COLIN HOWSON

THE LOGIC OF BAYESIAN PROBABILITY

For the last eighty or so years it has been generally accepted that the theory of Bayesian probability is a theory of partial belief subject to rationality constraints. There is also a virtual consensus that both the measure of belief and the constraints to which it is subject can only be provided via utility theory. It is easy to see why this should be so. The underlying idea, accepted initially by both de Finetti and Ramsey in their seminal papers ([1964] and [1931] respectively, though the paper 1964, first published in 1937, built on earlier work), but going back at least as far as Bayes' Memoir [1763], is that an agent's degree of belief in or uncertainty about a proposition A can be assessed by their rate of substitution of a quantity of value for a conditional benefit [S if A is true, 0 if not]. The natural medium of value is, of course, money, but the obvious difficulties with sensitivity to loss and the consequent diminishing marginal value of money seem to lead, apparently inexorably, to the need to develop this idea within an explicit theory of utility. This was first done only in this century, by Ramsey [1931]; today it is customary to follow Savage [1954] and show that suitable axioms for preference determine a reflexive and transitive ordering 'at least as probable as' and thence, given a further assumption about how finely the state space can be partitioned, a unique probability function.

The results of these various endeavours have all the hallmarks of a vigorously progressing research-programme. For all that, I do not myself think that it is the right way to provide a foundation for epistemic probability. I believe that the current state of utility theory itself is far from satisfactory, but underlying that concern is the feeling that one should not need a general theory of rational preference in order to talk sensibly about estimates of uncertainty and the laws these should obey. These estimates are *intellectual judgments*, and they are constrained by rules of consistency. In support of this view is an elementary mathematical fact seldom highlighted but of much significance. The probability of a proposition A is the expected value of its indicator function, that is the function defined on the space of relevant possibilities which takes the value 1 on those possible states of affairs that make A true, and 0 on the others. In other words, *probability is expected truth-value*. Truth and its logical neighbourhood are surely the right focus, not rationality (whatever that is).

These considerations suggest a view of mathematical uncertainty well-represented if not prominent among the seventeenth and eighteenth century pioneers, which is that the laws of epistemic probability are, in Leibniz's words, 'une nouvelle espèce de logique'[1]. And James Bernoulli, in the *Ars Conjectandi*, talked of measuring the probability of an uncertain proposition A in terms of the number of 'proofs' of A relative to the number of 'proofs' of not-A. The sort of logical

[1] 'I have more than once said that we should have a new kind of logic which would treat degrees of probability' (*New Essays*, bk. IV, ch. XVI).

analysis of probability proposed by Leibniz and Bernoulli was never properly developed, however. In retrospect it is easy to see the factors that hindered it: firstly, rapid technical development, and exploration of the problem-solving power of the mathematical theory, were primary, relegating 'philosophical' investigation to a relatively low-priority task; secondly, a satisfactory theory of *deductive* logic did not arrive until the beginning of the twentieth century; thirdly, probability became a puzzlingly equivocal notion, with two seemingly quite different aspects. These S-D. Poisson [1823] labelled respectively *chance*, signifying a property of events generated by repeatable random devices, measured by long-run frequencies (this is more or less Carnap's probability$_2$), and *probabilité*, signifying warranted degree of certainty relative to an agent's knowledge-state (cf. Carnap's probability$_1$).

The list of factors is not complete. Possibly the most influential, and one that seemed to many to prove decisive, was a principle additional to the standard probability axioms which was regarded as indispensable for actually determining probability-values. The principle is usually known now by the name Keynes (an advocate of it) gave it: the *Principle of Indifference*. Enough has been written about the difficulties and paradoxes attending the use of this principle (see, for example, [Howson and Urbach, 1993, Ch. 4] to make it unnecessary to undertake another long discussion here. But the connection between the principle and the logical programme is intimate, and when stated in the context of modern semantics it is very plausible. Ironically, therefore, it is also in such a setting that the basic problem with the principle is most easily identified and its gravity appreciated. Thus, suppose that a sentence B admits n models distinct up to isomorphism, in r of which A is also true. Then it seems obvious that there is a logical interpretation of a conditional probability $P(A|B)$ evaluated according to the rule: $P(A|B) = r/n$.[2] Such an interpretation is practically explicit in Bolzano [1850, Sec 66 et seq.], and fully explicit a century later in Carnap [1950], for whom this function (denoted by $c^\dagger$) explicated formally the idea of a *partial entailment* of A by B, measured by the proportion of B's models which are also models of A (though Carnap abandoned this measure almost immediately because of its inability to deliver a type of induction by enumeration; see [Howson and Urbach, 1993, Ch. 4].

Of course, there is a problem when B does not admit only finitely many models. In such cases it may still nevertheless be possible to exploit a 'natural' metric structure, such as when the possibilities are parametrizable by real numbers in a compact interval (such a structure was famously exploited by Bayes when he derived a conditional, 'posterior', probability distribution for a binomial parameter in his [1763]). However, it was in precisely this type of case that serious problems with the Principle of Indifference first became evident, with the discovery of what were called the 'paradoxes of geometrical probability', where 'geometrical'

[2]Laplace of course enunciated in effect just this rule when he defined the probability of an event to be the number of cases favourable to the event divided by the number of all possible cases, where these cases are 'equally possible'. The proviso has been much commented on, but in the semantic context its meaning is clear enough: that is certainly how Bolzano and the German school understood it later.

was the traditional word referring to the real-number continuum, or some compact subset thereof (the best-known of these 'paradoxes' is Bertrand's chord problem; see [Kac and Ulam, 1968, pp. 37–39]). This underlying problem is that how the elementary possibilities are conceived is not absolute but relative to some conceptual frame — in effect a *language* — and depending on how this is chosen the probability-values will themselves vary.

The mathematical subtleties of continuous possibility spaces rather conceal this point by suggesting that it is only in such spaces that real problems arise (still unfortunately a common point of view), so here is a very elementary example which shows that the problem lies right at the heart of the idea of taking Laplacean ratios to compute probabilities. Consider two simple first-order languages with identity and a one-place predicate symbol Q, and no other relation or function symbols. Language 1 has no individual names (constants), and language 2 has two individual names a, b. In both languages there are identical sentences

A: *Something has the property Q*

B: *There are exactly 2 individuals.*

There are only three models of B in language 1 distinguishable up to isomorphism: one containing no, one containing one and one containing two instances of Q. Two of these satisfy A. In language 2, on the other hand, the fact that the individuals can be distinguished by constants means that there are more than three distinct models of B: allowing for the possibility that the constants might name the same individual there are eight, six of which satisfy A. Using the Laplacean definition of $P(A|B)$ above we therefore get different answers for the value of $P(A|B)$ depending on which language we use. In language 1 the value is 2/3 and in Language 2 it is 3/4 (cf. Maxwell–Boltzmann vs. Bose–Einstein statistics). Relative to each language the models are of course all 'equally possible'.

To sum up: the 'equal possibilities' of the Principle of Indifference are equal relative to some conceptual frame, or language. This can in general be chosen in different ways and, depending on the choice, the ratios of numbers of favourable to possible cases *for the same event or proposition* may vary — where they can be computed at all. Moreover, not only does there seem to be no non-arbitrary way of determining the 'correct' language, or even being able to assign any meaning to the notion of a correct language, but in continuous spaces entirely equivalent frames will exist, related by one-to-one bicontinuous transformations, which will generate different probabilities.

This intractable knot was eventually untied, or rather cut, only in the last century, in the move from objectively to subjectively measured uncertainty, for with that move was jettisoned the Principle of Indifference, as inappropriate in a theory merely of consistent degrees of belief. Unfortunately there was no further systematic development of the idea that the probability axioms are no more than consistency constraints, at any rate within an explicitly logical setting: Ramsey went on to pioneer the development of subjective probability as a subtheory of utility theory (we shall see why shortly), and de Finetti employed the idea of penalties like Dutch Books to generate the probability laws. De Finetti's Dutch Book argument and its extension to scoring rules generally also lead in the wrong direction, of financial prudence in unrealistic circumstances (you always agree to take either side of a bet with any specified stake at your personal betting quotient). So despite its promising starting idea, from a strictly *logical* point of view Ramsey's and de Finetti's work represented a dead end. In what follows I propose to go back to the beginning, and combine the intuitions of Leibniz and J. Bernoulli with the conceptual apparatus of modern logic. We shall then get a 'rational reconstruction' of history closer to their intentions than to the actual course of events.

1 DEGREE OF BELIEF

There is a long tradition of informally expressing one's uncertainty about a proposition A in the odds which one takes to reflect the currently best estimate of the chances for and against A's being true. Mathematically, odds are a short step from probabilities, or at any rate the probability scale of the unit interval. The step is taken by normalising the odds, by the rule p = odds/(1+odds). p is called the *betting quotient* associated with those odds. The p-scale has the advantage as an uncertainty measure that it is both bounded and symmetrical about even-money odds (unlike the odds scale itself where the even-odds point, unity, is close to one end of the scale (0) and infinitely far from the other). Since the seventeenth century betting quotients as the measure of uncertainty have been called probabilities, and for the time being I shall do so myself (I am quite aware that they have not yet been shown to be probabilities in the technical sense: that will come later). Note that the inverse transformation gives odds $= p/(1-p)$.

To determine how such probabilities should be evaluated in specific cases was the function of the Principle of Indifference, a function which, as we have seen, it was unable to discharge consistently. However, abandoning the Principle, as Ramsey saw, seems to leave behind only *beliefs* about chances. This appears to signal a move into mathematical psychology, and particularly into measurement theory to develop techniques for measuring partial belief. Such at any rate was the programme inaugurated and partly carried out by Ramsey in his pioneering study 'Truth and Probability' [1931]. According to Ramsey the empirical data of partial belief are behaviourally-expressed preferences for rewards which depend on the outcomes of uncertain events: for example, in bets. Though Ramsey's idea seems to have a scope far beyond ordinary betting, as he pointed out we can always think in a general context of bets not just against human opponents but also against Nature. But preferences among bets will normally depend not only on the odds but also on the size of the stake: for large stakes there will be a natural disinclination to risk a substantial proportion of one's fortune, while for very small ones the odds will not matter overmuch. The only answer to this problem, Ramsey believed, was to give up the idea that odds, at any rate money odds, could measure uncertainty, and invoke instead a very general theory of rationally contrained preference: in other words, axiomatic utility theory.

But invoking the elaborate apparatus of contemporary utility theory, with its own more or less serious problems, seems like taking a hammer — and a hammer of dubious integrity and strength — to crack a nut. Why not simply express your beliefs by reporting the probabilities you feel justified by your current information, in the traditional way? The answer usually given is that to do so begs questions that only full-blown utility theory can answer. These relate to what these probabilities actually *mean*. They are odds, or normalised odds, so presumably they should indicate some property of bets at those odds. For example, I would be taken to imply that in a bet on A at any other odds than $p/(1-p)$, where p is my probability of A, I think one side of that bet would be positively disadvantaged, given what I know about A and the sorts of things that I believe make A more likely to be true than false — or not as the case may be. Thus my 'personal probability' (that terminology is due to Savage) determines what I believe to the *fair odds* on A: i.e. those odds which I believe give neither side an advantage calculable on the basis of my existing empirical knowledge. This is where the objections start.

The first is that your assessment of which odds, or betting quotients, do or do not confer advantage is a judgment which cannot be divorced from considerations of your own — hypothetical or actual — gain or loss and how you value these;

> for to ask which of two "equal" betters ask has the advantage is to ask which of them has the preferable alternative. [Savage, 1954, p. 63]

Granted that, we seem after all ineluctably faced with the task of developing a theory of uncertainty as part of a more general theory of preference, i.e. utility theory: precisely what it was thought could be avoided. But should we grant it? One hesitates to dismiss summarily a considered claim of someone with the authority of Savage, but nonetheless it simply is not true. Here is a simple counterexample (due in essence to [Hellman, 1997, p. 195]): imagine the bettors to be coprophiliacs and the stakes measures of manure. One's own preferences are *irrelevant* to judging fairness. They have only seemed relevant because gambles are traditionally paid in money and money is a universal medium of value.

Nevertheless, a Savagean objector might continue, to compute advantage you still need to know how the bettors themselves evaluate the payoffs, (a) in isolation and (b) in combination with what are perceived by those parties to be the chances of the payoffs; and both (a) and (b) may vary depending on the individual. For example, one party may be risk-averse and the other risk-prone, and a fair bet between such antagonists will be quite different from a fair bet between two risk-neutral ones. The answer to this is that the concept of advantage here is that of *bias*: on such a criterion a bet is fair simply if the money odds match what are perceived to be the fair odds, whatever the beliefs or values of the bettors. This is easily seen to imply an expected value criterion. For suppose R and Q are the sums of money staked, and that the odds measure of your uncertainty is $p : (1-p)$. The money odds $R : Q$ match your fair odds just in case $pQ = (1-p)R$, i.e.

(1) $pQ - (1-p)R = 0$

tells us that the bet is fair just when what is formally an expected value is equal to 0.

Now things do not look at all promising, however, for (1) seems to lead immediately to the notorious St Petersburg Problem. For those unfamiliar with it, this involves a denumerably infinite set of bets where a fair coin is repeatedly tossed and the bettor pays 1 ducat to receive 2^n ducats if the nth toss is the first to land heads, $n = 1, 2, \ldots$. Given no other information, the coin's assumed equal tendency to land heads or tails will presumably determine the fair odds. In that case the associated probability of getting the first head at the nth toss is 2^{-n}, and the expected value of each bet is clearly 0. Yet everyone intuitively feels that no bettor would be wise in accepting even a large finite number of these bets, let alone all of them — which would of course mean staking an infinite sum of money. The *inequity* of such bets, according to practically all commentators from Daniel Bernoulli onwards, is due to the diminishing marginal utility of money, and in particular the inequality in value between losing and gaining the same sum: the loss outweighs the gain, the more noticeably the larger the sum involved. In the St Petersburg game you are extremely likely to lose most of your 100 ducats if you accept the first 100 of the bets, a considerable sum to lose on a trifle. Your opponent could even less afford to pay out if the 100th bet won. Either way you would be silly to accept the bets even though they are fair by the criterion of money expectation. Nowadays it is taken for granted that the only solution to the problem is to use a utility function which is not only concave (like Daniel Bernoulli's logarithmic function) but also bounded above (unlike the logarithm).

I do not think we should worry too much about the St Petersburg Problem, for it begs a question in its turn, namely that a bet cannot be fair which it would be highly imprudent for one side to accept. But this is exactly what is being questioned, and is I think just false. A contract between a prince and a pauper is not unfair just because one can pay their debt easily and the other cannot. That is to confuse two senses of fairness: as lack of bias (the sense intended here), in which payoffs are balanced against probabilties according to (1), and as lack of differential impact those payoffs will have taking into consideration the wealth of the players. And indeed these quite distinct ideas have become confused in the Bayesian literature, to the extent that probability has become almost uniformly regarded as necessarily a subtheory of a theory of prudent behaviour.

The idea that the expected money-gain principle was vulnerable to the St Petersburg Problem was already challenged over two centuries ago by Condorcet, who pointed out that in repeated bets at odds $2^n - 1 : 1$ on heads landing first on the nth toss against, the average gain converges in probability to the expected value, i.e. 0, while, to quote Todhunter reporting Condorcet, "if any other ratio of stakes be adopted a proportional advantage is given to one of the players" [Todhunter, 1865, pp. 392–393]. Of course, this argument relies on (a) the full apparatus of probability theory (the quick modern proof would use Chebychev's Inequality) and

(b) assuming that the trials are uncorrelated with constant probability, neither of which assumption is appropriate here. But one doesn't need all that anyway: the moments equation (1) itself is a sufficient answer.

To sum up: the agent's probability is the odds, or the betting quotient, they currently believe fair, with the sense of 'fair' that there is no calculable advantage to either side of a bet at those odds. Despite opposing the widely accepted view that subjective probability can be coherently developed only within a theory of utility, this view is, I believe, quite unexceptionable, and certainly not vulnerable to what are usually taken to be decisive objections to it. Not only is it unexceptionable: it will turn out to deliver the probability axioms in a way that is both elegant and fully consonant with the idea that they are nothing less than conditions of consistency, and a complete set of such conditions at that.

2 CONSISTENCY

Now we can move on to the main theme of this paper. Ramsey claimed that the laws of probability are, with the probability function interpreted as degree of belief, laws of consistency and so of the species of logic. Unfortunately, as we saw, Ramsey then proceeded to divert the theory into the alien path of utility, where 'consistent' meant something like 'rational'. But rationality has nothing essentially to do with logic, except at the limits. Can we give the idea of consistent assignments of fair betting quotients an authentically logical meaning? The answer is that we can. We proceed in easy stages. A traditional sense of consistency for assignments of numbers is equation-consistency, or *solvability*. A set of equations is consistent if there is at least one single-valued assignment of values to its variables. The variables evaluated in terms of betting quotients are propositions. Correspondingly, we can say that an assignment of fair betting quotients is consistent just in case it can be solved in a analogous sense, the sense of being extendable to a single-valued assignment to all the propositions in the language determined by them (this is the notion of consistency Paris appeals to in a recent work on the mathematical analysis of uncertainty [Paris, 1994, p. 6]). But what, it might be asked, has the notion of consistency as solvability to do with logical consistency? Everything, it turns out. For *deductive consistency itself is nothing but solvability*. To see why, it will help to look at deductive consistency in a slightly different sort of way, though one still equivalent to the standard account, as a property not directly of sets of sentences *but of truth-value assignments*.

According to the standard (classical) Tarskian truth-definition for a first or higher-order language conjunctions, disjunctions and negations are homomorphically mapped onto a Boolean algebra of two truth-values, $\{T, F\}$, or $\{1, 0\}$ or however these elements are to be signified (here T or 1 signifies 'true' and F or 0 signifies 'false'). Now consider *any* attribution of truth-values to some set Σ of sentences of L, i.e. *any* function from Σ to truth-values. We can say that this assignment is consistent if it is capable of being extended to a function from the entire set of sentences of L to truth-values which satisfies those homomorphism constraints. For propositional languages the language of equation-solvability is sometimes explictly used: formulas can be regarded as representing boolean polynomial equations [Halmos, 1963, p. 8] in the algebra of two truth-values, and sets of them are consistent just in case they have a simultaneous solution.

The theory of 'signed' semantic tableaux or trees is a syntax perfectly adapted to seeing whether such equations are soluble and if so, finding all the solutions to them. ('Signing' a tableau just means appending Ts and Fs to the constituent sentences. The classic treatment is Smullyan [1968, pp. 15–30], a simplified account is in [Howson, 1997b].) Here is a very simple example:

$A \;\; T$
$A \rightarrow B \;\; T$
$B \;\; F$

The tree rule for $[A \rightarrow B \;\; T]$ is the binary branching

$F \; A \qquad B \; T$

Appending the branches beneath the initial signed sentences results in a *closed tree*, i.e. one on each of whose branches occurs a sentence to which is attached both a T and an F. A soundness and completeness theorem for trees [Howson, 1997b, pp. 107–111] tells us that any such tree closes if and only if the initial assignment of values to the three sentences $A, A \rightarrow B$ and B is inconsistent, i.e. unsolvable over L subject to the constraints of the general truth-definition.

To sum up: in deductive logic (semantic) consistency can be equivalently defined in the equational sense of a truth-value assignment being solvable, i.e. extendable to a valuation over all sentences of L satisfying the general rules governing truth-valuations. By a natural extension of the more familiar concept we can call such an extension a *model* of the initial assignment. Note that this sense of consistency does not pick out a different concept from the more usual one of a property of sets of sentences. Indeed, the two are essentially equivalent, as can be seen by noting that an assignment of truth-values to a set Σ of sentences is consistent in the solvability sense above just in case the set obtained from Σ by negating each sentence in Σ assigned F is consistent in the standard (semantic) sense.

We have become accustomed to understand by consistency deductive consistency and thereby something that exists, so to speak, only in a truth-centred environment. That this is not necessarily the case is now clear, for deductive consistency is seen to be merely an application of a much more general (and older) idea of consistency as solvability, having nothing necessarily to do with truth at all, but merely with assignments of values, not necessarily and indeed not usually truth-values, to variables in such a way that does not result in overdetermination.

What deductive and probabilistic consistency do have in common however is that the variables in question are propositional, and to proceed further we need to specify the language relative to which an assignment of fair betting quotients is solvable (if it is), subject to the appropriate constraints. For the sake of definitieness let us start, as in deductive logic, with a language relative to which the class of propositions will be determined. In fact, we can employ just the same sort of language, a first order language. Let L be one such, without identity. Let Ω = the class of structures $\Im$ interpreting the extralogical vocabulary of L. For any sentence A of L let $\mathrm{Mod}(A) = \{\Im : A \text{ is true in } \Im\}$. Let $F = \{\mathrm{Mod}(A)\text{: } A \text{ a sentence of } L\}$. Following Carnap [1971, pp. 35–37] F is the set of *propositions* of L. Note that F is a Boolean algebra isomorphic to the Lindenbaum sentence algebra of L [Paris, 1994, p. 34]. In fact, it will be better to work in a rather more extensive class of propositions, because F as it stands represents merely the propositions expressible by single sentences of L. But it is well-known that the mathematical theories incorporated in any minimally acceptable theory of physics, for example, are not expressible by single sentences of a first order language: they are not finitely axiomatisable (even the simplest of all mathematical theories, the theory of identity investigated by Euclid over 2000 years ago, is not finitely axiomatisable). The customary closing off of F under denumerably infinite unions (disjunctions) and intersections (conjunctions), generating the Borel field $B(F)$, (more than) allows such theories to be treated on a par with their finitely axiomatisable cousins.

In Ω and $B(F)$ we have two of the three ingredients of what mathematicians call a *probability space*. The third is a probability function defined on $B(F)$. Finding this will be the next task (in what follows I shall use $A, B, C, \ldots$ now to denote members of $B(F)$). The first step on the way is to determine the appropriate constraints on solutions of assignments of fair betting quotients. These will function like the purely *general* rules of truth in classical truth-definitions, as analytic properties of truth. In the context of fair betting quotients the constraints should presumably be analytic of the notion of fairness as applied to bets. At this point it is helpful to transform the payoff table

A	
T	Q
F	$-R$

into the well-known (betting-quotient, stake) 'coordinates' introduced by de Finetti in his seminal paper [1964]. The stake S is $R + Q$ and the betting quotient p^* is of course just R/S, and the table above becomes

A	
T	$S(1 - p^*)$
F	$-p^*S$

Where I_A is the indicator function of A, the bet can now be represented as a random quantity $S(I_A - p^*)$, and the equation (1) now transforms to

(1′) $$pS(1 - p^*) - p^*S(1 - p) = 0$$

where p is your fair betting quotient. Clearly, the left hand side is equal to 0 just when $p = p^*$, which is merely a different way of stating that a fair bet is one in which your estimate of the fair odds is identical with the money odds. Besides bets like the above there are also so-called *conditional bets*, i.e. bets on a proposition A which require the truth of some proposition B for the bet to go ahead: if B is false the bet on A is annulled. The bet is called a conditional bet on A given B. A betting quotient on A in a conditional bet is called a conditional betting quotient. A conditional bet on A given B with stake S and conditional betting quotient p clearly has the form $I_B S(I_A - p)$. If your uncertainty about A is registered by your personal fair betting quotient on A then your uncertainty, your conditional uncertainty, on A *on the supposition that* B *is true* will plausibly be given by your conditional fair betting quotient on A given B.

Let (F) be the set of formal constraints other than $0 \leq p \leq 1$ which fair betting quotients, including conditional fair betting quotients, should *in general* satisfy. This general content is contained in the claim that a fair bet is unbiased given the agent's own beliefs. These of course are unspecified, varying as they do from individual to individual. We can quickly infer

(a) If p is the fair betting quotient on A, and A is a logical truth, then $p = 1$; if A is a logical falsehood $p = 0$. Thus logical truth, logical falsehood and entailment relations correspond to the extreme values of fair betting quotients. Similarly if B entails A then the conditional betting quotient on A given B should be 1, and 0 if B entails the negation of A.

(b) Fair bets are invariant under change of sign of stake.

The reason for (a) is not difficult to see. If A is a logical truth (i.e. $A = \Omega$) and p is less than 1 then in the bet $S(I_A - p)$ with betting quotient p, I_A is identically 1 and so the bet reduces to the positive scalar quantity $S(1-p)$ received come what may. Hence the bet is not fair since one side has a manifest advantage. Similar reasoning shows that if A is a logical falsehood then p must be 0. Similar reasoning accounts for the conditions relating to entailment. As to (b), (1′) shows that the condition for a bet to be fair is independent both of the magnitude and sign of S.

But there is something else to (F) besides (a) and (b), a natural closure condition which can be stated as follows: *if a set of fair bets determines a bet on a proposition* B *with betting quotient* q *then* q *is the fair betting quotient on* B. What is the justification for this apart from 'naturalness'? It is well-known by professional bookmakers that certain combinations of bets amount to a bet on some other event, inducing a corresponding relationship between the betting quotients. For example, if A and B are mutually inconsistent then simultaneous bets at the same stake are extensionally the same as a bet on $A \vee B$ with that stake, and if p and q are the betting quotients on A and B respectively, we easily see that $S(I_A - p) + S(I_B - q) = S(I_{A\vee B} - r)$ if and only if $r = p + q$. Now add to this the thesis that if each of a set of bets gives zero advantage then the net advantage of anybody accepting all of them should also be zero (though this thesis is not *provable*, it seems so fundamentally constitutive of the ordinary notion of a fair game that we are entitled to adopt it as a desideratum to be satisfied by any formal explication; and, of course, when it is explicated as zero expected value within the fully developed mathematical theory we have the elementary theorem that expectation is a linear functional and hence all expectations, zero or not, add over sums of random variables). Putting all this together we obtain the closure principle above.

To proceed further, note that bets obey the following arithmetical conditions:

(i) $-S(I_A - p) = S(I_{\neg A} - (1-p))$.

(ii) If $A\&B = \bot$ then $S(I_A - p) + S(I_B - q) = S(I_{A\vee B} - (p+q))$.

(iii) If $\{A_i\}$ is a denumerable family of propositions in $B(F)$ and $A_i\&A_j = \bot$ and p_i are corresponding betting quotients and $\sum p_i$ exists then $\sum S(I_{A_i} - p_i) = S(I_{\vee A_i} - \sum p_i)$.

(iv) If $p, q > 0$ then there are nonzero numbers S, T, W such that $S(I_{A\&B} - p) + (-T)(I_B - q) = I_B W(I_A - p/q)$ (T/S must be equal to p/q). The right hand side is clearly a conditional bet on A given B with stake W and betting quotient p/q.

Closure tells us that if the betting quotients on the left hand side are fair then so are those on the right. The way the betting quotients on the left combine to give those on the right is, of course, just the way the probability calculus tells us that probabilities combine over compound propositions and for conditional probabilities.

Now for the central definition. Let Q be an assignment of personal fair betting quotients to a subset X of $B(F)$. By analogy with the deductive case, we shall say that Q is *consistent* if it can be extended to a single-valued function on all the propositions of L satisfying suitable conditions.

The final stage in our investigation is to generate interesting properties of consistency. If we suggestively signify a fair betting quotient on A by $P(A)$ closure tells us

(i′) $P(\neg A) = 1 - P(A)$.

(ii′) If $A\&B = \bot$ then $P(A \vee B) = P(A) + P(B)$.

(iii′) If $\{A_i\}$ is a denumerable family of propositions in $B(F)$ and $A_i\&A_j = \bot$ and $\sum P(A_i)$ exists then $P(\vee A_i) = \sum P(A_i)$.

(iv′) If $P(A\&B)$ and $P(B) > 0$ then $P(A|B) = P(A\&B)/P(B)$.

It might seem slightly anomalous that in (iv′) both $P(B)$ and $P(A\&B)$ should be positive, but it will turn out that only the former condition need be retained. At any rate, from the equations (i′)–(iv′) and (F) it is a short and easy step to proving the following theorem:

THEOREM 1. *An assignment Q of fair betting quotients (including conditional fair betting quotients) to a subset X of $B(F)$ is consistent (has a model) if and only if Q satisfies the constraints of the countably additive probability calculus; i.e. if and only if there is a countably additive probability function on $B(F)$ whose restriction to X is Q.*

Proof. The proof of the theorem is straightforward. Necessity is a fairly obvious inference from the closure property. It is not difficult to see that the condition that $P(A\&B)$ be positive in (iv′) can be jettisoned once we can assume invariance under sign of stake. Also, given finite additivity, the condition in (ii′) that $\sum P(A_i)$ exists is provable (using the Bolzano–Weierstrass Theorem). For sufficiency, all we really have to do is show that closure follows, but this is easy. For suppose P is a probability function on $B(F)$ and that X_i are bets on a finite or countable set of propositions A_i in which the betting quotients are the corresponding probabilities $P(A_i)$. Suppose also that $\sum X_i = S(I_B - q)$ for some proposition B. The expected value relative to P of each X_i is 0, and since expectations are linear functionals the expected value of the sum is also 0. Hence the expected value of $S(I_B - q)$ must be 0 also and so $q = P(B)$. So we have closure. ■

I pointed out earlier that there is a soundness and completeness theorem for trees (signed or unsigned), establishing an extensional equivalence between a semantic notion of consistency, as a solvable truth-value assignment, and a syntactic notion, as the openness of a tree from the initial assignment. In the theorem above we seem therefore to have an analogous soundness and completeness theorem, establishing an extensional equivalence between a semantic notion of consistency, i.e. having a model, and a syntactic one, deductive consistency with the probability axioms when the probability functor P signifies the fair betting quotients in Q. The deductive closure of the rules of the probability calculus is now seen as the complete theory of generally valid probability-assignments, just as the closure of the logical axioms in a Hilbert-style system is the complete theory of generally valid assignments of truth. My proposal is to take the theorem above as a soundness and completeness theorem for a logic of uncertainty, of the sort Leibniz seems to have had in mind when he called probability 'a new kind of logic'. To complete the discussion we must consider briefly what qualifies a discipline for the title 'logic'.

3 LOGIC (WHAT IS IT?)

Wilfrid Hodges's well-known text on elementary deductive logic tells us that 'Logic is about consistency' [Hodges, 1974]. This raises a question. Should 'consistency' just mean 'deductive consistency', or might there be other species of consistency closely kindred the deductive variety entitling their ambient theories to the status of logic or logics? It may well be the case that logic is about consistency without foreclosing the possibility of there being logics other than deductive. To answer these questions we first need an answer to the question 'What is logic?'.

My own belief is that there is no fact of the matter about what entitles a theory of reasoning to logical status, and one has to proceed as one does in extending common law to new cases, by appeal to precedent and common sense. Here again, of course, one must be selective, but with modern deductive logic in mind I propose — hesitantly — the following criteria for a discipline to be a logic:

(a) Its field is statements and relations between them.

(b) It adjudicates a mode of non-domain-specific reasoning in its field.

(c) Ideally it should incorporate a semantic notion of consistency extensionally equivalent to a syntactic one: it should have a *soundness and completeness theorem.* First order logic famously has a soundness and completeness theorem; so of course do many modal systems. Second order logic does not, but one could argue that that is the exception proving the rule, for it is largely for this reason that second order logic is generally regarded as not being a logic.

(a) and (b) are certainly satisfied by Bayesian, i.e. evidential, probability theory: any factual statement whatever can be in the domain of such a probability function. An interesting fact implicit in the discussion above is that the the statements involved here are not tied to any language: a Borel field does not of course have to be generated by the sets of model-structures of a first, or indeed any, order language. It can be completely language-free, and still nevertheless be regarded as a set of propositions, propositions in the most general sense of a class of subsets of a possibility-space, closed under the finite and countable Boolean operations. This creates a great deal of freedom, e.g. to assign probabilities to the measurable subsets of Euclidean n-space, a freedom not available in any of the classical logical languages..

As to (c), the theorem above establishes an extensional equivalence between a semantic notion, having a model, and a syntactic one (the probability calculus is a purely syntactic theory, of a function assigning values to elements of an algebra), and as such, I have claimed, is in effect a soundness and completeness theorem for Bayesian logic. The question is whether fulfilling (a)–(c) is sufficient to warrant the title 'logic'. It is of course quite impossible in principle to prove this, just as it is impossible in principle to prove the Church–Turing Thesis that the partial recursive functions exhaust the class of computable (partial) functions. In the latter case 'computable function', and in the former 'logic', have no precise mathematical definition. In addition, new theories of reasoning are increasingly marketed under the general title 'logic' as the information technology revolution gets under way (whereas even twenty years ago most logic texts were produced in university philosophy departments, now probably most are produced in computer science departments). Under any reasonable definition of 'theory of general reasoning' the rules of evidential probability would qualify as such, and hence as logic in this broad, liberal construal. But what has been shown is that a much tighter criterion, having an authentic semantics provably equivalent extensionally to an equally authentic syntax, applies to evidential probability in much the same sort of way that it applies to first order logic. Of course, it is open to anyone to deny that a completeness result is essential to a genuine logic; this is, of course, just what advocates of second order logic say (assuming that only full models are counted). Be that as it may, there seems little doubt that the completeness, i.e. axiomatisability, of first order logic has been a major factor in its widespread acceptance not only as logic but pretty well as *the* (classical) logic.

I have made consistency the focus of my discussion. It might well be objected that central to logic is the idea of *consequence*, or deduction. In a recent collection [Gabbay, 1994] dedicated to the discussion of what is to count as logic we find that this is a view common to almost all the authors. For example, 'logic is the science of deduction' [Hacking, 1994, p. 5]; 'a logical system is a pair $(\mid\!\sim, S_{\mid\sim})$ where $S_{\mid\sim}$ is a proof theory for $\mid\!\sim$ [$\mid\!\sim$ is a consequence relation]'. [Gabbay, 1994, p. 181]; 'Logic is concerned with what follows from what i.e. with logical consequence' [Aczel, 1994, p. 262]; and so on. I think that the reflections above show that this view, though widespread, is nevertheless incorrect. It arose because traditionally logic has been about conditions necessary and sufficient for the preservation of just one of the values in a two-valued system, the truth-value 'true'. In this sense, and indeed quite naturally, logic has traditionally been *deterministic*. It is true that there have been proposals for various sorts of many-valued logics, discrete and continuous, but even there the tendency has been to retain as far as possible something like a traditional concept of consequence. Even Adams's explicitly probabilistic system does this [Adams, 1998]. I believe that it is misguided because it is in effect a denial of the freedom such a mutli-valued system affords to get away from what is, I believe, nothing more than an artifact of two-valued systems. Of course, even in the account proposed here there is a consequence relation, but it is only the trivial one of *deductive* consequence from the probability axioms, telling us that if such and such are the probability-values of a specified set of propositions, then so and so is the probability of some other proposition. Williamson, discussing the account I have given above, points out that a relation of probabilistic consequence emerges naturally by analogy with the usual deductive notion of semantic consequence. A sentence A is a semantic consequence of a set Σ of sentences iff every model of Σ is a model of A. This transforms to: an assignment $r(A)$ is a consequence of an assignment $q(B_1), ..., q(B_n)$ iff every probability function extending q also extends r, i.e. iff every model, in the sense I have given above, of q is a model of r [Williamson, 2001]. But this does not generate any notion of consequence between A and $B_1, \ldots, B_n$ themselves. As Williamson notes, it generates a notion of probabilistic consequence, but only in the deductive sense above: $r(A)$ is a consequence of $q(B_1), \ldots, q(B_n)$ iff $P(A) = q(A)$ follows deductively from the probability axioms together with the 'assumption formulas' $P(B_1) = q(B_1), \ldots, P(B_n) = q(B_n)$.

Contemporary discussions of the relation between probability and formal deductive logic take a quite different approach to the one I regard as implicit in the theorem above. Some, e.g. [Gaifman, 1964; Scott and Krauss, 1970], take the logical aspect of probability to be exhausted by defining a probability function on the sentences of a suitable formal language, either a standard first order language or an infinitary one (as with Scott and Krauss), and showing how standard measure-theoretic arguments have to be correspondingly modified, in particular the extension theorem that states that a countably additive probability function on a field of sets has a unique countably additive extension on the Borel closure [Kolmogorov, 1956] p. 17). Gaifman provides an analogue of this for finitely additive probability functions defined on the class of sentences of a first order language with equality, showing that if a condition that has consequently come to be known as the Gaifman condition is satisfied then there is a unique extension from the quantifier-free sentences to all the sentences of the language (the Gaifman condition states that the supremum of the probabilities of all the disjunctions of instances of a formula is equal to the probability of its existential quantification; in terms of the Lindenbaum algebra the Gaifman condition is that probabilities commute with suprema).

Others, like Fagin and Halpern [1988], and in a different way Heifetz and Mongin [forthcoming], incorporate probability into the *syntax* of a formal language. 'Pulling down' probability into the object language is of course very much in the spirit of modal logic, and indeed Heifetz and Mongin introduce what they call a modal operator, 'the probability is at least α that ...', for all rational α in [0,1], which they interpret according to a generalisation of Kripke semantics incorporating a probability distribution over a class of possible worlds. What distinguishes my own account most of all from the modal one(s) and the others that I have mentioned is that they take the probability axioms as pretty much given: in Gaifman and Scott and Krauss probability is just a weakening of the usual two-valued semantic valuation, while in Heifetz and Mongin the axioms are, as in Savage, derivative from a set of rationality constraints over preferences. I believe I have shown in the foregoing that the probability axioms for epistemic probability are naturally, if not inescapably, interpreted as being of the same general species of consistency constraint as the rules of deductive logic itself, and to that extent *intrinsically* logical in nature.

We now move to consider some collateral benefits accruing from construing the (epistemic) probability calculus as a species of logical axioms. Where a title, like 'logic', is already uncontroversially bestowed elsewhere, discharging some genuine explanatory function should be a condition of its extension to a putative new case. As we shall see below, there are in addition to (a)–(c) above other interesting points of contact, or near-contact, with classical deductive logic, and a logical understanding of the rules of probability will, I hope, be seen to bring with it a considerable explanatory bonus.

The following topics in Bayesian probability are all regarded as problematic to some extent or other: countable additivity; strict coherence versus coherence;

rationality; completeness; conditionalisation; sharp versus fuzzy or interval-valued probability; inductive inference; penalties for infringing the rules. The logical view at the least usefully illuminates these and at best solves them. I shall deal with them in turn.

3.1 *Countable Additivity*

There has been a good deal of controversy concerning the status of the principle of countable additivity within the theory of epistemic probability. Most workers in the field, including, famously, de Finetti himself, reject it, while a much smaller number accept it. I do not think that it is necessary to go into the details of the protagonists' arguments. The theorem above shows that it must be adopted within any adequate view of the rules of probability as consistency constraints. The fact is that if we wish to assign probabilities as widely as possible then consistency in their assignment over compound propositions can then be guaranteed only by adding countable additivity to the axioms.

3.2 *Rationality*

I started out by remarking that the recent history of subjective probability has tended to neglect the logical aspect identified by Leibniz, favouring instead a rationality interpretation of the constraints as prudential criteria of one type or another. The trouble with adopting this line is that it is very difficult to *demonstrate* in any uncontroversial and non-question-begging way that violation of any of the constraints is positively irrational. Take the requirement of transitivity for preferences, for example: it is not evident that certain types of intransitive preference are necessarily irrational, especially when it is considered that the comparisons are always pairwise (for a counterexample see Hughes [1981]). The logical view, on the other hand, need not in principle be troubled by links with rationality of only doubtful strength, since logic is not *about* rational belief or action as such. Thus, deductive logic is about the conditions which sets of sentences must satisfy to be capable of being simultaneously true (deductive consistency), and the conditions in which the simultaneous truth of some set of sentences necessitates the truth of some given sentence (deductive consequence): in other words, it specifies the conditions regulating what might be called consistent truth-value assignments. This objectivism is nicely paralleled in the interpretation of the probability axioms as the conditions regulating the assignment of consistent fair betting quotients.

3.3 *Completeness*

Under the aspect of logic the probability axioms are as they stand complete. Hence any extension of them — as in principles for determining 'objective' prior probabilities — goes beyond pure logic. This should come as something of a relief: the principles canvassed at one time or another for determining 'objective' priors have

been the Principle of Indifference, symmetry principles including principles of invariance under various groups of transformations, simplicity, maximum entropy and many others. All these ideas have turned out to be more or less problematic: at one extreme inconsistent, at the other, empty. It is nice not to have to recommend any.

3.4 *Coherence versus Strict Coherence*

A hitherto puzzling qustion posed first by Shimony [1955] and then repeated by Carnap [1971, pp. 111–114] is easily answered if we accept that the probability axioms are laws of consistency. Consider a set of bets on n propositions $A_1, ..., A_n$ with corresponding betting quotients p_i. The classic Dutch Book argument shows that a necessary and sufficient condition for there being, for every set of stakes S_i, a distribution of truth-values over the A_i such that for that distribution there is a non-negative gain to the bettor (or loss: reverse the signs of the stakes), is obedience to the probability axioms. However, if we substitute 'positive' for 'non-negative' we also get an interesting result: the necessary and sufficient condition now becomes that the probability function is in addition *strictly positive*, i.e. it takes the value 0 only on logical falsehoods. Which of these two Dutch Book arguments should we take to be the truly normative one: that we should always have the possibility of a positive gain, or that we should always have the possibility of a non-negative gain? It might seem that the second is the more worthwhile objective: what is the point of going to a lot of trouble computing and checking probability values just to break even? On the other hand, strictly positive probability functions are very restrictive. There can be no continuous distributions, for example, so a whole swathe of standard statistics seems to go out of the window. There does not seem to be a determinately correct or incorrect answer to the question of what to do, which is why it is a relief to learn that the problem is purely an artifact of the classic Dutch Book argument. Give up the idea that the probability laws are justified in those terms and the problem vanishes. Indeed, we now have a decisive objection to specifying *any* conditions additional to the probability axioms: the laws of probability as they stand are complete.

3.5 *Unsharp Probabilities*

We seldom if ever have personal probabilities, defined by Bayes' procedure of evaluating uncertain options, which can be expressed by an exact real number. My value for the probability that it will rain some time today is rather vague, and the value 0.7, say, is no more than a very rough approximation. In the standard Bayesian model the probability function takes real-number values. But if we are trying to use the model to understand agents' actual cognitive decisions it would seem useful if not mandatory to assume that they have more or less diffuse probabilities — because they mostly if not invariably do in the real world.

If I am correct then probabilistic and deductive models of reasoning are intimately related, suggesting that considerations which prove illuminating in one can be profitably transferred, *mutatis mutandis*, to the other. So ask: what corresponds in deductive models to consistent probability-values? Answer: truth-values. Well, deductive models, or at any rate the standard ones, equally fail to be realistic through incorporating 'sharp' truth-values, or what comes to the same thing, predicates having sharp 'yes'/'no' boundaries. Thus, it is assumed in standard deductive logic that for each predicate Q and individual a in the domain, a definitely has Q or it definitely does not. An equivalent way of stating the assumption is in terms of *characteristic functions*: the characteristic function of Q is a function f_Q on the domain of individuals such that for each a, $f_Q(a) = 1$ (i.e. a has Q) or $f_Q a) = 0$ (a does not have Q). No intermediate values are permitted. And this apparatus is used to model reasoning in natural languages which by nature are highly unsharp, except in the very special circumstances when technical vocabularies are employed. There are actually good functional reasons why natural predicates are not sharp: their flexibility in extending beyond know cases is an indispensable feature of the adaptive success of natural languages. Not surprisingly the modelling of these languages by artificially sharp ones results in 'paradoxes' of the Sorites type (whose classic exemplar is the Paradox of the Heap).

Such unpalatable results have prompted the investigation of more accurate methods of modelling informal deductive reasoning by means of 'vague' predicates, and in particular the use of so-called 'fuzzy' ones, where {0,1}-valued characteristic functions are replaced by continuous functions, with appropriate rules for their use. The analogue for blunting 'sharp' probability values is to replace them with unsharp, interval-valued ones, and the theory of these is well-understood (see [Walley, 1991]). But 'sharp' probability models find their justification in investigations of how evidence, in the form of reports of observations, should affect estimates of the credibility of hypotheses. It is quite difficult to answer this and related questions, e.g. how sensitive y is to x where the data are particularly numerous or varied or both, without using a theory that can say things like 'Suppose the prior value of the probability is x', and then use the machinery of the point-valued probability calculus (in particular Bayes' Theorem) to calculate that the posterior value is y. So we need a fairly strong theory which will tell us things like this; and in the standard mathematical theory of probability we have a very rich theory indeed. At the same time, the model is not too distant from reality; it is quite possible to regard it as a not-unreasonable approximation in many applications, for example where the results obtained are robust across a considerable range of variation in the probability parameters. Many of the limiting theorems in particular have this property.

Similar sorts of considerations apply to the usual formal models of deductive reasoning. There are non-sharp models, but it is partly the sharpness itself of the more familiar structures that explains why they still dominate logical investigations: nearly all the deep results of modern logic, like the Completeness Theorem for the various formulations of first order logic, and the limitative theorems of Church, Gödel, Tarski etc., are derived within 'sharp' models. Much more could be written on this subject, but space and time are limited and enough has, I hope, now been said to convey why sharp models are not the unacceptable departures from a messier reality that at first sight they might seem to be.

3.6 *Conditionalisation*

If the rules of the probability calculus are a complete set of consistency constraints, what is the status of conditionalisation, which is not one of them, though it is standardly regarded as a 'core' Bayesian principle? Recall that conditionalisation is the rule that

$$\frac{P(B|A) = r \quad P'(A) = 1}{P'(B) = r}$$

where '$P'(A) = 1$' signifies an exogenous 'learning' of A; P is your probability function up to that point, and P' after. There is a well-known Dutch Book argument for this rule due to David Lewis (reported in [Teller, 1973]). I have given detailed reasons elsewhere [Howson and Urbach, 1993, Ch. 6] why I believe any Dutch Book argument in the 'dynamic' context to be radically unsound and I shall not repeat them all here. What I will do is show how consideration of a corresponding, obviously unsound, deductive analogue enables us to translate back and see why the probabilistic dynamic rule should in principle be unsound too. Consider first a possibly slightly unfamiliar — though sound — version of modus ponens, where v is a truth-valuation and $r = 1$ (true) or 0 (false):

$$(2) \quad \frac{v(A \to B) = r \qquad v(A) = 1}{v(B) = r}.$$

But now suppose v and w are distinct and consider

$$\frac{v(A \to B) = r \qquad w(A) = 1}{w(B) = r}.$$

This 'dynamic' version of (2), where v and w represent earlier and later valuations, is clearly invalid. Indeed, suppose A says that B is false, i.e. $A = \neg B$, and you now (v) accept $\neg B \to B$ (i.e. $v(\neg B \to B) = 1$): it means accepting B), but later (w) accept $\neg B$ (i.e. $w(\neg B) = 1$). If you try to 'conditionalise' and infer B (i.e. $w(B) = 1$) you will obviously be inconsistent.

Here is a probabilistic analogue of that counterexample. Let A say not that B is false but that B is going to be less than 100% probable: A = "$P'(B) < 1$", where P' is your probability function at some specified future time t (e.g. you are imagining that you will be visited by Descartes's demon). Further, suppose $P(B) = 1$, and suppose $P(A) > 0$. If you are a consistent probablistic reasoner $P(B|A) = 1$. But suppose at t it is true that $P'(B) < 1$, and you realise this by introspection (i.e. $P'(A) = 1$). If you try to conditionalise on A and infer

$P'(B) = P(B|A)$ you will be inconsistent. In the deductive 'dynamic' modus ponens, only if $v(A \to B) = w(A \to B)$ can you validly pass from $w(A) = 1$ and $v(A \to B) = r$ to $w(B) = r$ ($r = 0$ or 1): i.e.

$$(3) \quad \frac{w(A \to B) = v(A \to B) \qquad w(A) = 1}{w(B) = v(A \to B)}$$

which is just a substitution instance of (2). This suggests the analogous rule

$$(4) \quad \frac{P'(B|A) = P(B|A) \qquad P'(B) = 1}{P'(B) = P(B|A)}$$

Indeed this rule is valid, but you don't need a new rule to tell you so: *it already follows from the probability calculus*! In other words, conditionalisation is only conditionally valid, and the conditions for validity are already supplied by the probability calculus. Similar considerations apply to 'probability kinematics'.

The rather striking similarities between (3) and (4) above suggest a further point of contact between Bayesian probability and deductive logic. Do not these rules imply that P(B|A) is the probability of a type of sentential conditional? But, as is well-known, this question seemed to have been answered firmly in the negative by David Lewis with his so-called 'triviality theorem' [Lewis, 1973]. In spite of the apparent finality of the answer there have been attempts to bypass it via various types of non-Boolean propositional logics incorporating a conditional B⇒A satisfying the so-called Adams Principle $P(B|A) = P(A \Rightarrow B)$ (from [Adams, 1975]). Long before Lewis's result de Finetti had suggested a non-Boolean, three-valued logic of conditionals [de Finetti, 1964, Section 1] satisfying that condition, with values 'true', 'false'and 'void', where a conditional is 'void' if the antecedent is false (cf. called-off bets). This is not the place for a discussion of these attempts and I refer the reader to Milne [1997] for a relatively sympathetic account of this and other theories of non-Boolean conditionals, and to Howson [1997a] for a somewhat less sympathetic discussion.

3.7 Inductive Inference

This is a vast topic and I can only sketch here the account I have given at length elsewhere [Howson, 2000, Ch. 8]. The logical view of the principles of subjective probability exhibits an extension of deductive logic which still manages to remain non-ampliative. It therefore also respects Hume's argument that there is no sound inductive argument from experiential data that does not incorporate an inductive premise, and it also tells us what the inductive premise will look like: *it will be a probability assignment that is not deducible from the probability axioms*. Far from vitiating Bayesian methodology I believe this Humean view, which arises naturally from placing epistemic probability in an explictly logical context, strengthens it against many of the objections commonly brought against it (see Howson *loc. cit.*).

3.8 Sanctions

A logical view of the probability axioms is all very well, but what sanction attaches to infringing them on this view? Indeed, doesn't the analogy with deductive logic break down at precisely this point? There is after all an obvious sanction to breaking the rules of deductive consistency: what you say cannot be true if it is inconsistent. Perhaps surprisingly, however, there seems to be as much — or as little — sanction in both cases. In the probabilistic one there certainly are sanctions: not just the (usually remote) theoretical possibility of being Dutch-Booked were you to bet indiscriminately at inconsistent betting quotients (note that there is no presumption in the earlier discussion that you will and certainly not that you ought to bet at your fair betting quotients, or indeed that you ought to do anything at all), but those arising in general from accepting fallacious arguments with probabilities: we have only to look at the notorious Harvard Medical School Test to see what these might be (see [Howson, 2000, pp. 52–54]). Morover, probabilistic inconsistency is as self-stultifying as deductive: as we saw in the previous chapter, inconsistency means that you differ *from yourself* in the uncertainty value you attach to propositions, just as deductive inconsistency means that you differ from yourself in the truth-values you attach to them.

As to the positive sanctions penalising deductive inconsistency, these on closer inspection turn out to be less forceful than they initially seemed. Here is Albert rehearsing them:

> Logical [deductive] consistency serves a purpose. Beliefs cannot possibly be true if they are inconsistent. Thus, if one wants truth, logical consistency is necessary. An analogous argument in favor of Bayesianism would have to point out some advantage of coherence unavailable to those relying on non-probabilistic beliefs and deductive logic alone. Such an argument is missing. [Albert, 2001, p. 366]

The two principal assertions here are both incorrect. Firstly, logical consistency is *not* necessary for truth. False statements are well known to have true consequences, lots of them, and inconsistent statements the most of all since every statement follows from a contradiction. Current science may well be be inconsistent (many distinguished scientists think it is), but it has nevertheless provided a rich bounty of truths. The benefits of maintaining deductive consistency are more complex and less direct than is often supposed, but the principal penalty is the same in both the deductive and the probabilistic case: inconsistency amounts to evaluating the same proposition in incompatible ways. It is self-stultifying. More generally, the practical penalties attaching to deductive inconsistency are not obviously greater than those attaching to probabilistic inconsistency. What practical consequences flow from either will vary with one's situation.

4 CONCLUSION

In the foregoing pages I have tried to carry through Leibniz's programme for understanding the rules of evidential probability as a species of logic as authentic as those of deductive logic. In both cases these rules are conditions of solvability of value-assignments, with similar completeness properties. I have also argued that this leads to a quite different, and much more coherent and fruitful, view of Bayesian probability than the usual one which is as a theory of prudential rationality. One thing that Lakatos's well-known theory of scientific research programme emphasises is that a programme may be almost written off yet eventually come back to win the field (see Lakatos [1970]). It is early perhaps to predict that this will happen with the logical programme, but I hope I have shown that its resources are more than adequate to the task.

ACKNOWLEDGEMENT

I would like to thank Oxford University Press for allowing me to reproduce passages from *Hume's Problem: Induction and the Justification of Belief*, C. Howson, 2000.

Department of Philosophy, Logic and Scientific Method, London School of Economics, London, UK.

BIBLIOGRAPHY

[Aczel, 1994] P. Aczel. Schematic consequence. In [Gabbay, 1994, pp. 261–273].
[Adams, 1975] E. W. Adams. *The Logic of Conditionals*, Reidel, Dordrecht, 1975.
[Adams, 1998] E. W. Adams. *A Primer of Probability Logic*, CSLI, Stanford, 1998.
[Albert, 2001] M. Albert. Bayesian learning and expectations formation: anything goes. This volume, pp. 347–368.
[Anscombe and Aumann, 1963] F. J. Anscombe and R. J. Aumann. A definition of subjective probability, *Annals of Mathematical Statistics*, **34**, 199–205, 1963.
[Bayes, 1763] T. Bayes. An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, 1763.
[Bernoulli, 1715] J. Bernoulli. *Ars Conjectandi*, Basel, 1715.
[Bolzano, 1850] B. Bolzano. *Theory of Science*, 1850
[Carnap, 1971] R. Carnap. A basic system of inductive logic. In *Studies in Inductive Logic and Probability*, Volume I, R. Carnap and R. C. Jeffrey, eds. pp. 33–167. University of California Press, 1971.
[Carnap, 1950] R. Carnap. *The Logical Foundations of Probability*, Chicago: University of Chicago Press, 1950.
[Couturat, 1901] L. Couturat. *La Logique de Leibniz*, Paris, 1901.
[de Finetti, 1964] B. de Finetti. Foresight; its Logical Laws, its Subjective Sources', *Studies in Subjective Probability*, H. Kyburg and H. Smokler, eds. pp. 93–159. Wiley, 1964. (de Finetti's paper was published originally in 1937 in French.)
[Fagin and Halpern, 1988] R. Fagin and J. Y. Halpern. Reasoning about Knowledge and Probability: Preliminary Report. In *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, M. Y. Vardi, ed. pp. 277–293. Morgan Kaufmann, 1988.
[Gabbay, 1994] D. M. Gabbay, ed. *What is a Logical System*?, Oxford: Oxford University Press, 1994

[Gabbay, 1994a] D. M. Gabbay. What is a logical system? In [Gabbay, 1994, pp. 179–217].
[Gaifman, 1964] H. Gaifman. Concerning measures in first order calculi. *Israel Journal of Mathematics*, **2**, 1–18, 1964.
[Hacking, 1994] I. Hacking. What is logic? In [Gabbay, 1994, pp. 1–35].
[Halmos, 1963] P. Halmos. *Lectures on Boolean Algebras*, Van Nostrand, Princeton,, 1963.
[Heifetz and Mongin, forthcoming] A. Heifetz and P. Mongin. The Modal Logic of Probability, forthcoming.
[Hellman, 1997] G. Hellman. Bayes and beyond. *Philosophy of Science*, **64**, 190–205, 1997.
[Hodges, 1974] W. Hodges. *Logic*, Harmondsworth: Penguin Books, 1974.
[Howson, 1997a] C. Howson. Logic and probability', *British Journal for the Philosophy of Science*, **48**, 517–531, 1997.
[Howson, 1997b] C. Howson. *Logic With Trees*, London: Routledge, 1997..
[Howson, 2000] C. Howson. *Hume's Problem: Induction and the Justification of Belief*, Oxford: Oxford University Press, 2000.
[Howson and Urbach, 1993] C. Howson and P. Urbach. *Scientific Reasoning: the Bayesian Approach*, 2nd edition, Chicago: Open Court, 1993.
[Hughes, 1980] R. I. G. Hughes. Rationality and intransitive preferences, *Analysis*, **40**, 132–134, 1980.
[Kac and Ulam, 1968] M. Kac and S. Ulam. *Mathematics and Logic*, New York: Dover, 1968.
[Kolmogorov, 1956] A. N. Kolmogorov. *Foundations of the Theory of Probability*, New York: Chelsea, 1956.
[Lakatos, 1970] I. Lakatos. Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*, I. Lakatos and A. Musgrave, eds. pp. 91–197. Cambridge: Cambridge University Press, 1970.
[Lewis, 1973] D. Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, vol. LXXXV, 297–315, 1973.
[Milne, 1997] P. M. Milne. Bruno de Finetti and the logic of conditional events. *British Journal for the Philosophy of Science*, **48**, 195–233, 1997.
[Paris, 1994] J. Paris. *The Uncertain Reasoner's Companion. A Mathematical Perspective*, Cambridge: Cambridge University Press, 1994.
[Poisson, 1823] S.-D. Poisson. *Recherches sur la probabilité des jugements en matière civile et en matière criminelle*, Paris, 1823.
[Ramsey, 1931] F. P. Ramsey. Truth and probability. In *The Foundations of Mathematics*, R.B. Braithwaite, ed. London: Kegan Paul, 1931.
[Savage, 1954] L. J. Savage. *The Foundations of Statistics*, New York: Wiley, 1954.
[Scott and Krauss, 1970] D. Scott and P. Krauss. Assigning probabilities to logical formulas. In *Aspects of Inductive Logic*, J. Hintikka and P. Suppes, eds. pp. 219–264, 1970.
[Shimony, 1955] A. Shimony. Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, **20**, 1–28, 1955.
[Smullyan, 1968] R. Smullyan. *First Order Logic*, New York: Dover, 1968.
[Teller, 1973] P. Teller. Conditionalisation and Observation, *Synthese*, **26**, 218–58, 1973.
[Todhunter, 1865] I. Todhunter. *A History of the Mathematical Theory of Probability*, Cambridge and London, 1865.
[Walley, 1991] P. Walley. *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall, 1991.
[Williamson, 2001] J. Williamson. Probability logic. In 'Handbook of the Logic of Inference and Argument: The Turn Toward the Practical', D. Gabbay, R. Johnson, H.J. Ohlbach & J. Woods, eds. Elsevier, 2001.