H_1 \cap H_2 \in \mathscr{X} \Rightarrow P(A \cap H_1|H_2) = P(A|H_1 \cap H_2) \cdot P(H_1|H_2)$.

In other words, conditions (i)–(iv) [(v)–(viii)] are necessary for P to be a prevision (probability). In general, they are not sufficient since, for instance, it may happen that $X_1|H$, $X_2|H \in \mathscr{X}$ and $(X_1 + X_2)|H \notin \mathscr{X}$. On the other hand, such conditions play an important role in the axiomatic theory of finitely additive conditional probabilities [Mazurkiewicz (1932), Hosiasson-Lindenbaum (1940), de Finetti (1949), Section 5.13, Krauss (1968) and Dubins (1975), Section 3]. In fact, the two groups of conditions mentioned above, jointly with the definition of suitable restrictions on $\mathscr{X}$, are generally thought of as axiomatic properties of a conditional expectation and of a conditional probability, respectively. Restrictions on $\mathscr{X}$ must be given so that it can be closed under the operations involved by (i)–(iv) or by (v)–(viii). Obviously, there are many suitable algebraic structures for this purpose; among them we will mention:

(S_1) $\mathscr{X}_1 = \{X|H;\ X \in \mathscr{S},\ H \in \mathscr{H}^0\}$, *where* $\mathscr{S}$ *is a real linear space of bounded r.q.'s containing a nonzero constant and* $\mathscr{H}^0$ *is the class of nonzero elements of the algebra of events* $\mathscr{H}$; *furthermore*, $\mathscr{H}$, *thought of as a class of indicators, is included in* $\mathscr{S}$ *and* $X \cdot H \in \mathscr{S}$ *for all* $X \in \mathscr{S}$ *and* $H \in \mathscr{H}$.

(S_2) $\mathscr{X}_a = \{E|H;\ E \in \mathscr{E},\ H \in \mathscr{H}^0\}$, *where* $\mathscr{E}$ *is an algebra of events including the algebra* $\mathscr{H}$.

The following theorem, stated by Williams (1975), 1.2.2, and, in a less general framework, by Lehman (1955), Theorem 3, points out that previsions on $\mathscr{X}_1$ (probabilities on $\mathscr{X}_a$) may be characterized via the usual axioms of expectations (probability charges).

THEOREM 1.3. *P is a prevision on* $\mathscr{X}_1$ *(a probability on* $\mathscr{X}_a$*) if, and only if, it satisfies* (i)–(iv) [(v)–(viii)] *of Theorem* 1.2. *In such a case, P is said to be a conditional expectation* [*a conditional probability charge*].

Theorem 1.3, jointly with Definition 1.1, is useful for underlining various aspects of the usual axioms by which one defines expectations and/or probability charges [BR (1983), Definition 2.1.1, (7)].

The results we have considered thus far assume that, under more or less general conditions, a prevision exists. We have not yet considered the general problem *of determining whether, given an arbitrary class* $\mathscr{X}$, *there exists at least one P*: $\mathscr{X} \to \mathbb{R}$ *satisfying the conditional coherence principle stated by Definition* 1.1. In order to answer this question we can make use of the following propositions, the first of which is an obvious consequence of Definition 1.1: *if*

$\mathscr{X} = \{X|H\}$, *then there exists at least one prevision on* $\mathscr{X}$. The second provides a basic *extension theorem* proved, in different settings and independently, by several authors [cf. Regazzini (1985)].

THEOREM 1.4. *Let* P_1 *be a prevision on* $\mathscr{X}_1$ *and* $\mathscr{X}_2$ *be a class of bounded r.q.'s properly including* $\mathscr{X}_1$. *Then there exists at least one prevision on* $\mathscr{X}_2$ *which agrees with* P_1 *on* $\mathscr{X}_1$.

Now, given an arbitrary class $\mathscr{X} \ni X|H$, one can assess a prevision on $\{X|H\}$ and, subsequently, extend it to $\mathscr{X}$, preserving coherence, by virtue of Theorem 1.4. Conclusions of this type do not generally hold when one demands that prevision be continuous [cf. de Finetti (1970), Section 18.5 of the Appendix] or, likewise, that probability be σ-additive. However, a prevision P_1 on $\mathscr{X}_1$ generally admits several coherent extensions to $\mathscr{X}_2 \supset \mathscr{X}_1$, possibly unique or more or less arbitrary depending on the structure of $\mathscr{X}_2$ in relation to that of $\mathscr{X}_1$ and on the analytically distinctive features of P_1 and P_2 [cf. de Finetti (1949), Section 5.10 and BR (1983), Chapter 3].

We conclude the present section by recalling that the limit of a sequence of previsions is itself a prevision. Such a result will be used to illustrate a few peculiarities of an inference stated according to Definition 1.1 (cf. Section 3).

THEOREM 1.5. *Let* $\{P_n\}_{n=1}^{\infty}$ *be a sequence of previsions on* $\mathscr{X}$ *and let* $\mathscr{X}^* \subset \mathscr{X}$ *be the set on which* $\lim_{n\to\infty} P_n$ *exists. If* $\mathscr{X}^*$ *is nonempty, then* $P^* = \lim_{n\to\infty} P_n$ *is a prevision on* $\mathscr{X}^*$.

Various of the previous definitions and results establish that σ-additivity of probabilities and continuity of previsions are not necessary conditions for coherence; in fact, they are sufficient, provided some precautions are taken in order that conditional expectations and probabilities defined by Radon–Nikodym derivatives can satisfy conditions (i)–(viii) of Theorem 1.2 [cf. Blackwell and Ryll-Nardzewski (1963), Blackwell and Dubins (1975) and Dubins (1977)].

2. Coherent inferences. The present section aims at analyzing the distinctive features of inferences made in accordance with the theory given in Section 1. First, the meaning of a few widely used symbols will be clarified. $\mathfrak{X}$ and Θ are non-empty sets to be thought of as the set of possible observations and possible states of nature, respectively; $\mathscr{A}_\Theta$ ($\mathscr{A}_\mathfrak{X}$) is a class of subsets of Θ ($\mathfrak{X}$) which represent events relative to the random subject of inference (relative to the experimental results). The set $\Omega \subset \mathfrak{X} \times \Theta$ of all the logically possible couples (x, θ) is thought of as the sure event. Given any subset C of Ω, $C^x = \{\theta: (x, \theta) \in C\}$ and $C_\theta = \{x: (x, \theta) \in C\}$ denote the sections of C at x and at θ, respectively; by $\{x\}$ ($\{\theta\}$) we will denote the event which comes true if, and only if, the experimental results is x (the state of nature is θ). *Henceforth, it is assumed that*: $\{x\} \in \mathscr{A}_\mathfrak{X}$, $\{\theta\} \in \mathscr{A}_\Theta$ *for all* x *and* θ. Finally, by $\mathscr{A}$ we denote a class of subsets of Ω and by $\mathscr{L}$ and $\mathscr{T}$ the classes of conditional events $\{A|\{\theta\};$

$A \in \mathscr{A}_{\mathscr{X}}$, $\theta \in \Theta\}$ and $\{B|\{x\};\ B \in \mathscr{A}_{\Theta},\ x \in \mathscr{X}\}$. According to the Bayesian approach to statistics, an inference takes the form of a probability on $\mathscr{T}$ (*posterior distribution*) which is assessed before knowing the outcome of the experiment. Once the inferrer has become acquainted with the experimental result, say x', he will operate by the restriction of the previous probability to $\{B|\{x'\};\ B \in \mathscr{A}_{\Theta}\}$. Under these circumstances it is desirable that the probability on $\mathscr{T}$ be assessed in such a way that it cannot produce uniformly positive losses, whatever $\{x\}$ may come true. This request leads us to assess inferences according to Definition 1.1 (cf. Theorem 1.1). In order to reduce the arbitrariness, which might be present in a direct assessment of an inference, one generally assumes that inferences are stated after having assigned a probability τ on $\mathscr{A}_{\Theta}$ (*prior distribution*) and a probability P_{θ} on $\mathscr{L}$ (generally expressed through the *likelihood function*). If one wants to avoid sure losses of money, then one must assign P_{θ} and τ in such a way that they produce a probability P on $\mathscr{E} = \mathscr{L} \cup \mathscr{A}_{\Theta}$ in the sense of Definition 1.1. Consequently, inferences will be chosen within the class of probabilities on $\mathscr{T}$ which, besides being coherent by themselves, are coherent relative to P on $\mathscr{E}$ also. All treatments of the Bayesian approach, inspired by Kolmogorov's theory of probability, extend P to a distinguished class $\mathscr{A}$ of subsets of Ω by integrating P_{θ} w.r.t. τ and, subsequently, they evaluate posterior distributions via Radon–Nikodym derivatives of such an extension w.r.t. its restriction to $\mathscr{A}_{\mathscr{X}}$. These standard procedures require restrictions on P and $\mathscr{E}$ which are not necessary in order to fulfill coherence in the sense of Definition 1.1. On the other hand, given a probability P on $\mathscr{E}$ and any class $\mathscr{A}$ of subsets of Ω, one can assert that, in view of Theorem 1.4, there exists at least one probability P^* on $\mathscr{E}^* = \mathscr{A} \cup \mathscr{L}$ such that

$$P^*(A|\{\theta\}) = P_{\theta}(A), \quad \text{if } A|\{\theta\} \in \mathscr{L}, \qquad P^*(C) = \pi(C), \quad \text{if } C \in \mathscr{A},$$

$$P^*(B) = \tau(B) = \pi(\mathscr{X} \times B), \quad \text{if } B \in \mathscr{A}_{\Theta};$$

henceforth it is assumed that $A \times \Theta$, $\mathscr{X} \times B \in \mathscr{A}$ *for all* $A \in \mathscr{A}_{\mathscr{X}}$ *and* $B \in \mathscr{A}_{\Theta}$.

At any rate, since such an extension of P need not be unique, if one decides to assess an inference after having assigned a probability P^* on $\mathscr{E}^*$, then one has to specify a choice of P^*. If P^* is coherent, then at least one coherent inference relative to P^* exists (by virtue of Theorem 1.4); the class of coherent inferences relative to P^* is included in the class of coherent inferences relative to P. This discussion leads us to introduce

DEFINITION 2.1. Let P (P^*) be a probability on $\mathscr{E}$ ($\mathscr{E}^*$). Then the restriction to $\mathscr{T}$ of any coherent extension of P (P^*) to $\mathscr{E} \cup \mathscr{T}$ ($\mathscr{E}^* \cup \mathscr{T}$) is said to be a *dF-coherent posterior relative to* P (P^*).

The abbreviation dF recalls that this approach owes much to de Finetti's theory of probabilities. Definition 2.1 points out that the prior distribution, the posterior and P_{θ} are restrictions of the same probability, whose distinctive feature consists in avoiding sure losses of money. Consequently, the terms prior and posterior are merely conventional and they do not mean that the posterior represents a correction of the corresponding prior. The comments which precede Definition 2.1 emphasize a few statements which move away from the usual treatment of the Bayesian approach (but not, we believe, from the Bayesian approach) and which, consequently, deserve further explanation. In this connection we shall try to answer the following questions:

(1) Is the standard extension from P to P^* coherent in the sense of Definition 1.1?
(2) Is it possible to assign a few simple rules in order to decide whether P is a probability on $\mathscr{E}$?
(3) Is it possible to characterize the class of dF-coherent posteriors relative to P^*?
(4) Are the standard procedures conceived in order to produce Bayesian inferences admissible in the framework outlined by Definition 2.1?

Question (1) makes sense on condition that we impose a few restrictions in order to carry out the involved standard procedure. The following hypotheses are weaker than the classical ones according to which τ is a probability measure and P_θ a transition probability measure [Barra (1981), page 6]:

(C_1) $\mathscr{A}_{\mathscr{X}}$, $\mathscr{A}_\Theta$ and $\mathscr{A}$ are algebras such that $\mathscr{A} \supset \mathscr{A}_{\mathscr{X}} \times \mathscr{A}_\Theta$; $C_\theta \in \mathscr{A}_{\mathscr{X}}$, for all $\theta \in \Theta$ and $C \in \mathscr{A}$; P_θ is a probability charge on $\mathscr{A}_{\mathscr{X}}$ for each $\theta \in \Theta$, such that $P_\theta(A) = 1$ whenever $\{\theta\}$ implies A; τ is a probability charge on $\mathscr{A}_\Theta$.

(C_2) The function $\theta \to P_\theta(A)$ is integrable w.r.t. τ on Θ for all $A \in \mathscr{A}_{\mathscr{X}}$; henceforth, integrals are thought of as Stieltjes integrals in the framework of finitely additive probabilities [cf. BR (1983), Section 4.5].

The following result answers (1) positively.

THEOREM 2.1. *If conditions* (C_1)–(C_2) *are satisfied, then* P_θ *and* τ *generate a probability* P *on* $\mathscr{E}$; *furthermore, the function* P^* *which agrees with* P *on* $\mathscr{E}$ *and with*

$$\pi(C) = \int_\Theta P_\theta(C_\theta)\tau(d\theta), \quad \textit{for all } C \in \mathscr{A} \tag{2.1}$$

is a probability on $\mathscr{E}^*$.

PROOF. Since P is a restriction of P^*, it suffices to show that P^* satisfies (1.1). We will consider two alternative situations: (a) the finite combination of bets includes both events belonging to $\mathscr{L}$ and events belonging to $\mathscr{A}$; (b) the finite combination of bets includes events belonging to $\mathscr{L}$ only.

(a) In this case we have to analyze the sign of

$$G = \sum_{i=1}^{n} s_i\{P_{\theta_i}(A_i) - A_i\}\{\theta_i\} + \sum_{i=1}^{m} s_i'\{\pi(C_i) - C_i\},$$

since $H_0 = \Omega$. In view of (C_1), the constituents $\{w_1, \dots, w_s\}$ of the events involved in G belong to $\mathscr{A}$ and, consequently, their probability is determined by (C_1) and (2.1): $\pi(w_j) \geq 0$, $\Sigma_1^s \pi(w_j) = 1$. If by g_j we denote the value of G when

w_j comes true, then from (C_1) and (2.1) one obtains $0 = \Sigma_1^s g_j \pi(w_j)$ and, consequently, (1.1) comes true.

(b) In this case $G = \Sigma_1^n s_i\{P_{\theta_i}(A_i) - A_i\}\{\theta_i\}$, $H_0 = \cup_1^n \{\theta_i\}$. It suffices to show that the inequalities of Definition 1.1 hold conditionally on $\{t\}$, where $t \in \{\theta_1, \dots, \theta_n\}$. Given $\{t\}$, the realization of G is given by

$$G_t = \sum_1^q s_{i,t}\{P_t(A_{i,t}) - A_{i,t}\},$$

where $A_{i,t}$, $s_{i,t}$, $i = 1, \dots, q$, are the events and the stakes associated with $\{t\}$ in the original betting combination. By denoting the constituents of $\{A_{1,t}, \dots, A_{q,t}\}$ compatible with $\{t\}$ by $w_{1,t}, \dots, w_{r,t}$, and the corresponding values of G_t by $g_{j,t}$, one obtains

$$0 = \sum_{j=1}^r g_{j,t} P_t(w_{j,t}), \quad \text{with } 1 = \sum_{j=1}^r P_t(w_{j,t}).$$

Hence, G_t cannot turn out to be uniformly negative or positive under the hypothesis that H_0 comes true. □

As far as question (2) is concerned, a clear answer can be given, which holds without any restriction on $\mathscr{E}$.

THEOREM 2.2. *Let P be a map from $\mathscr{E}$ to $\mathbb{R}$. Then P is a probability if, and only if, the following conditions hold:*

(i) *$A \to P_\theta(A)$ is a probability on $\mathscr{A}_\mathscr{X}$ for each $\theta \in \Theta$;*
(ii) *$P_\theta(A) = 1$, whenever $\{\theta\}$ implies A;*
(iii) *τ: $\mathscr{A}_\Theta \to \mathbb{R}$ is a probability on $\mathscr{A}_\Theta$.*

PROOF. In view of Theorem 1.2, it suffices to show that (i)–(iii) imply coherence of P. By virtue of (iii) and Theorem 1.4, τ admits an extension $\tilde{\tau}$ to the power set $\tilde{\mathscr{A}}_\Theta$ of Θ; P_θ also can be extended to the algebra $\tilde{\mathscr{A}}_\mathscr{X}$ generated by $\mathscr{A}_\mathscr{X}$, preserving (i)–(ii). Now, the algebra $\tilde{\mathscr{A}}$ generated by $\tilde{\mathscr{A}}_\mathscr{X} \times \tilde{\mathscr{A}}_\Theta$, $\tilde{\tau}$ and $\tilde{P}_\theta$ satisfy (C_1)–(C_2). By defining $\tilde{P}$ on $\tilde{\mathscr{A}} \cup \{A|\{\theta\};\ A \in \tilde{\mathscr{A}}_\mathscr{X},\ \theta \in \Theta\}$ according to (2.1), $\tilde{P}$ is a probability, by virtue of Theorem 2.1, and P is a probability since it is a restriction of $\tilde{P}$. □

As regards question (3), we start by characterizing dF-coherent posteriors relative to P^* when $\mathscr{A}_\mathscr{X}$, $\mathscr{A}_\Theta$ and $\mathscr{A}$ are supposed to be algebras such that $\mathscr{A} \supset \mathscr{A}_\mathscr{X} \times \mathscr{A}_\Theta$. In such a case, conditions (i)–(iii) of Theorem 2.2 reduce to (C_1) and the map $q_x(B)$ from $\mathscr{X} \times \mathscr{A}_\Theta$ to $\mathbb{R}$, in order that it may be a dF-coherent posterior, has to satisfy conditions (v)–(viii) of Theorem 1.2; hence,

(Q_1) *$B \to q_x(B)$ is a probability charge on $\mathscr{A}_\Theta$ for all $x \in \mathscr{X}$;*
(Q_2) *$q_x(B) = 1$, whenever $\{x\}$ implies B;*
(Q_3) *$q_x(B)\pi(\{x\} \times \Theta) = \pi(\{x\} \times B)$ for all $x \in \mathscr{X}$ and $B \in \mathscr{A}_\Theta$.*

Notice that π need not be assessed according to (2.1) and that (Q_1)–(Q_3) do not suffice, in general, to generate a dF-coherent posterior relative to P^*. In fact, if there exist $C = \{\theta_1, \dots, \theta_n\} \in \mathscr{A}_\Theta$ and $D = \{x_1, \dots, x_q\} \in \mathscr{A}_{\mathfrak{X}}$ such that

$$(2.2) \qquad P_{\theta_i}(D) \textit{ and } q_{x_j}(C) > 0 \textit{ for all } i = 1, \dots, n, \ j = 1, \dots, q,$$

then any extension π^* of P^* and q_x has to satisfy the following conditions in which H stands for $C \cup D$:

$$\pi^*(\{x_j\}|H) \geq 0, \qquad \pi^*(\{\theta_i\}|H) \geq 0,$$

$$\pi^*(\{x_j\}|\{\theta_i\} \cap H) = P_{\theta_i}(\{x_j\}), \qquad \pi^*(\{\theta_i\}|\{x_j\} \cap H) = q_{x_j}(\{\theta_i\}),$$

$$(*) \quad \pi^*\big(\{(x_j, \theta_i)\}|H\big) = q_{x_j}(\{\theta_i\})\pi^*(\{x_j\}|H)$$
$$= P_{\theta_i}(\{x_j\})\pi^*(\{\theta_i\}|H), \quad \text{for all } i = 1, \dots, n; \ j = 1, \dots, q,$$

$$1 = \pi^*(H|H) = \sum_{i=1}^{n} \pi^*(\{\theta_i\}|H) + \sum_{j=1}^{q} \{1 - q_{x_j}(C)\}\pi^*(\{x_j\}|H).$$

In other words, coherence requires that there exist assessments ρ_i's of $\pi^*(\{\theta_i\}|H)$, $i = 1, \dots, n$, and assessments σ_j's of $\pi^*(\{x_j\}|H)$, $j = 1, \dots, q$, such that

(Q_4) *if* (2.2) *holds, then the system*

$$\sigma_j \geq 0, \qquad \rho_i \geq 0, \qquad i = 1, \dots, n; \ j = 1, \dots, q,$$

$$(2.3) \qquad q_{x_j}(\{\theta_i\})\sigma_j = P_{\theta_i}(\{x_j\})\rho_i \qquad i = 1, \dots, n; \ j = 1, \dots, q,$$

$$1 = \sum_{i=1}^{n} \rho_i + \sum_{j=1}^{q} \{1 - q_{x_j}(C)\}\sigma_j$$

is consistent.

Condition (Q_4) implies that, for a single H, any solution of (2.3) represents a coherent extension of P^* and q_x to $\mathscr{A}_H = \{\{x_j\}|H, \{\theta_i\}|H, \{(x_j, \theta_i)\}|H; i = 1, \dots, n, j = 1, \dots, q\}$. Reciprocally, it confirms that q_{x_j} can be assessed via the classical Bayes' theorem by replacing the probabilities of $\{x_j\}$ and $\{\theta_i\}$ by the corresponding probabilities conditioned on H. Under the guidance of (Q_4), one may assess an actual coherent extension of P^* and q_x to $\bigcup \mathscr{A}_H$; in that case, we must make sure that for each H, the corresponding system (2.3) admits a solution S_H such that $\bigcup S_H$ is a coherent assessment in the sense of Definition 1.1. The following theorem, by stating that (Q_1)–(Q_4) are sufficient conditions in order that P^* and q_x may generate a probability on $\mathscr{E}^* \cup \mathscr{T}$, establishes, through Theorem 1.4, that solutions of this type do exist. As far as sufficiency of (Q_4) is concerned, it is worth remarking that results of this kind already exist in the literature [cf. Dubins (1975), Lemma 8 and Regazzini (1985), lemma]. The role of (Q_4) as regards the assessment of q_x by standard procedures will be emphasized in the next sections. Clearly, the above-mentioned theorem answers question (3) also.

THEOREM 2.3. *Let $\mathscr{A}_{\mathfrak{X}}$, $\mathscr{A}_{\Theta}$ and $\mathscr{A}$ be algebras such that $\mathscr{A} \supset \mathscr{A}_{\mathfrak{X}} \times \mathscr{A}_{\Theta}$ and P^* be a probability on $\mathscr{E}^*$. Under these circumstances, q_x is a dF-coherent posterior relative to P^* if, and only if, it satisfies* (Q_1)–(Q_4).

PROOF. In view of the previous comments, it suffices to prove that (Q_1)–(Q_4) imply (1.1). Let us first consider combinations of bets including at least one element of $\mathscr{A}$. In such a case, since $H_0 = \Omega$, it is easy to deduce from the coherence of P^* and (Q_1)–(Q_3) that $\pi(G) = 0$; this implies (1.1).

Whenever the combination includes conditional events only, the inferrer's gain is given by

$$G = \sum_1^n s_i\{P_{\theta_i}(A_i) - A_i\}\{\theta_i\} + \sum_1^q t_j\{q_{x_j}(B_j) - B_j\}\{x_j\}, \tag{2.4}$$

and the sign of G has to be analyzed conditionally on $H_0 = \{\theta_1, \dots, \theta_n, x_1, \dots, x_q\}$. If $\pi(H_0) > 0$, then (1.1) is immediately deduced as in the previous case. If $\pi(H_0) = 0$, then we will consider the following alternative situations: (a) either $x \in \{x_1, \dots, x_q\}$ or $\theta \in \{\theta_1, \dots, \theta_n\}$ or both exist such that $q_x(C) = 0$, $p_\theta(D) = 0$; (b) (2.2) holds. In the first case of (a), if one supposes that $\{x\} \cap \bar{C}$ occurs, then (2.4) reduces to a r.q. G^* whose values are those of G which are compatible with $\{x\} \cap \bar{C}$. Since $q_x(\bar{C}) = 1$, any extension to $\{\cdot | \{x\} \cap \bar{C}\}$ has to satisfy $\text{Prob}(B | \{x\} \cap \bar{C}) = q_x(B)$ for all $B \in \mathscr{A}_\Theta$; hence, $q_x(G^*) = 0$ and this implies (1.1). The same conclusion is reached in the second case of (a). When (b) holds, we may choose a solution $\rho_1, \dots, \rho_n, \sigma_1, \dots, \sigma_q$ of (2.3) and, after putting $\pi^*(\{\theta_i\}|H) = \rho_i$, $\pi^*(\{x_j\}|H) = \sigma_j$, assess π^* according to $(*)$. From this we deduce that $\pi^*(G|H) = 0$ and, consequently, that (1.1) holds. □

Theorem 2.3 may be extended to arbitrary classes of events by arguing as in the proof of Theorem 2.2. Such a theorem points out that one can conceive situations in which, given P^*, q_x may be assessed in an almost totally arbitrary manner [e.g., when $p_\theta(\{x\}) = 0$ for all $(x, \theta) \in \Omega$]. Since question (4) involves a different kind of problem, we shall answer it separately in the next section.

3. Construction of dF-coherent posteriors. Since a complete solution of the problem would require an extremely long and technical treatment, we will confine ourselves to outlining the basic steps. This section can be thought of as split up into two parts: The first concerns inferences deduced through Bayes' theorem; the second deals with inferences from *improper priors*.

Given P^* on $\mathscr{E}^*$, the applicability of the standard procedures conceived in order to produce Bayesian inferences depends on the algebraic structure of $\mathscr{E}^*$, $\mathscr{T}$, and on the analytical properties of P^*. If one supposes, for example, that Θ is finite and that $P^*(\{x\}) > 0$, then (Q_1)–(Q_4) imply that there exists a *unique* dF-coherent posterior, determined through the classical Bayes' theorem; the same conclusion holds whenever Θ is countably infinite, $1 = \sum_1^\infty P^*(\{\theta_j\})$ and $P^*(\{x\}) > 0$. In order to deal with more general situations, it is useful to make some remarks about dF-coherent posteriors assessed through Radon–Nikodym derivatives of π with respect to its restriction μ to $\mathscr{A}_{\mathscr{X}}$. In that case, q_x ought to satisfy equation

$$\pi(\{A \times \Theta\} \cap C) = \int_A q_x(C^x)\mu(dx), \quad \text{for all } A \in \mathscr{A}_{\mathscr{X}}, C \in \mathscr{A}; \tag{3.1}$$

clearly, it is supposed that $\mathscr{A}_{\mathscr{X}}$, $\mathscr{A}_\Theta$ and $\mathscr{A}$ are algebras such that $C^x \in \mathscr{A}_\Theta$ for al $C \in \mathscr{A}$ and $x \in \mathscr{X}$. By virtue of (Q_2), this equation is the same as

$$\pi(C) = \int_{\mathscr{X}} q_x(C^x)\mu(dx), \quad \text{for all } C \in \mathscr{A}.$$

Unfortunately, it may happen that "exact derivatives" q_x do not exist for arbitrary P^* [cf. Example 4.3 and BR (1983), Section 6.3]. Maynard (1979) stated necessary and sufficient conditions under which "exact" Radon–Nikodym derivatives exist in the framework of finitely additive set functions. Obviously, these conditions do not imply that any solution of (3.1) satisfies (Q_1)–(Q_4) (cf. Example 4.2); in fact, any solution of (3.1) is a dF-coherent posterior relative to P^* if, and only if, it satisfies (Q_1), (Q_2) and (Q_4). In this connection, one can appreciate the role of (Q_4) in reducing the arbitrariness connected with the choice of a dF-coherent posterior within the set of the solutions of (3.1).

In line with standard practice, we will say that an inference q_x is assigned by *Bayes' theorem* if there exists a nonnegative and integrable function l on $\mathscr{X} \times \Theta$ and vanishing on $\mathscr{X} \times \Theta \setminus \Omega$, such that

$$q_x(B) = \left\{\int_\Theta l(x, \theta)\tau(d\theta)\right\}^{-1} \int_B l(x, \theta)\tau(d\theta) \tag{3.2}$$

satisfies (3.1).

In our approach to statistical inference, two questions arise with regard to this definition. First, under what restrictions on P^* does there exist a function l satisfying these requirements? If this is possible, what about the coherence of the resulting q_x? We tried to answer the first question [Regazzini (1984)] by suggesting a set of sufficient conditions which generalize those commonly assumed in Kolmogorov's theory of probability [cf. Kallianpur and Striebel (1968), Theorem 2.1]. As far as the second question is concerned, one can prove that a posterior from Bayes' theorem is dF-coherent whenever the denominator is not zero. On the other hand, the possible subset A_0 of $\mathscr{X}$ on which the denominator vanishes is μ-null and the posterior has to be defined on it in such a way as to satisfy (Q_1), (Q_2) and (Q_4); in fact, (Q_3) certainly holds [Regazzini (1984), Theorem 2.1]. We take the opportunity to point out a mistake contained in the wording of this theorem, in which q_x is defined on A_0 in a substantially arbitrary manner. We are indebted to a referee for calling our attention to this oversight.

To sum up, we may assert that, given a coherent P^* satisfying suitable regularity conditions, one may assess a dF-coherent posterior according to the algorithms in use among Bayesian statisticians. However, these algorithms do not generally exhaust all the admissible procedures in order to assess dF-coherent posteriors relative to a given P^*.

One of these is connected with the use of improper priors. Let $\mathscr{A}_\Theta$ be a σ-algebra and ρ a *measure* on it such that $\rho(\Theta)$ is infinite and $0 < \int_\Theta l(x,\theta)\rho(d\theta) < \infty$ for all $x \in \mathscr{X}$. It is well known that many statisticians, after replacing τ by ρ in (3.2), consider the corresponding result as a Bayesian inference from an *improper prior*. We will see that, under suitable conditions, this procedure can lead to a dF-coherent posterior w.r.t. a probability P on $\mathscr{E}$. Suppose that there exists a sequence $\{\Theta_n\} \subset \mathscr{A}_\Theta$ such that $\Theta_n \uparrow \Theta$ and $0 < \rho(\Theta_n) < \infty$ for all n and consider the corresponding sequence of probability measures on $(\Theta, \mathscr{A}_\Theta)$:

$$\tau_n(B) = \rho(B \cap \Theta_n)/\rho(\Theta_n), \qquad B \in \mathscr{A}_\Theta.$$

Let $\mathscr{A}_\mathscr{X}$ be a σ-algebra and $\mathscr{A}$ the σ-algebra generated by $\mathscr{A}_\mathscr{X} \times \mathscr{A}_\Theta$. For the sake of simplicity, we will assume that P_θ is a transition probability measure dominated by a σ-finite measure λ; l indicates the density of P_θ w.r.t. λ and $\int_{\Theta_n} l(x,\theta)\rho(d\theta) \in (0, +\infty)$ for all $x \in \mathscr{X}$. In view of the comments about Bayes' theorem, it follows that

$$q_{x,n}(B) = \left\{\int_{\Theta_n} l(x,\theta)\rho(d\theta)\right\}^{-1} \int_{B\cap\Theta_n} l(x,\theta)\rho(d\theta)$$

is a dF-coherent posterior on $\mathscr{A}_\Theta \cap \Theta_n$ relative to τ_n and P_θ. If $0 < \int_\Theta l(x,\theta)\rho(d\theta) < \infty$ for all $x \in \mathscr{X}$, then, a straightforward application of the monotone convergence theorem yields

$$q_x(B) := \lim_{n\to\infty} q_{x,n}(B)$$

$$(3.3)\qquad = \left\{\int_\Theta l(x,\theta)\rho(d\theta)\right\}^{-1} \int_B l(x,\theta)\rho(d\theta), \quad \text{for all } B \in \mathscr{A}_\Theta \text{ and } x \in \mathscr{X}.$$

Setting $\mathscr{G}_\Theta$ for the subclass of $\mathscr{A}_\Theta$ on which $\lim \tau_n$ exists and τ for the limit of $\{\tau_n\}$, by virtue of Theorem 1.5 we may state

> (3.3) *is a dF-coherent posterior on* $\mathscr{A}_\Theta$ *relative to* P^* *where* $P^*(A|\{\theta\}) = P_\theta(A)$ *for all* $A \in \mathscr{A}_\mathscr{X}$ *and* $\theta \in \Theta$, $P^*(B) = \tau(B)$ *for all* $B \in \mathscr{G}_\Theta$.

This statement, however, does not entitle us to believe either that an improper prior can be thought of as a probability on $\mathscr{A}_\Theta$ or that it can be reinterpreted as a coherent probability. It merely says that, under suitable conditions, the right-hand side of (3.3) produces posteriors which are dF-coherent relative to a prior on $\mathscr{A}_\Theta$, which cannot be confused with ρ if $\rho(\Theta) \neq 1$. From a practical viewpoint, given a measure ρ on $\mathscr{A}_\Theta$ with $0 < \int_\Theta l(x,\theta)\rho(d\theta) < \infty$ for all $x \in \mathscr{X}$, if one succeeds in singling out a sequence $\{\Theta_n\} \subset \mathscr{A}_\Theta$, converging to Θ and such that $0 < \rho(\Theta_n) < \infty$ for all $n = 1, 2, \dots$, then the right-hand side of (3.3) represents a dF-coherent posterior relative to P_θ and to a suitable *probability charge on* $\mathscr{A}_\Theta$ *agreeing with* $\lim_{n\to\infty}\rho(B\cap\Theta_n)/\rho(\Theta_n)$ *for all* $B \in \mathscr{A}_\Theta$ *for which the limit exists*. For a different approach to the problems connected with the use of improper priors, see Hartigan (1983).

We conclude the present section by exhibiting an example which shows that dF-coherent posteriors assigned according to the procedure just described need not satisfy the classical theory of conditional probability even if $q_{x,n}$ do so for all n. As a matter of fact, σ-additivity and other properties connected with σ-additivity are not generally preserved in a passage to the limit.

EXAMPLE 3.1 [adapted from Stone and Dawid (1972)]. Let $\mathscr{X} = \Theta = (0, +\infty)^2$; $\mathscr{A}_{\mathscr{X}} = \mathscr{A}_{\Theta}$ = class of Borel subsets of $(0, +\infty)^2$; $\mathscr{A}$ = product σ-algebra of $\mathscr{A}_{\mathscr{X}}$ and $\mathscr{A}_{\Theta}$. Let P_θ be the probability measure on $\mathscr{A}_{\mathscr{X}}$ characterized by the density

$$l(x, \theta) = \theta_1 \theta_2^2 \exp\{-\theta_2(\theta_1 x_1 + x_2)\}, \quad \text{where } x = (x_1, x_2),\ \theta = (\theta_1, \theta_2).$$

Let us set

$$\Theta_n = (0, +\infty) \times (0, n], \qquad \tau_n(B) = \int_{B \cap \Theta_n} \pi(\theta_1)\, d\theta_1\, d\theta_2/n, \quad \text{for all } B \in \mathscr{A}_{\Theta},$$

where $\pi(\theta_1) \geq 0$ and $\int_0^\infty \pi(t)\, dt = 1$.

Resorting to (3.3) one obtains

$$(3.4)\qquad \begin{aligned} q_x(\theta_1 \leq \tilde{\theta}) &= \left\{ \int_0^\infty \frac{\theta_1 \pi(\theta_1)}{(\theta_1 + z)^3}\, d\theta_1 \right\}^{-1} \int_0^{\tilde{\theta}} \frac{\theta_1 \pi(\theta_1)}{(\theta_1 + z)^3}\, d\theta_1, \qquad z = x_2/x_1, \\ P(\theta_1 \leq \tilde{\theta} | z) &= \left\{ \int_0^\infty \frac{\theta_1 \pi(\theta_1)}{(\theta_1 + z)^2}\, d\theta_1 \right\}^{-1} \int_0^{\tilde{\theta}} \frac{\theta_1 \pi(\theta_1)}{(\theta_1 + z)^2}\, d\theta_1 \neq q_x(\theta_1 \leq \tilde{\theta}). \end{aligned}$$

According to the classical theory of conditional expectation, since $q_x(\theta_1 \leq \tilde{\theta}) = \phi(z)$, one would obtain $q_x(\theta_1 \leq \tilde{\theta}) = P(\theta_1 \leq \tilde{\theta}|z)$ a.s. Consequently, (3.4) is seen as a paradox, usually referred to as a *marginalization paradox*. It is closely related to the *failure of conglomerability*.

4. Conglomerability and H-coherence. We have already noticed that Definition 1.1 states a strengthening of the classical coherence principle for unconditional r.q.'s. In the present section we will deal with the strengthening suggested and analyzed by Heath, Lane and Sudderth.

After assigning P^* and q_x according to Definition 2.1, the inferrer's gain from a finite system of bets on events belonging to $\mathscr{E}^* \cup \mathscr{T}$ is given by

$$(4.1)\qquad \begin{aligned} G = {} & \sum_{i=1}^{n} s_i \{P_{\theta_i}(A_i) - A_i\}\{\theta_i\} + \sum_{i=1}^{q} s_i' \{\pi(C_i) - C_i\} \\ & + \sum_{i=1}^{k} s_i'' \{q_{x_i}(B_i) - B_i\}\{x_i\}. \end{aligned}$$

Coherence assures that $G|H_0$ cannot result in a uniformly positive loss. Heath, Lane and Sudderth argue that this condition is not sufficient in order to define a well-behaved inference. In fact, since just one value of the unknown parameter will turn out to be the true one, the prevision of (4.1), evaluated under the hypothesis that $\{\theta\}$ is true, must not be uniformly negative w.r.t. $\theta \in \Theta$. In view of the assumption of Section 2, according to which odds are posted before knowing the experimental result, if one sticks to the point of view expounded above, then one argues that the prevision of (4.1) conditioned on $\{x\}$ must not be uniformly negative w.r.t. $x \in \mathscr{X}$. Since the sign of s_i, s_i', s_i'' is arbitrary, these conditions may be synthesized by the following inequalities:

$$\inf_\theta P_\theta(G) \le 0 \le \sup_\theta P_\theta(G), \qquad \inf_x q_x(G) \le 0 \le \sup_x q_x(G). \tag{4.2}$$

In conformity with the quoted authors, let us now assume that $\mathscr{A}_{\mathscr{X}}$, $\mathscr{A}_\Theta$ are σ-algebras and $\mathscr{A}$ is the corresponding product σ-algebra. From this hypothesis, Theorem 4.7.4 of BR (1983) and Theorem 2.2 of Darst (1961), one deduces that, given P^* on $\mathscr{E}^*$ and any bounded $\mathscr{A}$-measurable function f: $\mathscr{X} \times \Theta \to \mathbb{R}$, the previsions $P_\theta(f)$ and $\pi(f)$ are uniquely determined according to

$$P_\theta(f) = \int_{\Omega_\theta} f_\theta(x) P_\theta(dx), \qquad \pi(f) = \int_\Omega f(u)\pi(du),$$

f_θ being the θ-section of f. On the other hand, in this framework, condition (4.2) is the same as

$$\inf_\theta P_\theta(f) \le \pi(f) \le \sup_\theta P_\theta(f), \qquad \inf_x q_x(f) \le \pi(f) \le \sup_x q_x(f), \tag{4.3}$$

for all simple functions f on $(\mathscr{X} \times \Theta, \mathscr{A})$; that is, *the prevision π is conglomerative w.r.t. the partitions $\{\{\theta\};\ \theta \in \Theta\}$ and $\{\{x\};\ x \in \mathscr{X}\}$ on the class of simple functions on $(\mathscr{X} \times \Theta, \mathscr{A})$* [cf. de Finetti (1930) and Dubins (1975), Section 1].

By a slight modification of the proof of Theorem 1 of Dubins (1975), we are now able to state

THEOREM 4.1. *Let $\mathscr{A}_{\mathscr{X}}$, $\mathscr{A}_\Theta$ be σ-algebras and $\mathscr{A}$ be the corresponding product σ-algebra. Let P_θ and q_x be probability charges on $\mathscr{A}_{\mathscr{X}}$ and $\mathscr{A}_\Theta$, respectively, for each x and θ, such that $P_\theta(A) = 1 = q_x(B)$ if $\{\theta\}$ implies A and $\{x\}$ implies B; $\theta \to P_\theta(A)$, $x \to q_x(B)$ are $\mathscr{A}_\Theta$, $\mathscr{A}_{\mathscr{X}}$-measurable functions for each A and B in $\mathscr{A}_{\mathscr{X}}$, $\mathscr{A}_\Theta$, respectively. Then the following statements are equivalent*:

(i) *Inequalities* (4.2) *hold.*
(ii) $\pi(C) = \int_\Theta P_\theta(C_\theta)\tau(d\theta) = \int_{\mathscr{X}} q_x(C^x)\mu(dx)$ *for all* $C \in \mathscr{A}$.

Each of these statements implies that the unique value of the prevision of every bounded, $\mathscr{A}$-measurable function f admits the representation

$$\pi(f) = \int_\Omega f\,d\pi = \int_\Theta P_\theta(f)\tau(d\theta) = \int_{\mathscr{X}} q_x(f)\mu(dx). \tag{4.4}$$

It follows from this result that (4.2) induces conglomerability on the class of bounded, $\mathscr{A}$-measurable functions, w.r.t the given partitions. On the other hand, (4.4) confirms that under suitable conditions, in order to satisfy (4.2), one has to evaluate π according to the classical procedure of Theorem 2.1 and to define the corresponding posterior as a Radon–Nikodym derivative w.r.t. μ. Analogous results may be found in the quoted papers of Heath, Lane and Sudderth [Heath and Sudderth (1978), Theorem 1, Lane and Sudderth (1983), Section 2 and (1984), Section 3 and Lane and Sudderth (1985), Theorem 1], who use (4.4) in order to characterize coherent inferences. These authors mention neither (Q_2) nor (Q_4), so that their definition, on one hand, is stronger than Definition 2.1 but, on the other hand, is weaker. The following examples illustrate this.

EXAMPLE 4.1 [from Lane (1981), Section 2]. *dF-coherent posteriors need not be coherent in the sense of Heath, Lane and Sudderth.* Set $\Theta = \mathscr{X} = \mathbb{Z}$ (the set of integers); $\mathscr{A}_\Theta = \mathscr{A}_\mathscr{X} =$ the power set of $\mathbb{Z}$; $\mathscr{A} =$ the power set of $\mathbb{Z}^2$; $P_\theta(\{x\}) = \frac{1}{2}$ if $x = \theta - 1, \theta + 1$ and $\theta \in \mathbb{Z}$.

From $\Theta_n = \{-n, \dots, 0, \dots, n\}$,

$$\tau_n(B) = \left\{ \frac{(\sqrt{2})^{2n+1} - 1}{\sqrt{2} - 1} \right\}^{-1} \sum_{K \in \Theta_n \cap B} 2^{(n+K)/2}$$

and (3.3) we obtain that $q_x(\{\theta\}) = \frac{1}{3}, \frac{2}{3}$ according to whether $\theta = x - 1$ or $\theta = x + 1$ is a dF-coherent posterior relative to P_θ and to a suitable extension to $\mathscr{A}_\Theta$ of $\tau = \lim \tau_n$. It follows that the values of the gain G from a bet on $\{\theta = x - 1\}$, conditionally on $\{x\}$, are given by $-\frac{2}{3}$ if $\theta = x - 1$ and by $\frac{1}{3}$ if $\theta = x + 1$ for all $x \in \mathbb{Z}$; hence, $P_\theta(G) = -\frac{1}{6}$ for all $\theta \in \Theta$. This result contradicts (4.2), even though it satisfies dF-coherence. Stone (1976) designates this phenomenon by the term of *strong inconsistency*.

EXAMPLE 4.2 (from an anonymous referee). *Coherent inferences in the sense of Heath, Lane and Sudderth need not satisfy* (Q_4). Set $\Theta = [0, 1]$; $\mathscr{X} = \{-1, 0, 1, \dots, n + 1\}$; $\mathscr{A}_\Theta =$ the class of Borel sets of $[0, 1]$; $\mathscr{A}_\mathscr{X} =$ the power set of $\mathscr{X}$; $\tau(\cdot) =$ the probability measure with uniform density on $[0, 1]$; $P_\theta(\{x\}) = \binom{n}{x}\theta^x(1 - \theta)^{n-x}$ if $\theta \in (0, 1)$ and $x = 0, 1, \dots, n$, $P_0(\{-1\}) = P_1(\{n + 1\}) = \frac{2}{3}$, $P_0(\{n + 1\}) = P_1(\{-1\}) = \frac{1}{3}$. These functions define a coherent P on $\mathscr{E}$. In view of (4.4), the probability measure q_x generated by the distribution function $F_x(\theta) = \{B(x + 1, n - x + 1)\}^{-1}\int_0^\theta t^x(1 - t)^{n-x}\,dt$ whenever $x = 0, 1, \dots, n$, and such that $q_{-1}(\{0\}) = q_{n+1}(\{1\}) = 1$, is a coherent posterior in the sense of Heath, Lane and Sudderth, but it does not satisfy (Q_4). On the other hand, the same q_x for $x = 0, 1, \dots, n$ with $q_{-1}(\{0\}) = q_{n+1}(\{0\}) = 1$ is a dF-coherent posterior.

Since the failure either of (Q_2) or of (Q_4) produces sure losses, it seems appropriate to restate the definition of coherence in the sense of Heath, Lane and Sudderth according to

DEFINITION 4.1. Let P^* be a probability on $\mathscr{E}^*$, where $\mathscr{A}_\mathscr{X}$, $\mathscr{A}_\Theta$ are σ-algebras and $\mathscr{A}$ is the corresponding product σ-algebra. We will say that q_x is an

H-coherent posterior relative to P^* if it is a dF-coherent posterior which satisfies (4.2) for all G.

Under the "regularity" conditions of Theorem 4.1, q_x *is an H-coherent inference if, and only if, besides* (Q_1), (Q_2) *and* (Q_4), *it satisfies* (4.4). We say again that in the original works of Heath, Lane and Sudderth, H-coherence is stated through (4.4); henceforth, we will refer to the less restrictive Definition 4.1. At any rate, H-coherence implies dF-coherence, but the converse does not generally hold (Example 4.1). Furthermore, there are problems which do not admit H-coherent posteriors as happens in the following:

EXAMPLE 4.3. Set: $\mathscr{X} = \mathbb{Z}^+$ (the set of strictly positive integers); $\Theta = \mathbb{R}$; $\mathscr{A}_{\mathscr{X}} =$ the σ-algebra generated by the finite–cofinite algebra on $\mathscr{X}$; $\mathscr{A}_\Theta =$ the power set of Θ; $\mathscr{A} =$ the σ-algebra generated by $\mathscr{A}_{\mathscr{X}} \times \mathscr{A}_\Theta$. Let $\{p_n\}$ be a sequence of strictly positive real numbers such that $1 = \sum_{n=1}^{\infty} p_n$. For every $\theta \in \mathbb{R}$, set $P_\theta(A) = \sum_{k \in A} p_k$ (for finite A and $\theta > 0$); $P_\theta(A) = 0$ (for finite A and $\theta \le 0$); $P_\theta(A) = \sum_{k \in A} p_k$ (for cofinite A and $\theta > 0$); $P_\theta(A) = 1$ (for cofinite A and $\theta \le 0$). Choose τ in such a way that $\tau((-\infty, c)) = \tau((d, +\infty)) = \frac{1}{2}$ for all $c, d \in \mathbb{R}$ and define π according to (2.1). Then $\mu(\{n\}) = p_n/2$ for all $n \in \mathbb{Z}^+$. In view of (Q_3), given $B_c = (-\infty, c)$, one obtains $0 = \pi(\{n\} \times B_c) = p_n q_n(B_c)/2$; hence, $q_n(B_c) = 0$ for all $n \in \mathbb{Z}^+$. On the other hand, $\pi(\mathscr{X} \times B_c) = \tau(B_c) = \frac{1}{2}$ and this contradicts (4.2). This is another manifestation of strong inconsistency.

5. Concluding remarks. It is plain that dF-coherence as well as H-coherence aims at fixing restrictions on inferences in order that they cannot produce uniformly positive losses. The brief analysis of the notion of H-coherence that we have given in Section 4 stresses that such a notion, under mild "regularity" conditions, provides a substantial justification for the usual procedures, which consist in extending P to $\mathscr{E}^*$ by integration of P_θ w.r.t. τ and, subsequently, in assessing a posterior by choosing a derivative of π w.r.t. its marginal distribution on $\mathscr{A}_{\mathscr{X}}$. On the other hand, we have noticed that, given a probability P on $\mathscr{E}$, dF-coherence is generally compatible with assessments yielding very different conclusions from the ones just now quoted. Therefore, it is natural that some ticklish questions about the adequacy of our Definition 2.1 may arise. We think that the main need for caution with regard to dF-coherence stems from situations such as those considered in Examples 3.1, 4.1 and 4.3, and which have been criticized by several authors [Stone and Dawid (1972), Dawid and Stone (1972, 1973), Stone (1976, 1982) and Lane (1981)]. On the other hand, these situations cannot occur in the framework of H-coherence [Sudderth (1980) and Lane and Sudderth (1983), Proposition 2.3]; in view of Section 4, they may be seen as manifestations of the failure of conglomerability. In this connection some comments on Example 3.1 may be helpful. Suppose that an inferrer becomes acquainted with the value z of x_2/x_1 but not with that of x_1. Under this circumstance, one may write $P(\cdot\,|z) = \zeta(\cdot)$ and $q_x(\cdot) = \zeta_{x_1}(\cdot)$. In view of (3.4), since ζ_{x_1} is constant w.r.t. x_1, we have that $\zeta(\cdot)$ is *not* conglomerative w.r.t. the partition $\{\{x_1\};\ x_1 > 0\}$.

The criticism of the above-mentioned authors hinges on *two points*: a particular interpretation of conditional probability surfacing when they comment upon marginalization paradox, and the remarks quoted at the beginning of Section 4. As regards the *first point*, let us start with this passage from Stone (1982), page 417, where the notation has been changed to that of Example 3.1.

"The marginalization paradox... is simply that:

(a) q_x is a function of z only, implying that we need only be informed of the values of $z = x_2/x_1$ to construct the posterior distribution of the parameter of interest;

(b) a Bayesian so informed could never agree that q_x represents Bayesian inference." [In fact, his distribution $P(\cdot|z)$ differs from q_x, as it follows from (3.4)].

These statements stress an attitude of this kind: Whenever q_x depends on z only, then q_x ought to coincide with $P(\cdot|z)$. In other words and in a more elementary situation, "if we are sure that one of the events of a sequence $H_1, H_2, \ldots$ will occur, and if we assume that in each of these possible cases the probability of E is p, then it must also be true that $P(E) = p$" [de Finetti (1949), Section 5.31]. This interpretation of conditional probability does not correspond to the definitions given in Section 1 and 2 which regard the phrase "E occurs, conditionally on the occurrence of H" as a new single concept and not as a logical deduction. Therefore, according to this formulation "there is nothing contradictory in assuming that the bets on an event E, conditionally on each of a sequence of hypotheses that are all nearly impossible by themselves, could be made on terms different from those of an unconditional bet" [de Finetti (1949), Section 5.31]. Inequality (3.4) states that the adopted inference is "sensitive" to the change of information which occurs when, z being known already, we become acquainted with the value of x_1; on the other hand, since q_x is a function of z only, x_1 affects the inference, in an indirect manner, through $z = x_2/x_1$. This explains why a Bayesian, who knows z only, cannot reach q_x: z is not a sufficient statistic even though q_x is a function of z only. It would seem unsatisfactory to consider this type of sensitiveness as contradictory because it is incompatible with the usual (and coherent) *method of conditional probability assessing* which, in the framework of dF-coherence, cannot be confused with a *definition* of conditional probability. On the other hand, there is a case in which the usual definition of conditional probability need not yield the same inference, even if, in our view, the distinctive features of that case request the equality of inferences. Suppose that X_1, X_2 are random elements such that the events $\{X_1 = x_1\}, \{X_2 = x_2\}$ *coincide*; then, by denoting an inference, under the condition that $X_i = x_i$, by $q_{x_i}^{(i)}$, $i = 1, 2$, one would expect $q_{x_1}^{(1)}(B) = q_{x_2}^{(2)}(B)$ for every $B \in \mathscr{A}_\Theta$. A straightforward application of Definitions 1.1 and 2.1 shows that such an equality is really necessary in order that the q's may be dF-coherent. On the contrary, if the q's are defined through Radon–Nikodym derivatives, without taking dF-coherence into due consideration, the two assessments derived from the Radon–Nikodym derivatives might differ and, moreover, the classical theory provides no compelling argument obliging an inferrer to adopt equal "variants" of those derivatives. One might object that the major flexibility of dF-coherence is achieved by introducing "strange" distributions, assigning probability one, for example, to events of this type: $\{|\theta| > M\}$, whatever M may be (Examples 4.1 and 4.3). But this cannot be regarded as a deficiency of the theory; in fact, dF-coherence does not compel one to adopt that type of distributions; it merely opens up the possibility, since there are problems (in quantum physics, in the theory of numbers, etc.) in which these distributions are appropriate [Rényi (1955)]. Hill (1980) expounds analogous arguments *pro* de Finetti's theory.

We will now deal with the *rationale of condition* (4.2) in the framework of dF-coherent inference by starting with this passage from Lane (1981), page 84, which refers to Example 4.1:

"Now with this betting system, the gambler achieves an advantage over the bookie... if the bookie assesses his economic situation before x is revealed, from the point of view of any of the possible values for θ, his horizon is cloudy; he foresees a loss of $1/6$... ."

In order to decide whether dF-coherence may be an adequate concept, we must decide whether it is convenient to introduce (4.2) as a compelling rule of conduct. In fact, such a condition, seen as a rule that one is free to adopt or not, is not prejudicial to dF-coherence. We try to explain why it is unsuitable for us to strengthen dF-coherence by introducing (4.2) as a new axiom. We note that the gain G of Example 4.1 is evaluated under the hypothesis that x and θ are unknown; therefore, fairness of rules governing the game has to be judged under the same hypothesis. In other words, one has to consider $\pi(G)$ which, by virtue of dF-coherence, comes out to be equal to zero. The analysis of the sign of $P_\theta(G)$ $[q_x(G)]$ is of importance whenever G expresses the consequences of an action that one is free to undertake or not, under the hypothesis that θ is the true state of nature (x is the experimental result). Obviously, in the circumstances, we are not judging the fairness of a game but the convenience of taking part. At any rate, whenever x and θ are supposed to be unknown, any judgment of convenience has to be expressed in terms of π and not of P_θ and/or q_x. Furthermore, we think that if one adheres to the point of view which leads to (4.2), then one must introduce a stronger axiom in order to define coherence. Indeed, if an inferrer sees this condition as a compelling rule because he is sure that one of the events $\{x\}, \{\theta\}$ will occur, then he ought to introduce an analogous rule for every possible partition of Ω since one of the events of a given partition will occur, whatever the partition may be. Therefore, in view of Theorems 3.1 and 3.3 of Schervish, Seidenfeld and Kadane (1984), one would conclude that π has to be σ-additive [see Hill and Lane (1985)]. Lane and Sudderth (1985), page 1246, meet this objection by saying that "it is unreasonable to require a predicter to make predictions which are not based on natural partitions such as that induced by an initial observation x." In fact, *one does not require* a predicter to make predictions conditionally on the events of arbitrary partitions; one is simply saying that a partition $\mathscr{P}$ and a simple function G (= gain from a finite betting system) might exist such that $\sup_{h \in \mathscr{P}} P_h(G) < 0$, and that a phenomenon of this kind does not seem to suit the rationale of (4.2).

Acknowledgments. The author wishes to thank the Editor, Associate Editor and referees for suggestions that improved this paper.

REFERENCES

BARRA, J.-R. (1981). *Mathematical Basis of Statistics*. Academic, New York.

BHASKARA-RAO, K. P. S. and BHASKARA-RAO, M. (1983). *Theory of Charges*. Academic, London.

BLACKWELL, D. and DUBINS, L. E. (1975). On existence and non-existence of proper, regular, conditional distributions. *Ann. Probab.* **3** 741–752.

BLACKWELL, D. and RYLL-NARDZEWSKI, C. (1963). Non-existence of everywhere proper conditional distributions. *Ann. Math. Statist.* **34** 223–225.

DARST, R. B. (1961). A note on abstract integration. *Trans. Amer. Math. Soc.* **99** 292–297.

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

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

DE FINETTI, B. (1930). Sulla proprietà conglomerativa delle probabilità subordinate. *Rend. R. Ist. Lombardo (Milano)* **63** 414–418.

DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7** 1–68. (English translation in *Studies in Subjective Probability* (1980) (H. E. Kyburg, Jr. and H. E. Smokler, eds.). Krieger, Huntington, N.Y.)

DE FINETTI, B. (1949). Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Triestini dell'Un. di Trieste* **19** 29–81. (English translation in de Finetti, B. (1972). *Probability, Induction and Statistics*, Chapter 5. Wiley, London.)

DE FINETTI, B. (1970). *Teoria delle probabilità*. Einaudi Editore, Torino. (English translation in *Theory of Probability* **1** (1974), **2** (1975). Wiley, London.)

DUBINS, L. E. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probab.* **3** 89–99.

DUBINS, L. E. (1977). Measurable, tail disintegrations of the Haar integral are purely finitely additive. *Proc. Amer. Math. Soc.* **62** 34–36.

FREEDMAN, D. A. and PURVES, R. A. (1969). Bayes' method for bookies. *Ann. Math. Statist.* **40** 1177–1186.

HARTIGAN, J. A. (1983). *Bayes Theory*. Springer, New York.

HEATH, D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345.

HILL, B. M. (1980). On some statistical paradoxes and nonconglomerability. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). University Press, Valencia.

HILL, B. M. and LANE, D. (1985). Conglomerability and countable additivity. *Sankhyā Ser. A* **47** 366–379.

HOSIASSON-LINDENBAUM, J. (1940). On confirmation. *J. Symbolic Logic* **5** 133–148.

KALLIANPUR, G. and STRIEBEL, C. (1968). Estimation of stochastic systems: arbitrary system process with additive white noise observation error. *Ann. Math. Statist.* **39** 785–801.

KRAUSS, P. H. (1968). Representation of conditional probability measures on Boolean algebras. *Acta Math. Acad. Sci. Hungar.* **19** 229–241.

LANE, D. A. (1981). Coherence and prediction. *Bull. ISI Proc. 43rd Session* **1** 81–96.

LANE, D. A. and SUDDERTH, W. D. (1983). Coherent and continuous inference. *Ann. Statist.* **11** 114–120.

LANE, D. and SUDDERTH, W. D. (1984). Coherent predictive inference. *Sankhyā Ser. A* **46** 166–185.

LANE, D. and SUDDERTH, W. D. (1985). Coherent predictions are strategic. *Ann. Statist.* **13** 1244–1248.

LEHMAN, R. S. (1955). On confirmation and rational betting. *J. Symbolic Logic* **20** 251–262.

MAYNARD, H. B. (1979). A Radon–Nikodym theorem for finitely additive bounded measures. *Pacific J. Math.* **83** 401–413.

MAZURKIEWICZ, S. (1932). Zur Axiomatik der Wahrscheinlichkeitsrechnung. *C. R. Soc. Sci. Lett. Varsovie III* **25** 1–4.

REGAZZINI, E. (1983). Coherent conditional probabilities, finite additivity, extensions. Technical Report, Dipartimento di Scienze Statistiche, Univ. of Bologna.

REGAZZINI, E. (1984). A Bayes' theorem for finitely additive probabilities and some considerations about the use of improper priors. Technical Report S 30/1984, Dip. Matematica, Univ. of Milano.

REGAZZINI, E. (1985). Coherent conditional probabilities. *Rend. Sem. Mat. Fis. Milano* **55**. To appear.

RÉNYI, A. (1955). On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hungar.* **6** 285–335.

SCHERVISH, M. J., SEIDENFELD, T. and KADANE, J. B. (1984). The extent of nonconglomerability of finitely additive probabilities. *Z. Wahrsch. verw. Gebiete* **66** 205–226.

STONE, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.* **71** 114–125.

STONE, M. (1982). Review and analysis of some inconsistencies related to improper priors and finite additivity. In *Logic, Methodology and Philosophy of Science* (L. J. Cohen, J. Łoś, H. Pfeiffer and K. P. Podewski, eds.) **6** 413–426. North-Holland, Amsterdam.

STONE, M. and DAWID, A. P. (1972). Un-Bayesian implications of improper Bayes inference in routine statistical problems. *Biometrika* **59** 369–375.

SUDDERTH, W. D. (1980). Finitely additive priors, coherence and the marginalization paradox. *J. Roy. Statist. Soc. Ser. B* **42** 339–341.

WILLIAMS, P. M. (1975). Notes on conditional previsions. School of Mathematical and Physical Sciences, Univ. of Sussex.

DIPARTIMENTO DI MATEMATICA
"FEDERIGO ENRIQUES"
UNIVERSITÀ DI MILANO
VIA C. SALDINI, 50
20133 MILANO
ITALY