Fine-Grained Opinion, Probability, and the Logic of Full Belief
Author(s): Bas C. van Fraassen
Source: *Journal of Philosophical Logic*, Vol. 24, No. 4 (Aug., 1995), pp. 349-377
Published by: Springer
Stable URL: http://www.jstor.org/stable/30226553

BAS C. VAN FRAASSEN

FINE-GRAINED OPINION, PROBABILITY, AND THE LOGIC OF FULL BELIEF

Personal or subjective probability entered epistemology as a cure for certain perceived inadequacies in the traditional notion of belief. But there are severe strains in the relationship between probability and belief. They seem too intimately related to exist as separate but equal; yet if either is taken as the more basic, the other may suffer.

After explaining the difficulties in some detail I will propose a single unified account which takes conditional personal probability as basic. Full belief is therefore a defined, derivative notion. Yet it is easiest to explain the resulting picture of opinion as follows: my subjective probability is only a grading of the possibilities left open by my beliefs. My conditional probabilities generally derive from the strongest belief I can maintain when admitting the relevant condition Appendices will survey the literature.

1. FULL BELIEF AND PERSONAL PROBABILITY

The most forceful answer I can give if asked for my opinion, is to say what I fully believe. The point of having beliefs is to construct a *single* (though in general incomplete) picture of what things are like. One obvious model of this part of my opinion is a set of propositions.[1] Their intersection is the proposition which captures exactly that single picture of the world which has my full assent. Clearly a person's full beliefs leave open many alternatives. Alternatives left open by belief are then also represented by (sets of) propositions, namely ones that imply my beliefs. But these alternatives do not all have the same status for me, though they are all "possible for all I [know or] believe." Some seem more or less likely than others: enter personal (subjective) probability, as a grading of the possibilities left open by one's beliefs.

I will take for granted that the probability of a proposition is a real number in the interval [0, 1], with the empty proposition Λ (self-contra-

Journal of Philosophical Logic **24**: 349–377, 1995.

diction) receiving 0 and the universal proposition U (tautology) receiving 1. The assignment is a *measure*, that is, it is additive and continuous (equivalently, countably additive). It follows from this that the assignment of probabilities respects the ordering by logical implication:

$$\text{If } A \subseteq B \text{ then } P(A) \leqslant P(B)$$

though we must be careful in any extrapolation from propositions to sets of propositions unless they are countable. That is essentially because at most countably many disjoint propositions can receive finite positive probability. (*Reason*: at most one can receive probability greater than 1/2, at most two can receive more than 1/3, ... etc. The question of infinitesimal positive probability will be taken up in an Appendix.)

The so-called lottery paradox shows that we cannot equate belief with probability $\geqslant p$, if $p < 1$. For example, suppose $p = 0.99$ and a lottery which I believe to be fair has 1000 tickets, then my probability that the k^{th} ticket will not win the (single) prize equals 0.999. Hence for each $k = 1, \ldots, 1000$, I would believe that the k^{th} ticket will not win. My beliefs would then entail that all tickets will fail to win, which conflicts with my original belief that the lottery is fair. This argument is more important for what it presupposes than for what it shows. It is clearly based on the assumed role of full belief: to form a single, unequivocally endorsed picture of what things are like.[2]

In fact, the thesis that probabilities grade exactly the alternatives left open by full belief guarantees that all full beliefs have maximal personal probability.

So what if we simply set $p = 1$, that is, *identify* our full beliefs with the propositions which are maximally likely to us? The first problem with this is that we seem to be treating full beliefs as on a par with tautologies. Are there no distinctions to be drawn among the maximally likely propositions? There is a second problem for this proposal as well. In science we deal with continuous quantities. Therefore, in general, if we let science guide our opinion, the maximally likely propositions will not form a single picture – they will just give us a family of rival maximally likely pictures.

EXAMPLE 1. Consider the mass of the moon reckoned in kilograms, and suppose I am sure that it is a number in the interval $[a, b]$. If my probability follows Lebesgue measure then my probability is *zero* that the number equals x, for $a \leqslant x \leqslant b$. Hence my probability equals 100% that the number lies in the set $[a, b] - \{x\}$, for each such number x. Yet no real number belongs to all these sets – their intersection is empty. Probability measures of this sort (deriving from continuous probability densities) are ubiquitous in science, and informed opinion must be allowed to let itself be guided by them. We have here a transfinite lottery paradox, and we can't get out of it in the way that worked for the finite case (see Maher, 1990).

2. SUPPOSITION AND TWO-PLACE PROBABILITY

There is a third aspect of opinion, besides belief and subjective grading, namely *supposition*. Much of our opinion can be elicited only by asking us to suppose something, which we may or may not believe. The respondent imaginatively puts himself in the position of someone for whom the supposition has some privileged epistemic status. But if his answer is to express his present opinion – which is surely what is requested – then this "momentary" shift in status must be guided by what his present opinion is. How does this guidance work?

One suggestion is that the respondent moves to a state of opinion derived from his own in two steps: (1) discarding beliefs so that the supposition receives more than minimal likelihood; (2) then (without further change in beliefs) regrading the alternatives left open so as to give the supposition maximal likelihood. This makes sense only if both steps are unambiguous. We can imagine a simple case. Suppose Peter has as "primary" beliefs A and B, and believes exactly what they jointly entail; he is asked to entertain the supposition $C - A$. In response he imaginatively moves into the epistemic position in which (1) B is the only primary belief, and (2) he assigns 0 to all alternatives left open by B which conflict with $(C - A)$ and then regrades the others in the same proportions as they had but with the maximum assigned to $(B \cap C - A)$.

This simple case already hinges on a certain hierarchical structure in Peter's opinion. Moreover it presupposes that those alternatives which were left open by B, but which conflict with his initial equally primary belief that A, had been graded proportionately as well. Even more structure must be present to guide the two steps in less simple cases. What if the beliefs had been, say, A, B, and D, and their joint consequences, and the supposition was compatible with each but not with the conjunction of any two? The discarding process can then be guided only if some hierarchy among the beliefs determines the selection.

Let us consider conditional personal probability as a possible means for describing structure of this sort. The intuitive Example 1 above about the mass of the moon is the sort often given to argue for the irreducibility of conditional probability. I could continue the example with: the mass of the moon seems to me to equally likely to be x as $(x + b)/2$, on the supposition that it is one of those two numbers. The two possibilities at issue here are represented by the degenerate intervals $[x]$, $[(x + b)/2]$, so both they and the supposition that one or other is the case (represented by set $\{x,\ (x+b)/2$ their union) receive probability 0. The usual calculation of conditional probability, which would set $P(B \mid A)$ equal to $P(B \cap C \mid A)$ divided by $P(C \mid A \cap C)$, can therefore not be carried out. The suggestion that conditional probability is irreducible means that *two-place probability* $P(\ \mid\ )$ – probability of one thing *given (on supposition of)* another – is autonomous and cannot be defined in terms of the usual one-place ("absolute") probability. Rather the reverse: we should define $P(\ ) = P(\ \mid U)$, probability conditional on the tautology.

There is a good deal of literature on two-place ("irreducible conditional") probability (see Appendix). Despite many individual differences, general agreement concerning two-place probability extends to:

I. If P is a 2-place probability function then $P(- \mid A)$ is "normally" a (1-place) probability function with $P(A \mid A) = 1$.

II. These derivative 1-place probability functions [described in I.] are related at least by the *Multiplication Axiom*:

$$P(B \cap C \mid A) = P(B \mid A)P(C \mid A \cap C)$$

where $A, B, C, \ldots$ are assumed to be in the domain and co-domain of the function. The "normally" restriction (eliminating at least $A = \Lambda$) is to be discussed below.

New non-trivial relationships between propositions are now definable. De Finetti suggested relations of local comparison of the following type:

$$A \text{ is } \textit{superior} \text{ to } B \text{ iff } P(A \mid A + B) = 1$$

where '+' marks exclusive disjunction: $A + B = (A - B) \cup (B - A)$.[3]

EXAMPLE 2. Given any probability measure P it is easy to produce a 2-place function that has that character:

$$\begin{aligned} \text{define } P(A \mid B) &= P(A \cap B)/P(B) \text{ if } P(B) > 0 \\ &= 1 \text{ if } P(B) = 0 \end{aligned}$$

That is a trivial 2-place function since it is definable from a 1-place function.

EXAMPLE 3. Let U be the set of natural number $\{0, 1, 2, \ldots\}$. For index $n = 0, 1, 2, \ldots$ let p_n be the probability measure defined on all subsets of U by the condition that it assigns 0.1 to $\{x\}$ if x is in the set $\{10n, \ldots, 10n + 9\}$, and 0 otherwise. Define:

$$P(A \mid B) = p_n(A \cap B)/p_n(B) \text{ for the first index } n \text{ such that}$$

$$p_n(B) > 0, \text{ if there is any such index; } = 1 \text{ otherwise.}$$

To verify the Multiplication Axiom note for instance that if $A \cap C$ is not empty, and $P(A \mid C) > 0$ then the first index n for which $p_n(A \cap C) > 0$ is the same as the first index m such that $p_m(C) > 0$. The "otherwise" clause will apply here only if $B = \Lambda$.

These examples are instances of the "lexicographic" probability models which I will discuss at some length below. We make the ideas of one- and two-place probability precise as follows.

> A *space* is a couple $S = \langle U, F \rangle$ with U a non-empty set (the *worlds*) and F (the family of *propositions*) a sigma-field on U, that is:
>
> (a) $U \in F$
>
> (b) if $A, B \in F$ then $A - B \in F$

(c) if $\{A_i\colon\ i = 1, 2, \ldots\} \subseteq F$ then $\cup\{A_i\} \in F$

A (1-place) *probability measure* P on space $S = \langle U, F\rangle$ is a function mapping F into the real numbers, subject to

1. $0 = P(\Lambda) \leqslant P(A) \leqslant P(U) = 1$
2. $P(A \cup B) + P(A \cap B) = P(A) + P(B)$ (*finite additivity*)
3. If $E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n \subseteq \ldots$ has union E, then $P(E) = \sup\{P(E_n)\colon\ n = 1, 2, \ldots\}$ (*continuity*)

Property 3 is in this context equivalent to *countable additivity*:

4. If $\{E_n\colon\ n = 1, 2, \ldots\}$ are disjoint, with union E, then $P(E) = \Sigma\{P(E_n)\colon\ n = 1, 2, \ldots\}$

and also to the dual continuity condition for countable intersection. The general class of two-place probability measures to be defined now will below be seen to contain a rich variety of non-trivial examples.

A *2-place probability measure* $P(- \mid -)$ on space $S = \langle U, F\rangle$ is a map of $F \times F$ into real numbers such that

I. (Reduction Axiom) The function $P(- \mid A)$ is either a probability measure on S or else has constant value $= 1$.

II. (Multiplication Axiom)

$$P(B \cap C \mid A) = P(B \mid A)P(C \mid B \cap A)$$

for all A, B, C, in F.

If $P(\mid A)$ is a (1-place) probability measure, I shall call A *normal* (for P), and otherwise *abnormal*. ("Absurd" might have been a better name; it is clearly a notion allied to self-contradiction.) The definition of 2-place probability allows for the totally abnormal state of opinion ($P(A \mid B) = 1$ for all A and B). It should not be excluded formally, but I shall tacitly exclude it during informal discussion. Here are some initial consequences of the definition. The variables range of course over propositions (members of family F in space $S = \langle U, F\rangle$).

[T2.1] $\quad P(X \mid A) = P(X \cap A \mid A)$

[T2.2] If A is normal, so are its supersets

[T2.3] If A is abnormal, so are its subsets

[T2.4] B is abnormal iff $P(-B \mid B) = 1$; iff $P(B \mid A) = 0$ for all normal A.

Let us call the case in which only Λ is abnormal the "Very Fine" case; that there are Very Fine 2-place probability measures on infinite fields will follow from results below.

We add one consequence which is related to De Finetti's notion of conglomerability:

[T2.5] Condition Continuity: If $\{E_n\}$ is a countable increasing chain – i.e. E_n part of E_{n+1} – with union E, and $P(E_n \mid E) > 0$ for all n, then $P(X \mid E)$ is the limit of the numbers $\{P(X \mid E_n)\}$.

To prove this: $P(X \mid E_n) = P(X \cap E_n \mid E)/P(E_n \cap E)$, so since E is normal, this follows from the principle of condition continuity which can be demonstrated for (one-place) probability measures.[4]

3. THE IDEA OF THE *A PRIORI*

In any conception of our epistemic state there will be propositions which are not epistemically distinguishable from the tautology U – let us say these are *a priori* for the person. This notion is the opposite of the idea of abnormality:

A is *a priori for* P iff $P(A \mid X) = 1$ for all X,
iff $U - A$ is abnormal for P.

What is *a priori* for a person is therefore exactly what is certain for him or her on any supposition whatsoever. This notion generalizes unconditional certainty, i.e. $P(A) = 1$. The strongest unconditional probability equivalence relation between A and B is that their symmetric difference $(A + B)$ has measure zero. We can generalize this similarly. As our strictest epistemic equivalence relation between two propositions we have *a priori* equivalence (their symmetric difference has probability 0 on all normal suppositions):

$A\langle P\rangle B$ iff $A + B$ is abnormal.[5]

The abnormal propositions are the ones *a priori* equivalent to the empty set (the self-contradiction) and the *a prioris* are the ones *a priori* equivalent to the tautology. (Of course these are subjective notions: we are speaking of what is *a priori* for the person with this state of opinion.)

Note now that $A\langle P\rangle B$ iff $P(A \mid A+B) = 1$ and $P(B \mid A+B) = 1$, since additivity would not allow that if $A + B$ were normal. We can divide this equivalence relation into its two conjuncts:

DEFINITION. A $P > B$ iff $P(A \mid A+B) = 1$.

This is the relationship of "superiority" mentioned above.

[T3.1] If A logically implies B then $B\ P > A$.

It follows from the definition that $A\ P > A$. In a later section I shall also show that $P >$ is transitive. Clearly if $A + B$ is normal, then $A\ P > B$ means that A is comparatively superior to B, in the sense that A is certainly true and B certainly false, on the supposition that one but not both are the case. But if $A + B$ is abnormal then the relationship $A\ P > B$ amounts to $A\langle P\rangle B$. The right reading for "$P >$" is therefore "is superior to or *a priori* equivalent to". To be brief, however, I'll just say "A is superior to B" for "$A\ P > B$", and ask you to keep the qualifications in mind.

4. FULL BELIEF REVISITED

The beliefs I hold so strongly that they are *a priori* for me are those whose contraries are all abnormal. There is a weaker condition a proposition K can satisfy: namely that any normal proposition which implies K is superior to any that are contrary to K. Consider the following conditions and definitions:

(A1)	Normality:	K is normal
(A2)	Superiority:	If A is a non-empty subset of K while B and K are disjoint, then $A\ P > B$
(A3)	Contingency:	The complement $U - K$ of K is normal.

We can restate the "Superiority" condition informally as follows:

> *Superiority*: the alternatives K leaves open are all superior to any alternative that K excludes.

We can deduce from these conditions something reminiscent of Carnap's "regularity" (or Shimony's "strict coherence"):

(A4) Finesse: all non-empty subsets of K are normal.

DEFINITION. K is a *belief core* (for P) iff K satisfies (A1)–(A3).

Note that the *a priori* propositions satisfy (A1) and (A2), though definitely not (A3), but rather its opposite. However, the following elementary results show that all the *a prioris* are among the propositions implied by belief cores.

[T4.1] If K is a belief core then $P(K \mid U) = 1$.

[T4.2] If K is a belief core then A $P > K$ iff K implies A.

[T4.3] If K is a belief core and A is *a priori* then K is a subset of A.

To characterize the full beliefs we need take into account the extreme possibility of there being no belief cores; in that case we still want the *a prioris* to be full beliefs of course. (This corresponds to what I have elsewhere called "Zen minds": states of opinion in which nothing is fully believed if it is subjectively possible to withhold belief.) In view of the above we have several equivalent candidates for this characterization, of which we can choose one as definition:

DEFINITION. A is a *full belief* (for P) iff (i) there is a belief core, and $A\ P > K$ for some belief core K; or (ii) there is no belief core, and A is *a priori*.

[T4.4] The following conditions are equivalent:
(a) A is a full belief (for P).

(b) Some proposition J which is either *a priori* or a full belief core is such that $A\ P > J$.

(c) A is implied either by an *a priori* or by a belief core (for P).

Very little in this discussion of full belief hinges on the peculiarities of probability. Indeed, (conditional) probability enters here only to give us intelligible, non-trivial resources for defining the notions of subjective superiority and a-prioricity (equivalently, [ab]normality). If those notions could be acceptably primitive, the account of full belief here would just take the following form:

A belief core is a proposition K such that:

(a) K and its complement are both normal; (b) K does not leave open any abnormal alternatives; (c) any alternatives left open by K are superior to any alternatives K excludes.

A belief is any proposition implied by a belief core (or *a priori*).

As before, a proposition is here called an alternative left open by K exactly if it is non-empty and implies K.

5. IDENTIFICATION OF FULL BELIEFS

By the definition I gave of full belief, beliefs are clustered: each belongs to a family $\{A\colon\ A\ P > K\}$ for some belief core K (if there are any at all), which by [T4.2] sums up that cluster exactly:

$$K \text{ is the intersection of } \{A\colon\ A\ P > K\}$$

$$\{A\colon\ A\ P > K\} = \{A\colon\ K \text{ implies } A\}.$$

We can now prove that these clusters form a chain, linearly ordered by set inclusion (implication).

[T5.1] If K, K' are belief cores, then either K is a subset of K' or K' is a subset of K.

For the proof let $A = K - K'$ and $B = K' - K$. Each is normal if non-empty, by (A4). If B is not empty then, since A is disjoint from K', it follows by (A2) that $B\ P > A$. By parity of reasoning, if A is not empty, then $A\ P > B$. But we cannot have both unless $A + B$ is abnormal, and hence empty by [T2.2] and (A4). So we conclude that either A or B or both are empty; hence at least one of K and K' is a subset of the other.

This result is crucial for the characterization of full belief. It is therefore worthwhile to note that the only ingredients needed for the proof were the features of Superiority and Finesse of belief cores, plus the following characteristic of the superiority relationship: if $A\ P > B$ and $B\ P > A$ then $A + B$ is abnormal. Here is an illustrative example:

EXAMPLE 4. For the center of mass of the remains of Noah's ark, Petra has subjective probability 1 for each of the following three propositions: that it lies in the Northern Hemisphere (which part of the earth includes the Equator), that it lies in the Western Hemisphere, and thirdly that it lies North of the Equator. But only the first two of these propositions are among her full beliefs; the third is not. On the supposition that one but only one of these beliefs is true, she gives 100% probability to the first proposition, that it lies in the Northern Hemisphere.

Note that the last sentence implies that the first proposition is superior to the second, although both are full beliefs. I will give a formal reconstruction of this example later on.

Writing K^* for the intersection of all the belief cores, we conclude that if A is a full belief, then K^* implies A. But is K^* itself a belief core? Does it have 100% probability? Is it even non-empty? This is the problem of transfinite consistency of full belief in our new setting.

[T5.2] The intersection of a non-empty countable family of belief cores is a belief core.

For proof, assume there is at least one belief core; call it K. Assume also that the belief cores are countable and form a chain (the latter by [T5.1]), and call the intersection K^*. Countable additivity of the ordinary probability measure $P(\,) = P(\mid U)$ is equivalent to just the right continuity condition needed here: the probabilities of the members of a countable chain of sets converge to the probability of its intersection. Since in our case all those numbers equal 1, so does

$P(K^*)$. Therefore also K^* is not empty, and thus normal because it is a subset of at least one belief core. Moreover, so are its non-empty subsets, so that they are normal too. Its complement $U - K^*$ includes $U - K$, and is therefore also normal.

We have now seen that K^* satisfies conditions (A1), (A3), and (A4), and need still to establish (A2). If A is a normal subset of K^*, and hence of all belief cores, and B is disjoint of K^*, we have $P(A \mid A + (B - K')) = 1$ for all belief cores K'. But the sets $B - K'$ form an increasing chain whose union is $B - K^* = B$. Hence also the sets $A + (B - K')$ here form such a chain with union $A + B$. To conclude now that $P(A \mid A+B) = 1$, we appeal to [T2.5], the principle of Condition Continuity. This ends the proof.

The significance of this result may be challenged by noting that the intersection of countably many sets of measure 1 also has measure 1. So how have we made progress with the transfinite lottery paradox? In four ways. The first is that in the representation of opinion we may have a "small" family of belief cores even if probability is continuous and there are uncountably many propositions with probability 1. The second is that no matter how large a chain is, its intersection is one of its members if it has a first (= "smallest") element. The third is that the following is a condition typically met in spaces on which probabilities are defined even in the most scientifically sophisticated applications:

(∗) Any chain of propositions, linearly ordered by set inclusion, has a countable subchain with the same intersection.

[T5.3] If (∗) holds and there is at least one belief core, then the intersection of all belief cores is also a belief core.

This is a corollary to [T5.2].

Fourthly, farther below I will also describe an especially nice class of models of fine-grained opinion for which we can prove that the intersection of the belief cores, if any, is always also a belief core ("lexicographic probability"). There are no countability restrictions there.

6. TO WHAT EXTENT DOES BELIEF GUIDE OPINION?

It is not to be expected that every two-place probability function is admissible as a representation of (possible) opinion. If we want to use this theory in descriptive epistemology, it is necessary to look for kinds of probability functions that have interesting structure. There are models in which there are no belief cores at all. Combining our previous Examples 1 and 2, take Lebesgue measure m on the unit interval, and trivially extend it to a two-place function by $P(A \mid B) = m(A \cap B)/m(B)$ if defined and $P(A \mid B) = 1$ if not (though A, B in domain of m). Then every unit set $\{x\}$ is in the domain and is abnormal. Therefore there is no set all of whose subsets are normal, and hence no belief cores. (The absence of belief cores in our present example derives from its triviality, and not from the continuity.) Obviously then, if this represents someone's opinion, his opinions are not guided or constrained by his beliefs (which include only the *a priori*).[6]

At the other extreme from this example, there is the Very Fine case of a probability function P for which every non-empty set is normal.

DEFINITION. P is *belief covered* if the union of the belief cores equals U.

In that case, P is Very Fine. For let A be any non-empty proposition; there will be some belief core K such that $K \cap A$ is not empty, hence normal, thus making A normal.

Example 3 in Section 2.1 furnishes us with a relatively simple example of this sort. Recall that P is there constructed from the series $p_0, p_1, \ldots, p_n, \ldots$ where the whole probability mass of p_n is concentrated (and evenly distributed) on the natural numbers $\{10n, \ldots, 10n + 9\}$. In this example, the belief cores are exactly the sets

$$K_0 = \{0, \ldots, 9\},\ K_1 = \{0, \ldots, 19\},$$
$$K_2 = \{0, \ldots, 29\}, \ldots K_i = \{0, \ldots, 10i + 9\}$$

Clearly K_i is normal, since $P(- \mid K_i) = P(- \mid K_0) = p_0$. The complement of K_i is normal too, for

$$P(- \mid U - K_i) = P(- \mid \{10(i + 1), \ldots\}) = p_{i+1}.$$

If A is a non-empty subset of K_i and B is disjoint from K_i, then A is superior to B. Specifically, the first n such that $p_n(A) > 0$ can be no higher than i in that case, while the first m such that $p_m(B) > 0$ can be no lower than $i + 1$. Therefore, the first k such that $p_k(A + B) > 0$ will assign a positive probability to A and *zero* to B.

These belief cores clearly cover U; P is belief covered and Very Fine. Indeed, the belief cores are well-ordered.

Define the *belief remnants*

$$R_0 = K_0$$
$$R_{j+1} = K_{j+1} - K_j \ (j = 0, 1, 2, \ldots).$$

Clearly $p_i = P(| \, R_i)$; for example, $p_1 = P(| \, \{10, \ldots, 19\})$. Probabilities conditional on belief remnants (beliefs remaining upon retrenchment to a weaker core) determine all probabilities in this case:

$$P(- \mid A) = P(- \mid A \cap R_i) \text{ for the first } i \text{ such that } P(A \mid R_i) > 0.$$

This says quite clearly that (in this case) belief guides opinion, for probabilities conditional on belief remnants are, so to speak, all the conditional probabilities there are.

7. THE MULTIPLICATION AXIOM VISUALIZED

In the basic theory of two-place probability, the Multiplication Axiom places the only constraint on how the one-place functions $P(| \, A)$, $P(| \, B), \ldots$ are related to each other. It entails that the proposition $A - B$ is irrelevant to the value of $P(B \mid A)$ – that this value is the same as $P(A \cap B \mid A)$ – and that the usual ratio-formula calculates the conditional probability when applicable. Indeed, the ratio formula applies in the generalized form summarized in the following:

If X is a subset of A which is a subset of B, then:

[T7.1] if $P(A \mid B) > 0$ then $P(X \mid A) = P(X \mid B) : P(A \mid B)$

[T7.2] if X is normal, then $P(X \mid B) \leqslant P(X \mid A)$.

There is another way to sum up how the Multiplication Axiom constrains the relation between $P(| \, A)$ and $P(| \, B)$ in general. When we consider the two conditional probabilities thus assigned to any proposition that implies both A and B, we find a proportionality factor, which is constant when defined.

[T7.3] If $P(A \mid B) > 0$ then there is a constant $k \geqslant 0$ such that for all subsets X of $A \cap B$, $P(X \mid A) = kP(X \mid B)$. The constant $k = k(A, B) = [P(B \mid A)/P(A \mid B)]$, defined provided $P(A \mid B) > 0$.

8. EQUIVALENCE RELATIONS ON PROPOSITIONS

In the main literature on two-place probability we find an equivalence relationship other than *a priori equivalence*, which I shall call *surface equivalence*:[7]

[T8.1] The following conditions are equivalent:
(a) $P(\cdot \mid A) = P(\cdot \mid B)$
(b) $P(A{+}B \mid A{\cup}B) = 0$ or else both A, B are abnormal
(c) $P(A \mid B) = P(B \mid A) = 1$.

(I use the dot for function notation: (a) means that $P(X \mid A) = P(X \mid B)$ for all X.) It is easy to prove that (a) implies (b), for if (a) and either A or B is normal then both are normal. Secondly suppose (b). If A, B abnormal then (c) follows. Suppose then that one of A, B is normal, so $A \cup B$ is normal. Then either $P(A \mid A \cup B)$ or $P(B \mid A \cup B)$ is positive; suppose the first. Since $A \cap B \subseteq A \subseteq A \cup B$ it follows by [T7.1] that $P(B \mid A) = P(B \cap A \mid A \cup B) : P(A \mid A \cup B)$. But $P(A - B \mid A \cup B) = 0$ so $P(A \cap B \mid A \cup B) = P(A \mid A \cup B) > 0$; hence we conclude that $P(B \mid A) = 1$. Accordingly, B too is normal and $P(B \mid A \cup B)$ is positive; the same argument leads *mutatis mutandis* to $P(A \mid B) = 1$. Therefore (b) implies (c).

Finally suppose (c). If A is abnormal, then so is B and (a) follows at once. If A is normal, then B and $A \cup B$ are also normal. But then $P(A \cap B \mid A \cup B) = P(A \mid A \cup B) = P(B \mid A \cup B) = 1$, using (c) and the Multiplication Axioms. Hence $P(X \cap A \cap B \mid A \cup B) = P(X \cap A \mid A \cup B) = P(X \mid A)$ and by similar reasoning $P(X \cap A \cap B \mid A \cup B) = P(X \mid B)$. Therefore (c) implies (a).

The relationship of *a priori* equivalence is much "deeper". As prelude let us introduce another operation on probability functions which is something like "deep conditionalization". Instead of raising a proposition to the status of subjective certainty it raises it to subjective apprioricity. To prevent confusion, I shall call this "relativization".[8]

DEFINITION. The *relativization of* P *to* A is the function $P//A$ defined by $P//A(X \mid Y) = P(X \mid Y \cap A)$ for all X, Y.

[T8.2] The following are equivalent:

(i) $P(A \mid \cdot) = P(B \mid \cdot)$

(ii) $P(- \mid \cdot \cap A) = P(- \mid \cdot \cap B)$ (i.e. $P//A = P//B$)

(iii) $P(A \mid A + B) = P(B \mid A + B) = 1$ (*a priori* equivalence)

(iv) $A + B$ is abnormal (i.e. $A\langle P\rangle B$)

(For proof that (ii) implies (iv), use the *Lemma*: if A, B are disjoint and abnormal then $A \cup B$ is abnormal.).

[T8.3] *A priori* equivalence implies surface equivalence.

The converse does not hold, as can be seen from our Example 3 in Section 2.1, where U is the set of natural numbers, and is surface equivalent, but not *a priori* equivalent, to $\{0, 1, \ldots, 9\}$. For $P(\{0, 1, \ldots, 9\} \mid \{10, \ldots, 19\}) = 0$ there.

9. IMPLICATION RELATIONS; SUPERIORITY IS TRANSITIVE

We are representing propositions by means of a field of sets, whose elements are thought of as alternative possible situations or worlds. Accordingly, "A implies B" can be equated only with "$A \subseteq B$." But when two propositions are *a priori* equivalent for P then they are not distinguishable as far as P is concerned. Therefore we can introduce a less sensitive partial ordering as a "coarser" implication relationship with respect to a given two-place probability measure.

[T9.1] The following are equivalent:

(a) A *P-implies* B: $P(A \mid \cdot) \leqslant P(B \mid \cdot)$

(b) $A - B$ is abnormal

(c) for all X, if $P(A \mid X) = 1$ then $P(B \mid X) = 1$.

The superiority relation is not a (converse) implication relationship, despite formal similarities. If A is superior to B, A may still have probability *zero* conditional on B, for example. It is just that the supposition that A has to be given up – in fact, denied – before B comes into play on our thinking. The hierarchy so indicated John Vickers (*loc. cit.*) calls the De Finetti hierarchy. As he points out, it is crucial to this role that we can describe the comparison in terms of a transitive relationship.

In fact, only one further point needs to be made concerning normality to show that the propositions form a partially ordered set under superiority (with the abnormal propositions forming an equivalence class at the bottom of this p.o.set).

[T9.2] If X is a subset of Y and Y a subset of normal set E, and $P(Y \mid E) > 0$, then $P(X \mid Y) = 0$ iff $P(X \mid E) = 0$.

[T9.3] $P>$ is transitive.

Let it be given that $A\ P > B$ and $B\ P > C$; we need to prove that $A\ P > C$. To visualize the proof, think of a Venn Diagram with the following labels for relevant propositions:

$$1 = (A - B - C),\ 2 = (AB - C),\ 3 = B - A - C,$$
$$4 = AC - B$$
$$5 = BC - A,\ 6 = C - B - A$$
$$E = (A + B) \cup (B + C) = (A + B) \cup (B + C) \cup (A + C)$$

I will denote unions by concatenation of the labels; thus $E = 123456$ and $A - C = 12$. We now consider all possible cases.

(1) $A + C$ is abnormal. Then $P(A \mid A + C) = 1$

(2) $A + C$ is normal; then also E is normal.

Hence $P(A + B \mid E)$ or $P(B + C \mid E)$ or both are positive; we proceed to the possible subcases:

(2.1) $P(A + B \mid E) > 0$. By the given and [T9.2] it follows that $P(B - A \mid E) = 0$, i.e. $P(3 \mid E) = P(5 \mid E) = 0$.

(2.11) Assume $P(B + C \mid E) > 0$. By [T9.2], and the given, also $P(C - B \mid E) = 0$, so $P(4 \mid E) = P(6 \mid E) = 0$. Altogether, $P(3456 \mid E) = 0$ hence $P(12 \mid E) = P(A - C \mid E) = 1$. It follows that

$P(A+C \mid E) > 0$, so by [T9.2] $P(56 \mid A+C) = P(C \mid A+C) = 0$, and therefore $P(A \mid A+C) = 1$.

(2.12) Assume $P(B+C \mid E) = 0$. Then $P(2346 \mid E) = 0$, so altogether $P(23456 \mid E) = 0$. Hence $P(1 \mid E) = 1$. It follows that $P(A+C \mid E) > 0$, therefore by [T9.2] again $P(56 \mid A+C) = 0$. It follows then that $P(A \mid A+C) = 1$.

(2.2) $P(A+B \mid E) = 0$ and $P(B+C \mid E) > 0$. The former entails that $P(1345 \mid E) = 0$. The latter entails by [T9.2] that $P(C-B \mid E) = P(46 \mid E) = 0$. Altogether therefore $P(13456 \mid E) = 0$ and $P(2 \mid E) = 1$. Therefore $P(A+C \mid E) > 0$, and so $P(C-A \mid A+C) = 0$ by [T9.2]; therefore $P(A \mid A+C) = 1$. This ends the proof.

Adding this to the fact that $P >$ is reflexive, we conclude that $P >$ is a partial ordering of the field of propositions. The abnormal propositions form a $\langle P \rangle$ equivalence class at the very bottom of this partial ordering.

10. A LARGE CLASS OF MODELS

I will define a class of models such that P satisfies principles I–II iff P can be represented by one of these models, in the way to be explained. The class will be chosen large; a special subclass ("lexicographic models") will yield nontrivial, easily constructed examples to be used in illustrations and refutations. (The term "lexicographic" is used similarly in decision theory literature; see Blume *et al.*, 1991a, 1991b.)

A model begins with a sample space $S = \langle U, F \rangle$, where U is a non-empty set (the universe of possibilities) and F a sigma-field of sets on U (the propositions). We define the subfields:

$$\text{if } A \text{ is in } F \text{ then } FA = \{E \cap A\colon\ E \text{ in } F\};$$

thus FA is a field on A. For each such field designate as PA the set of probability measures defined on FA. (When A is empty, $FA = \{A\}$ and PA is empty.) The restriction of a member p of PA to a subfield FB, with B a subset of A, will be designed $p \mid FB$. Finally let PS be the union of all the sets PA, A in F.

A model M will consist of a sample space S as above, and a function π defined on a subset of F, with range in PS. That is, π associates some probability measure on some subfield with certain propositions. (These will be the normal propositions.) I will abbreviate "$\pi(A)$" to "πA", and when p is in FB I will also designate B as Up (the universe of p). Thus $B = U\pi A$ means that π associates with A a probability measure defined on the measurable subsets of B, i.e. on the propositions which imply B, i.e. on FB. The function π is subject to the following conditions:

(M1) $\pi A(A)$ is defined and positive.

(M2) If $\pi B(A)$ is defined and positive, then πA is defined

(M3) If $\pi B(A)$ is defined and positive, then $\pi A \mid F(A \cap B)$ is proportional to $\pi B \mid F(A \cap B)$.

This does not entail that if $\pi B(A \cap B) > 0$ then $\pi A(A \cap B) > 0$, because the proportionality constant can be 0 (in which case πA gives 0 to all members of $F(A \cap B)$ – see further the discussion of [T7.3] which suggested this condition). It is easy to see what the constant of proportionality has to be:

[T10.1] If $\pi B(A)$ is defined and positive, then

$$\pi A \mid F(A \cap B)\colon\ \pi B \mid F(A \cap B) = \pi A(A \cap B)\colon\ \pi B(A \cap B).$$

Finally we define what it means for one of these functions to represent a two-place function:

DEFINITION. Model $M = \langle S, \pi\rangle$ with $S = \langle U, F\rangle$ represents binary function P iff the domain of P is F and for all A, B in F, $P(A \mid B) = \pi B(A \cap B)/\pi B(B)$ if defined, and $= 1$ otherwise.

It is easy to prove that:

[T10.2] If P is represented by a model, then P is a two-place probability measure.

Conversely, suppose that P is a two-place probability measure in the sense of satisfying I–II, defined on F in space $S = \langle U, F\rangle$. For all normal sets A of P define πA on FA by:

$$\pi A(B) = P(B \mid A).$$

That (M1) and (M2) are satisfied by $M = \langle S, \pi\rangle$ follows at once from the facts about normal sets. Suppose now, equivalently to the antecedent of (M3) that A and B are normal sets with $\pi(B) > 0$. To prove that (M3) holds, suppose that $\pi B(A)$ is defined and positive, so that B and A are normal sets, $P(A \mid B) > 0$. Then according to [T7.3], for each subset X of $A \cap B$ we have $P(X \mid A) = [P(B \mid A)/P(A \mid B)]P(X \mid B)$. Therefore here $\pi A(X) = [P(B \mid A)/P(A \mid B)]\pi B(X)$. In conclusion:

[T10.3] If P is a two-place probability measure satisfying the principles I–II, then P is represented by a model.

Having established this representation result, we now look for easily constructed models, for illustration, refutation of conjectures, and exploration of examples.

DEFINITION. Model $M = \langle S, \pi\rangle$ with $S = \langle U, F\rangle$ is *lexicographic* iff there is a sequence (well-ordered class) SEQ of 1-place probability measures defined on the whole of F, such that $\pi B(A) = q(A \cap B)/q(B)$ for the first member q of the sequence SEQ such that $q(B) > 0$; πB is undefined when there is no such q.

The members of SEQ correspond to the probabilities conditional on *belief remnants* (see discussion in Section 6 of Example 3). We will say that πA comes before πB in SEQ exactly when the first q in SEQ such that $q(A) > 0$ comes before the first q in SEQ such that $q(B) > 0$. It is easily checked that $M = \langle S, \pi\rangle$ is a model. Specifically, if A is a subset of B then πB will not come after πA, since whatever measure assigns a positive value to A will then assign one to B. Neither can πA come after πB if $\pi B(A) > 0$; in that case $\pi A = \pi B$. Consequently condition (M3) is easily verified: the proportionality constant $= 1$.

It is now very easy to make up examples of 2-place probability measures. Just take two or three or indeed any number, finite or infinite, of ordinary probability measures and well-order them. A special example, whose existence depends on the axiom of choice is this: let SEQ contain all one-place probability measures defined on given domain F. In that case, the only abnormal proposition is the empty set (the self-contradiction). Also the only *a priori* is the tautology. Short of this, we could of course have a sequence which does not contain literally all the definable probability measures, but contains all those which give 1 to a given set A. In that case, all propositions other than Λ that imply A are normal. Let us call P *Very Fine on A* in such a case. (The case of P Very Fine on U was already called "Very Fine" above.) Note that one of the defining conditions of a belief core K was that P had to be Very Fine on K.

11. BELIEF IN A LEXICOGRAPHIC MODEL

I will first show that in a lexicographic model, the intersection of all belief cores, if any, is always a belief core too. Since this does not depend on cardinality or the character of the sample space, the result adds significantly to the previous theorems. Then I will construct a lexicographic model to show that in general not all propositions with probability 1 are full beliefs. This model will be a reconstruction of Example 4 (Petra and Noah's Arc).

[T11.1] If P is represented by lexicographic model $M = \langle S, p \rangle$ defined by w.o. sequence SEQ, and A, B are disjoint normal sets for P, then the following are equivalent:
(i) $A\ P > B$ and not $B\ P > A$
(ii) πA comes before πB in SEQ.

[T11.2] If P is represented by lexicographic model $M = \langle S, \pi \rangle$ defined by w.o. sequence SEQ, and K, $K1$, $K2$ are belief cores with K a proper subset of $K1$ and $K1$ a proper subset of $K2$ then $\pi\ (K1 - K)$ comes before $\pi(K2 - K1)$ in SEQ

For proof note that $K1 - K$ and $K2 - K1$ are disjoint normal sets (*belief remnants*), so [T11.1] applies.

[T11.3] If P is represented by lexicographic model $M = \langle S, \pi \rangle$ defined by w.o. sequence SEQ, then the intersection of all its belief cores is also a belief core.

For convenience in the proof, call $\pi(K'' - K')$ a marker of K' when both are belief cores and K' is a proper subset of K''. This measure exists in the model since the set is normal. If no superset of K is a core, call $\pi(U - K)$ its marker. We now consider the whole class of markers; it must have a first member in SEQ. Let that be $p^* = \pi(K2 - K1)$. It remains to prove that K1 is the intersection of all the belief cores. Suppose to the contrary that core K is a proper subset of $K1$. Then by [T11.2] marker $p\ (K1 - K)$ comes before p^* in SEQ – contra the hypothesis. This ends the proof.

At this point we know enough about lexicographic models in general to exploit their illustrative uses. Recall Example 4. Petra has subjective probability 1 for N: the center of mass of Noah's Ark lies in the Northern Hemisphere and also for W: that it lies in the Western Hemisphere. I shall take it here that the Equator is part of the Northern Hemisphere; she has probability 1 that $-$EQ: it does not lie on the Equator. Let me add here that she also has probability 1 that $-$GR: it does not lie on the Greenwich Meridian (which I shall here take to be part of the Western Hemisphere). But she has probability 1 that it lies in Northern Hemisphere on the supposition that $N + W$: it lies either in the Northern or in he Western Hemisphere but not both (which supposition has 0 probability for her).

For the sample space, let U be the entire surface are of the Earth, and let F be the family of measurable subsets of U in the usual sense, so we can speak of area and length where appropriate. Let us first define some measures and classes of measures:

$$mA(X) = \text{ area of } X \cap A : \text{ area of } A$$

$$1A(X) = \text{ length of } X \cap A : \text{ length of } A$$

where XX is a subset of F

$$M(XX) = \text{ the class of all measures on } F$$

which give 1 to a member of XX.

The sequence SEQ is now pieced together from some other well ordered sequences as follows:

$$\text{SEQ} = m(N \cap W),\ l(\text{GR} \cap N),\ l(\text{EQ} \cap W),\ \text{S1},\ \text{S2},\ mN,\ mW,\ mU,\ \text{S3},\ \text{S4}$$

where the indicated subsequences are:

S1: a well-ordering of $\{1A$: A is a subset of $N \cap W$ with 0 area but positive length$\}$, S2: a well-ordering of $M\{A$ in F: A is a non-empty subset of $N \cap W$ with 0 area and 0 length$\}$,

S3: a well-ordering of $\{1A$: A is has 0 area but positive length$\}$,

S4: a well-ordering of $M\{A$ in F: A is a non-empty and has 0 area and 0 length$\}$.

Let us call the so constructed lexicographic model PETRA, in her honor. The construction is a little redundant: we would be constructing a Very Fine measure already if we just took the tail end mU, S3, S4. Preceding them with the others has the effect of establishing desired superiority relationships. For example, $N + W$ first receives a positive value from mN, which gives 0 to $W - N$, so that $P(N \mid N + W) = 1$. I made GReenwich similarly superior to EQuator.

[T11.4] $N \cap W$ is a belief core in PETRA.

[T11.5] No proper subset of $N \cap W - \mathrm{GR}$ is a belief core.

For let X be such a subset; to be a belief core it must have probability 1 tout court, so its area equals that of $N \cap W$. Let X' be one of its non-empty subsets with 0 area. Then X' as well as X itself are disjoint of $N \cap W \cap \mathrm{GR} = Y$. The first measure to assign positive value to $X' + Y$ is the second member of SEQ, namely $m(\mathrm{GR} \cap N)$, which assigns 1 to Y (because GR is part of W) and 0 to X'. Therefore X' is not superior to Y, and so X is not a belief core.

[T11.6] In PETRA some propositions not among its full beliefs have probability 1.

In view of the preceding, it suffices to reflect that $N \cap W - \mathrm{GR}$ has proper subsets with probability 1; for example $N \cap W - \mathrm{GR} - \mathrm{EQ}$.

APPENDIX

A1. *Previous literature*

The basic theory of two-place probability functions is a common part of a number of theories. Such probability functions have been called *Popper functions* because Popper's axioms originally presented in his *The Logic of Scientific Discovery* were adopted by other writers (see

Harper (1976), Field (1977), van Fraassen (1979, 1981a, 1981b)). Carnap used essentially the same axioms for his "c-functions", but concentrated his research on those which derive trivially from one-place probability functions ("m-functions"). Reichenbach's probability was also irreducibly two-place. I have mentioned De Finetti's paper (1936) which introduced the idea of local comparisons (like my "superior"; Vickers' "thinner"); see also Section 4.18 in his *Theory of Probability*, vol. 1. The most extensive work on two-place probability theory is by Renyi. The theory of two-place probability here presented is essentially as explored in my (1979), but with considerable improvement in the characterization of the described classes of models. Finally, the discussion of supposition in Section 2 is related to work on belief revision, much of it indebted to ideas of Isaac Levi; see Gardenfors 1988 for a qualitative version.

A2. *Transfinite consistency*

The ordering $P(A) \leqslant P(B)$ extends the partial ordering of logical implication: if $A \subseteq B$ then $P(A) \leqslant P(B)$. Unfortunately, the ordering $P(A) < P(B)$ does not extend in general the partial ordering of proper implication: $P(A) = P(B)$ is possible even when $A \neq B$. Indeed, this is inevitable if there are more than countably many disjoint propositions. As a corollary, the intersection of all propositions of maximal probability may itself even be empty. Kolmogoroff himself reacted to this problem by suggesting that we focus on probability algebras: algebras of propositions reduced by the relation of equivalence *modulo* differences of measure *zero*: $P(A + B) = 0$. (See Birkhoff (1967), XI, 5 and Kappos (1969), II, 4 and III, 3.)

The difficulty with this approach is that a probability algebra does not have the structure usually demanded of an algebra of propositions. For the latter, the notion of truth is relevant, so it should be possible to map the algebra homomorphically into $\{0, 1\}$. As example take the unit interval with Lebesgue measure, reduced by the above equivalence relation. This is a probability (sigma-)algebra. Let T be the class of elements designated as true, i.e. mapped into 1, and let A with measure x be in T. Then A is the join of two disjoint elements of measure $x/2$ each. Since the mapping is a homomorphism, one of these is in T. We conclude that T contains a countable downward chain $A_1, A_2, \ldots$ with the measures converging to *zero*. Therefore its meet is the *zero* element of the algebra. The meet should be in T because it is the countable meet of a family of "true" propositions; but it can't be in T, since the *zero* element is mapped into 0.

This "transfinite inconsistency" of the propositions which have probability *one*, was forcefully advanced by Patrick Maher (1990) as a difficulty for the integration of subjective probability and belief. My conclusion, contrary to Maher's, is that the role of subjective probability is to grade the alternatives left open by fully belief. That automatically bestows maximal probability on the full beliefs, but allows for other propositions to also be maximally probable. The question became then: how are the two classes of maximally probable propositions to be distinguished?

A3. *Rejection of the Bayesian paradigm?*

While I hold to the *probabilist* conviction that our opinion is to be represented by means of probability models, I reject many features associated with so-called "Bayesian" views in epistemology. In the present context, a main difference concerns the status of probability *one*. Conditionalization of an absolute (one-place) probability function cannot lower probability from *one* nor raise it from *zero*. As a result, such changes have often been relegated to epistemological catastrophes or irrational shifts of opinion. This is definitely not so in all probabilist work in epistemology (Isaac Levi and William Harper provide notable exceptions). In my view, probability *one* is easily bestowed, and as easily retracted, especially when it is *only* maximal unconditional probability (conditional on the tautology).

Obviously, then, I reject the naive Pascalian equation that a bet on A, with any payoff whatsoever, is worth to me my probability for A times that payoff. I think that Pascal's equation holds under restricted circumstances, with relevant assumptions kept fixed and in place. I mean this roughly in the sense of the "constructivist" view of subjective probability suggested in various ways by Glenn Shafer and Dick Jeffrey (and possibly meant by Dan Garber when he talks about the Bayesian hand-held calculator). In a given context I have a number of full beliefs which delimit the presently contemplated range of possibilities; it is the latter which I grade with respect to their comparative likelihood. The context may be anchored to a problem or type of problem, for which I go to this trouble. Some of the beliefs will indeed be "deepseated", and to some I subscribe so strongly that they would survive most any change of context. They are part of what I fall back on especially if I try to specify the context in which I am presently operating – for this involves seeing myself in a "larger" perspective.

A4. *Infinitesimals?*[10]

There is another solution on offer for most problems which two-place probability solves. That is to stick with one-place probability, but introduce infinitesimals. Any non-self-contradictory proposition can then receive a non-zero probability, though often it is infinitesimal (greater than zero but smaller than any rational fraction).

The infinitesimals solution is to say that all the non-self-contradictory propositions (that are not contrary to my full beliefs) receive not zero probability but an infinitesimal number as probability [in a non-standard model]. There is an important result, due to Vann McGee (1994) which shows that every finitely additive 2-place probability function $P(A \mid B)$ is the standard part of $p(A \cap B)/p(B)$ for some non-standard 1-place probability function p (and conversely). Despite this I see advantages to the present approach to conditional probability which eschews infinitesimals. First of all, there is really no such thing as "the" infinitesimals solution. If you go to non-standard models, there will in principle be many ways to generalize the old or ordinary concept of measure so as to keep agreement with measure in the embedded standard models. You don't have a specific solution till you specify exactly which reconstruction you prefer, and what its properties are.

In addition, in the present approach it is very easy to see how you can have "layering" of the following sort. Task: produce a two-place probability measure P such that for given A, B, C, the following is the case:

$$P(A) = 1,\ P(B \mid A) = 0,\ P(C \mid A) = 0$$
$$A,\ B,\ C \text{ are normal}$$

If $P(X \mid C) = P(X \mid B) = P(X \mid A) = 0$ then X is abnormal.

It is easy to construct a small lexicographic model in which this is the case. Let C be a subset of B and B a subset of A; let p_1 give 1 to A but 0 to B and to C; p_2 give 1 to B but 0 to C; and p_3 give 1 to C. If these are all the measures in the sequence, then subsets of C which receive probability 0 conditional on C are all abnormal. Intuitively it would seem that in the infinitesimal approach this would require the construction in which there are exactly two layers L and M of infinitesimals: x in L is infinitesimal in comparison to standard numbers, Y in M is infinitesimal in comparison (even) to any number in L, and no numbers at all are infinitesimal in comparison to numbers in M. I leave this as an exercise for the reader.

As to the problem of belief, I wonder if the nonstandard reconstruction would have the desirable features for which we naturally turn to infinitesimals. Suppose for example that I choose a model in which each non-empty set has a positive (possibly infinitesimal) probability. Then my full beliefs are not just those which have probability 1, since that includes the tautology only. On the other hand, I can't make it a requirement that my full beliefs have probability $> 1 - d$, for any infinitesimal d one could choose. For the intersection of the sets with probability $\geqslant 1 - d$ will generally have a lower probability. Hence the lottery paradox comes back to haunt us. We would again face the trilemma of either restricting full beliefs to the tautology, or specifying them in terms of some factor foreign to the degrees-of-belief framework, or banishing them from epistemology altogether.

NOTES

[1] This was essentially the model provided in Hintikka's book *Knowledge and Belief*. By propositions I mean the semantic content of statements; the same proposition can be expressed by many statements. I am not addressing how opinion is stored or communicated.

[2] This has been denied, e.g. by Henry Kyburg, and doubted, e.g. by Richard Foley (1993, Ch. 4).

[3] De Finetti (1936). I want to thank John M. Vickers for bringing this to my attention; De Finetti's idea is developed considerably further, with special reference to zero relative frequency, in Vickers (1988), Sections 3.6 and 5.4. The relation here

defined is slightly different from the so-named one in my (1979) – to which the name was somewhat better suited – for convenience in some of the proofs.

[4] See B. De Finetti (1972), Section 5.22.

[5] From this point on I shall drop the ubiquitous "for P" unless confusion threatens, and just write "a priori", "abnormal", etc. leaving the context to specify the relevant 2-place probability measure.

[6] Could a person's opinion be devoid of belief cores? Our definitions allow this, and it seems to me this case is related to the idea of a "Zen mind" which I have explored elsewhere (van Fraassen 1988).

[7] In earlier treatments of two-place probability this relationship has appeared as a special axiom: If $P(A \mid B) = P(B \mid A) = 1$ then $P(\cdot \mid A) = P(\cdot \mid B)$.

[8] As I have argued elsewhere (van Fraassen 1981a) this construction provides us with the "right" clue to the treatment of quantification and of intuitionistic implication in so-called probabilistic (or generally, subjective) semantics.

[9] In my (1981a), $P//A$ was designated as P^A and called "P conditioned on A." I now think this terminology likely to result in confusion, and prefer "P relativized to A."

[10] For related critiques of the 'infinitesimals' gambit, see Skyrms (1983, 1994), Hajek.

REFERENCES

Birkhoff, G. *Lattice Theory*, 3rd ed. Providence, 1967.

Blume, L., A. Brandenburger, and E. Deckel "Lexicographic probabilities and choice under uncertainty", *Econometrica* **59** (1991a), 61–79.

Blumme, L., A. Brandenburger, and E. Deckel "Lexicographic probabilities and equilibrium refinements", *Econometrica* **59** (1991b), 81–98.

De Finetti, Bruno "Les probabilités nulles", *Bulletin des sciences mathématiques* (1936), 275–288.

De Finetti, Bruno *Theory of Probability* (2 vols.). New York: John Wiley, 1972.

Field, Hartry "Logic, meaning and conceptual role", *Journal of Philosophy* **74** (1977), 374–409.

Foley, Richard *Working Without a Net.* Oxford: Oxford University Press, 1993.

Gardenfors, P. *Knowledge in Flux*: *Modeling the Dynamics of Epistemic States.* Cambridge, MA: MIT Press, 1988.

Hajek, Alan *"The conditional construal of conditional probability"*. Ph.D. Dissertation, Princeton University, 1992.

Harper, William, L. "Rational belief change, Popper functions, and counter-factuals", in C. Hooker and W. Harper (eds), *Foundations of Probability Theory*, vol. 1 (1976), 73–112.

Kappos, D. A. *Probability Algebras and Stochastic Spaces.* New York, 1969.

Maher, Patrick "Acceptance without Belief", *PSA* 1990, volume 1, 381–392.

McGee, Vann "Learning the impossible", in Skyrms (1994).

Popper, Karl *The Logic of Scientific Discovery.* New York: Basic Books, 1959.

Renyi, A. "On a new axiomatic theory of probability", *Acta Mathematica Hungarica* **6** (1955), 285–333.

Renyi, A. *Foundation of Probability.* San Francisco: Holden-Day, 1970.

Renyi, A. *Probability Theory.* Amsterdam: North-Holland, 1970.

Skyrms, Brian and Eells, Ellery (eds), *Probabilities and Conditionals: Belief Revision and Rational Decision.* Cambridge: Cambridge University Press, 1994.

Skyrms, Brian "Three ways to give a probability function a memory", pp. 157–161, in J. Earman (ed.), *Testing Scientific Theories* (Minnesota Studies in the Philosophy of Science, vol. X). Minneapolis: University of Minnesota Press, 1983.

Vickers, John, M. *Chance and Structure.* Oxford: Oxford University Press, 1988.

van Fraassen, Bas C. "Foundations of probability: a modal frequency interpretation", in G. Toraldo di Francia (ed.), *Problems in the Foundations of Physics* (1979), pp. 344–387.

van Fraassen, Bas C."Probabilistic semantics objectified: I. Postulates and logics", *Journal of Philosophical Logic* **10** (1981a), 371–394.

van Fraassen, Bas C."Probabilistic semantics objectified: II. Implications in probabilistic model sets", *Journal of Philosophical Logic* **10** (1981b), 495–510.

van Fraassen, Bas C. "Identity in intensional logic: Subjective semantics", pp. 201–219, in U. Eco *et al.* (eds), *Meaning and Mental Representation.* Bloomington: Indiana University Press, 1988.

Department of Philosophy,
1879 Hall,
Princeton University,
Princeton, NJ 08544-1006,
U.S.A.