J. FRANKLIN

RESURRECTING LOGICAL PROBABILITY

ABSTRACT. The logical interpretation of probability, or "objective Bayesianism" – the theory that (some) probabilities are strictly logical degrees of partial implication – is defended. The main argument against it is that it requires the assignment of prior probabilities, and that any attempt to determine them by symmetry via a "principle of insufficient reason" inevitably leads to paradox. Three replies are advanced: that priors are imprecise or of little weight, so that disagreement about them does not matter, within limits; that it is possible to distinguish reasonable from unreasonable priors on logical grounds; and that in real cases disagreement about priors can usually be explained by differences in the background information. It is argued also that proponents of alternative conceptions of probability, such as frequentists, Bayesians and Popperians, are unable to avoid committing themselves to the basic principles of logical probability.

Dennett's *Philosophical Lexicon* contains the definition:

Hume *v.* to commit to the flames, bury, or otherwise destroy a philosophical position. Hence, **exhume**, to revive a position generally believed to be humed.

What more appropriate as a candidate for exhumation than the position that once seemed to promise hope of answering Hume on induction, and providing an objective foundation for statistical inference generally, the thesis that there is a pure *logic* of probability?

On the surface, at least, it is natural to distinguish between *factual* or *stochastic* probability on the one hand, and *logical* or *epistemic* probability on the other. The statement:

> The probability that this die will come up 5 on the next throw is $\frac{1}{6}$

appears to express a physical fact about the die, or the manner of its throwing. By contrast,

> The probability that the big bang hypothesis is true, given present evidence, is high

seems to be a matter of evidential, or logical, relations between propositions.

There have been three main schools of thought as to the nature of logical or epistemic probability. According to the frequentists, it does not exist, and one can get by with only factual probability (whether factual probability itself is interpreted as relative frequencies, limits of relative frequencies, or propensities). The Bayesian (or more fully "subjective Bayesian") school holds that "logical" probability does exist, but is an essentially subjective matter of different agents' degrees of belief in propositions – subject however to the "consistency" requirements embodied in the sum and product laws of probability. The school of logical probabilists, including Keynes and Carnap, held that logical probability was genuinely logic, in the sense of an objective logical relation between propositions, a kind of "partial entailment". They held that there was a number between 0 and 1, denoted $P(h|e)$, which measures the objective degree of logical support that evidence e gives to hypothesis h. The theory has an obvious initial attractiveness, in explaining the reasonable level of agreement found in at least straightforward cases when scientists, juries, actuaries and so on evaluate hypotheses in the light of evidence.

1. THE ARGUMENT FOR THE IMPOSSIBILITY OF OBJECTIVE PRIORS

The word "held" in the last paragraph is in the past tense because logical probabilists have almost disappeared from the face of the earth. Although there is a small school of "objective Bayesian" or "maximum entropy Bayesian" statisticians and physicists which maintains the tradition, (e.g. Jaynes, to appear; Heidbreder 1996) in both the philosophical and the statistical communities the position is generally regarded as dead. Its decline began with some general considerations of F. P. Ramsey (of which more later). But its demise had another cause. The world was rendered uninhabitable for it by the impact of a single argument, though one appearing in various guises. It is that in order to assign a probability to a proposition after evidence comes in, it is necessary to first assign it a "prior" probability: the degree to which one ought to believe it initially. (One can then use Bayes' theorem to update its probability.) But the only natural way to assign prior probabilities is by symmetry arguments, that is, applications of the "principle of insufficient reason". For example, different faces of a die are equally likely to appear, prior to experiment, because there is no reason to prefer one to any other. But different symmetry arguments give different prior probabilities for the same problem, and there is no principled way of deciding which is right. Prior probabilities, it is concluded (and hence all "logical" probabilities) must either not exist, or be a matter of free choice by the investigator. (Standard recent presentations of the argument are in

Howson and Urbach 1993, pp. 59–72; van Fraasen 1989, ch. 12; Salmon 1966, pp. 66–8; Earman 1992, pp. 16–20, 139–41; Howson 1995; Howson 1997.) To take the simplest kind of example, if there are three kinds of balls in an urn, white, red and blue, what is the initial probability that a ball to be drawn will be white? Is it $\frac{1}{3}$ (because it could as easily be either white, red or blue) or $\frac{1}{2}$ (because it could as easily be white or non-white)? There are many similar cases. In a continuous case, where the value of some parameter is "completely unknown", should one express this by a uniform distribution over the parameter T, or a uniform distribution over T^2? For example, if one knows only that a factory process makes metal squares, should one express one's ignorance about their size by a uniform distribution over the possible lengths of the side of the squares, or over the possible areas? And Carnap famously agonised over whether, in looking at a simple universe in which objects could either have or lack an attribute A, one should regard as initially equally probable all possible "states", that is, assignments of A and non-A to all individuals, or all possible proportions of A to non-A, or something in between (Carnap 1952, pp. 1–2, 53–5; Howson and Urbach 1993, pp. 62–6). The most embarrassing case arose from Laplace's Rule of Succession. The Rule states that the probability that the next A will be a B, given that all of the previously observed n As have been Bs, is $\frac{n+1}{n+2}$. The Rule applies even when $n = 0$ or 1, in which case it asserts that the probability of the sun rising tomorrow is $\frac{1}{2}$ in the absence of evidence, increasing to $\frac{2}{3}$ after one successful observation. Even Keynes ridiculed Laplace over the oddity of the prescriptions of the Rule of Succession for small n (Keynes 1921, pp. 377, 383). To make the matter worse, the difficulty does not follow from the precise form of Laplace's Rule. It seems that almost whatever a rule of succession prescribed, it would look equally ridiculous for small n.

Now this argument is one that the logical probabilist will have to meet head-on. But there should be some initial doubts about the style of the argument, arising from the fact that in other branches of philosophy, difficulties in applying symmetry arguments do not immediately result in their abandonment. Consider metaethics, for example, where there is a familiar argument from the diversity of moral opinions in various cultures to the conclusion that there is no moral objectivity. Even those who, in the end, accept the argument recognise that there is a great deal of work to do in progressing from the simple fact of disagreement in values to the conclusion that "anything goes" (Snare 1980; Snare 1984). Almost everyone else believes that a certain amount of second-order criticism of moral systems is possible, which will reveal that some of them are less acceptable than others, even if there is no unique right one. Similarly with global scepticism.

The most substantial argument for global scepticism is the "symmetry argument", that demon worlds or vat worlds are indistinguishable from the real physical world, with respect to perceptual evidence; "there is no mark to distinguish the true and the false", as Cicero says (Franklin 1991). Nevertheless, it can hardly be pretended that a simple assertion or denial of symmetry will end the discussion. On the contrary, criticism and defence of the argument on logical grounds is exactly what the discussion consists in.

So there is reason to doubt the assumption, maintained equally by the opponents of logical probability and most of its defenders, that it is an essential claim of logical probability that there is always a *unique* probability assignable *a priori* and obviously by symmetry. The existence of logical probability would be established by something much weaker: if it appeared that there *were* different prior probabilities assignable by different symmetry arguments, but that one could through logical considerations partially order those priors as better or worse.

There are in fact three possible lines the defender of logical probability can take, short of defending a unique precise prior in every case. They are

- To maintain that any numbers given to the priors are, within limits, not to be taken seriously – because they are imprecise, fuzzy or vague, or unrobust in the face of evidence, or otherwise shaky for purely logical reasons.

- To maintain that although there is no uniquely specifiable prior, there is a principled distinction in the space of possible priors between reasonable and unreasonable ones, and that which is which is decidable, in large part at least, on logical grounds.

- To maintain that in real cases, there is always a good deal of background information to be taken into account, so that reasonable debate about what the correct prior is may often be explained by differences in the background information being called into play.

These three lines of defence are not entirely distinct, since vagueness in a prior is much the same thing as a distribution in a space of possible priors, and uncertainty as to the background information really present can also be expressible via a range of priors. The defences are also compatible and cumulative. All three will be pursued here. But it will also be maintained that, when all due allowance for these excuses has been made, there are many real cases where a uniform (or maximum entropy) prior is the unique most reasonable one.

2. THE WEIGHTLESSNESS OF PRIORS

Let us take first the defence involving the fuzziness of priors. It is obvious that, elsewhere, the logical probabilist should resist demands to supply numbers to all enquirers on all problems. If one is asked for

> P(the moon is made of green cheese | Marilyn Monroe was murdered by the CIA)

then the correct response should be that there is no such number. Logical probability was intended to express a "degree of partial entailment" (cf. Keynes 1921, p. 15) so that, if propositions h and e are irrelevant to each other (in the sense of relevant logic, of having no logical connections between their constituents), there is no partial implication, and hence no number expressing it.

At the other extreme, there are cases where a perfectly precise number may be attached to P(h | e), and that number ought to be taken seriously. A paradigm is the probability assignment sometimes called the "proportional syllogism" (Forrest 1986, ch. 8) or "statistical syllogism" (Hempel 1965, ch. 2, also Kyburg 1974, p. 247):

> P(Tex is rich | Tex is a Texan and 90% of Texans are rich) = 0.9

It is to be noted that this assignment, natural as it is, is available to logical probabilists only – subjective Bayesians can have no reason to choose it. For it involves a uniform distribution across the space of Texans. (To obviate misunderstandings, note that this assignment is not undermined by the fact that the number would be different if one knew, for example, that Tex was Mexican; the fact that

> P(Tex is rich | Tex is a Texan and 90% of Texans are rich and Tex is Mexican) $\neq$ 0.9

says nothing about the relation of the original evidence to the conclusion.)

The conceptual machinery to explain the difference between the two situations already exists, in Keynes concept of "weight" (Keynes 1921, ch. 6; see Cohen 1986; O'Donnell 1992; Shafer 1976, pp. 5–6; Achinstein 1994; a recent attempt to quantify weight in Jaynes, to appear, ch. 18) The standard example to introduce the concept of weight involves asking what the difference is between:

> P(this coin will come up heads | this coin appears symmetrical and about 500 of the 1000 throws with it have come up heads) = $\frac{1}{2}$

and

> P(this coin will come up heads | this coin appears symmetrical)
> $= \frac{1}{2}$.

Though the numerical value of the probability is the same in each case, there is plainly some important difference between the two. In the first case, the probability $\frac{1}{2}$ derives from the balance of a much greater amount of evidence than in the second. This amount of evidence that enters into the determination of the probability (but is not expressed in the numerical value of the probability) is what Keynes called "weight".

The relevance of weight to inference appears in the merely presumptive nature of inferences of low weight. They are easily overturned by substantial evidence (Runde 1991). In the extreme case where the inference from e to h has no weight, this means that if substantial evidence e′ for h appears, the original evidence e has no effect on how likely h now is; that is, although P(h | e) ≥ P(h | tautology),

> P(h | e′ & e) = P(h | e′)

Such judgements of "conditional independence" play a major role in "Bayesian networks" in Artificial Intelligence, whose considerable success in mechanising probabilistic reasoning seems to stem from the naturalness of representing conditional independence between propositions by not joining the nodes that represent those propositions (Charniak 1991; Pearl 1993; Spiegelhalter et al. 1993; Anderson and Hooker 1994).

Of interest in the present debate are intermediate cases, where the evidence e has some bearing on the hypothesis h, but not much. In that case, new relevant evidence e′ will dominate the effect of the old evidence of little weight, e, so that P(h | e′ & e) will be close to P(h | e′).

Probabilistic inference that involves evidence of low weight has always been recognised as particularly suspect. Theories with low weight are certainly known in pure science, such as the farcical scenarios that pass for "theories of the origin of life". (Horgan 1991) But at least in pure science one can afford the luxury of scepticism about all the alternatives so far thought of, if necessary. In applied science, there are realistic physical cases where the prior dominates, arising from the fact that measurement is often a partly random process, and repeating a measurement may be difficult. For example, if one must estimate the half-life of an isotope, given the observed decay time of a single atom, then one cannot pretend to know nothing, and the answer depends on a reasonable choice of prior (which cannot be uniform, because there is no upper limit on a possible half-life) (Steenstrup 1984). Inferences of low weight are also familiar in

television coverage of vote counting in elections, where the commentators warn against making too much of early returns. The most unsettling cases occur, however, in law, where one is sometimes forced to reach a decision on a fixed amount of evidence, even if that evidence is almost non-existent. When it is "your word against mine", there is little evidence, and what there is conflicts. The type of case that has been most discussed involves determining negligence in head-on collisions which kill both drivers. In *T.N.T. Management* v. *Brooks*, the plaintiff's husband was one of two drivers killed in a head-on collision on a straight road. There were no witnesses, and almost no further relevant evidence, and hence a symmetry in the evidence with respect to each driver. The legal situation required a decision as to whether, "on the balance of probabilities", the other driver was negligent (irrespective of any possible negligence on the part of the plaintiff's husband). It was argued, using the following diagram, that on the balance of probabilities the other driver *was* negligent (Eggleston 1983, p. 184; further on weight and legal evidence in Davidson and Pargetter 1987)
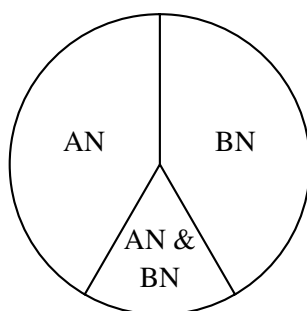
AN = plaintiff's husband alone negligent
BN = defendant's driver alone negligent
AN & BN = both drivers negligent

(The probability that neither driver was negligent has been neglected, on the grounds that such a collision could not happen unless at least one driver were negligent; but the result still follows even if there is some probability of neither driver being negligent, provided it is smaller than the probability of both being negligent.) There is no way to avoid the problem that the evidence is symmetrical, but of little weight.

Even more alarming cases are possible. One might be less inclined to make jokes at the expense of Laplace's Rule of Succession for small n if

one were on the receiving end of a case like the following one from the *Talmud*:

It was taught: If she circumcised her first child, and he died, and a second one also died, she must not circumcise her third child; thus Rabbi. R. Shimon ben Gamaliel, however, said: She circumcises the third, but must not circumcise the fourth child. . . . It once happened with four sisters at Sepphoris that when the first had circumcised her child he died; when the second [circumcised her child] he also died, and when the third, he also died. The fourth came before R. Shimon ben Gamaliel who told her, "You must not circumcise". But is it not possible that if the third sister had come he would also have told her the same? . . . It is possible that he meant to teach us the following: That sisters also establish a presumption.

Raba said: Now that it has been stated that sisters also establish a presumption, a man should not take a wife either from a family of epileptics, or from a family of lepers. This applies, however, only when the fact had been established by three cases. (*Talmud*, Yebamot 64b; cf. *Shabbath* 61a,b; modern versions in Walley et al. 1996; Greenland 1998)

Who is not relieved at the absence of frequentist and subjective Bayesian rabbis? There follows a discussion of whether it is safe to marry a woman who has had two husbands die. It is felt that if the deaths were obviously due to some chance event, such as falling out of a palm tree, there is no need to worry, but if not, there may be some hidden cause in the woman, which the prospective husband would do well to take into account.

The examples serve as a reminder that, while frequentists and Bayesians have derided logical probabilists over their problems with small samples, it is precisely there that logical probabilists have the strongest case (Verbraak 1990, Introduction; also in Richard Price's appendix to Bayes' original paper: Price 1763, at p. 312) For frequentism has little to offer, because there has been no long run in which to study relative frequencies, and Bayesianism is weak because the divergent priors it gives permission to choose freely will dominate what little experience there is, leading to completely arbitrary recommendations. It is when one cannot avoid dealing with examples such as those just discussed, where there is almost no experience, that one has only considerations of symmetry, possibly imprecise, to fall back on.

If one represents probabilities of low weight by imprecise numbers, calculations may take more trouble. But there are no major technical difficulties. The simplest method is to use probabilities with error bounds, (Walley 1991, Kyburg 1974, chs 9–10) but one can with also calculate with fuzzy probabilities if desired (Pan and Yuan 1997; cf. Levi 1974, section IV).

3. ARGUING ABOUT PRIORS

We now turn to the second defence, according to which the possibility of arguing rationally about the appropriateness of priors supports logical probability.

It is true that, as subjective Bayesians insist, one can construct cases where different symmetry requirements conflict (Milne 1983). But that simply confirms the desirability of relying on a single symmetry, when there is only one available, and calls attention to the need to balance values in logical space otherwise. As in ethics, the necessity in some cases to make tradeoffs between conflicting values gives more, not less, reason to believe those values exist.

If there really is doubt about which symmetry is appropriate, as there can be with the distribution of parameters, that doubt is itself part of the logical structure of the problem. One side cannot be allowed to trade on it implicitly, in favour of a sceptical conclusion, while denying the other side the chance to work explicitly in a second-order space of possibilities. In real statistical applications, one is doing this all the time, for example when taking one class of models as "reasonable" in the situation, and ignoring all others (that is, assigning them zero prior probability). Again, attempts to find priors by using higher symmetries are useful, even if they are not sufficient to produce a convincingly unique prior (Hartigan 1964). And, if a uniform distribution on T and one on T^2 are both reasonable in the context, it does not follow that a uniform distribution on T^{100} is also reasonable. Further, it is not alarming that mathematicians have a great deal of trouble in specifying *any* reasonable priors for continuous parameters, or that the priors they arrive at are uniform only in simple cases (Berger and Bernardo 1992, with many references). The possible values of a continuous parameter may occupy an infinite range, so that the prior probabilities may be thought to "tail off" at infinity. Or, even if the range of possible values is bounded, say to lie between 0 and 1, it is far from obvious whether the edges of the range are in the same logical situation as the middle; hence it may not be clear whether a uniform initial distribution is reasonable. It is not surprising to find, as one group of mathematicians puts it, that "There are often several candidates [as a prior], each fulfilling the different natural requirements, though not to the same extent" (Ghosh and Mukerjee 1992, p. 196; similar in Walley 1991, section 5.5; Kass and Wasserman 1996; Uffink 1996). The simplicity of competing hypotheses is another logical property of them that may need to be taken into account in some circumstances (Forster 1995). None of these problems, however,

gives any reason to doubt the assignment of a uniform prior to the six possible outcomes of throwing a die.

The relevance of subtle considerations in complicated logical spaces is confirmed by the existence of similar probabilistic reasoning in a space which everyone admits to be a matter of pure logic, namely mathematics. Pure mathematicians regularly consider numerical evidence for and against conjectures, and there are purely mathematical versions of such probabilistic argument forms as (ordinary, non-mathematical) induction (Franklin 1987). No-one takes this as reason to think that mathematical reasoning is not objective. Similar complications should not give reason to think the assignment of reasonable priors is not objective, either.

The logical probabilist is not claiming that the search for appropriate priors is easy. If topics were easy just in virtue of their subject matters being simple and abstract, we wouldn't need philosophers.

4. PRIOR TO WHAT? THE UBIQUITY OF BACKGROUND INFORMATION

We now consider the third defence of logical probability, arising from the almost ubiquitous presence of background information, whether explicit or implicit.

Underlying the objections to logical probability is the belief that it represents the "siren melody of empirical probabilities determined a priori by pure symmetry considerations", and is thus an attempt to make empirical predictions by "pure thought" (Van Fraasen 1989, pp. 317, 315). To determine the justice of such charges, one must explain what is meant by the "prior" in prior probabilities. Does it mean prior to all experience (that is, conditioned on a tautology)? Only in that case would it be just without qualification to accuse logical probability of metaphysical a priorism.

In fact, the situation is mixed. Typically, there is plenty of background evidence, explicit or implicit. It is just that the hypothesis is symmetrical with respect to it. Of course one knows many relevant things about dice and their falling, and about the initial conditions of each throw. But the "mixing" effect of the many rotations of the die between its cast and its fall blurs those facts out into a uniform background, making them indifferent with respect to the six possible outcomes. In general, symmetry arguments may be strong or weak, resulting in priors of high or low weight. The phrase "principle of insufficient reason" suggests weak ones, where we are in a state of ignorance, and "there is no reason to prefer one alternative to the other". But symmetry arguments may be strong: one should not confuse "there is no reason to prefer one alternative to the other" with "there is reason not to prefer one alternative to the other" (Verbraak 1990,

p. 43; similar in Castell 1998). A die which looks symmetrical to a casual glance results in a weak symmetry argument for the equal probability of the outcomes, whereas a carefully weighed die, with well-understood considerations about the physics of its throwing, provides a strong positive reason to assign the outcomes equal prior probability (Strevens 1998). But there is always the possibility of being convinced otherwise. Symmetry arguments, even strong ones, are defeasible.

More generally, background or contextual information need have nothing to do with symmetry, but may relate to, for example, independence of distributions, or any vague or precise constraints on probabilities (Festa, 1993, ch. 7.5). In any such case, what is of concern is how the background information bears on the hypothesis, so that we are still in the domain of logic. The ubiquity of background information is nicely illustrated by an example which is common in the folklore of the subject, but which has perhaps not been discussed in print. (But some similar examples in Swinburne 1971, pp. 326–7; Howson and Urbach 1993, p. 129; an ancient example in Diodorus Siculus bk. 3 chs 36–7.) Normally, instances of a generalisation are taken to confirm it; even among inductive sceptics, they are not usually taken to disconfirm it. Consider, however, the generalisation, "All humans are less than 5 metres tall". This is confirmed by present observations of people, all of whom have been observed to be less than 5 metres tall. Suppose that an expedition returns from a previously unexplored jungle and reports, with suitable evidence, that they have observed a human 4.99 metres tall. On this evidence, the probability that all people are less than 5 metres tall is nearly zero, although the generalisation still has only positive instances.

The lesson of the example is surely not that there is something wrong with confirmation by instances, but that information relevant to the probability is hidden in the structure of the concepts being used. Length, as everyone who uses the concept knows, is something that admits of continuous variation, which means that the existence of something 4.99 metres tall makes it probable that there is a similar thing at least 5 metres tall. There is no justification for pretending not to know such a fact, on the ground that logic ought to be formal (that is, that instance confirmation ought to apply in the same way for all concepts). The dogma that logic is formal or syntactic has enough difficulties even in deductive logic (for example, one cannot always substitute in inference schemas such as modus ponens concepts that have deductive-logical oddities, such as being inconsistent or self-referential (Stove 1986, ch. 10).) Nor can one apply logical probability blindly, or formally, just because it is logic, and treat concepts that have probabilistic-logical complexity as if they were simple. "Green",

"grue" and "less than 5 metres tall" have different logical structures, which can result in their behaving differently with respect to logical probability. Similarly, if the logical probabilist is asked to speculate on

P(this bird is red | a tautology)

he will of course refuse to do so, but may usefully say that there is certainly information in the meanings of "bird" and "red": both of these are words meant to label categories into which things in the world will fall, so one expects antecedently that only a minority of applicable items will fall into any one category. That is why a logical probabilist or Bayesian analysis of the Ravens Paradox will rightly help itself to such background information as that black ravens are expected to be rarer than non-black non-ravens (Earman 1992, pp. 70–3; Rosenkrantz 1994). This is not to maintain that the rarity of black ravens is itself a logical truth. It is merely to recall that logical probability deals in the relation between background knowledge and hypothesis, a relation which is sensitive to what the background knowledge actually is.

None of this requires special concessions to be made on behalf of non-deductive logic which would not be allowed in deductive logic. When discussing the truth of "all bachelors are unmarried", deductive logicians are usually prepared to abandon a strict formalism, and look inside the meanings of the terms to discover some relation of containment of meaning, or some surrogate. If such a relation exists, so may more complicated relations between meanings. One does not need to be a full-blown Saussurian structuralist about language to admit that the system in which meanings reside creates subtle relationships between them.

For much the same reason, the logical probabilist, while admiring Carnap's work on matters of principle, will believe his priors are a blind alley that he took because of his nominalism, and resulting fixation on syntax. All he could think of having a uniform prior over was either the space of states, or the space of proportions. Both of these assign zero prior probabilities to laws, which is wrong *a priori*. Plainly, someone believing in kinds or universals has no need to follow him. (alternatives in Changizi and Barber 1998)

5. RAMSEY'S CRITICISM OF KEYNES

To complete the defence of logical probability, it is necessary to consider briefly F. P. Ramsey's criticisms of Keynes, which have been influential in convincing many that no logical theory of probability is possible. They are

also clear versions of the initial qualms which it is natural to feel about the whole notion of logical interpretations of probability. His arguments are quite different to those concerning the principle of indifference. Their general tenor is that Keynes has failed to make intelligible the notion of a partial logical implication, and that a subjectivist interpretation of probability is supported by a definition of probabilities in terms of the betting odds an agent is willing to accept. The laws of probability are then justified in terms of a Dutch Book argument.

Ramsey's fundamental criticism is that "there really do not seem to be any such things as the probability relations he [Keynes] describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. I do not perceive them, and if I am to be persuaded they exist it must be by argument; moreover I shrewdly suspect that others do not perceive them either, because they are able to come to so very little agreement as to which of them relates any two given propositions" (Ramsey 1926, p. 57). As to the last assertion, about people other than Ramsey, it would seem that he exaggerates the disagreement existing about the relation between "The vast majority of trains arrive at their destination" and "This train will arrive at its destination", and that it will not be easy to find many people who think the former proposition tells against the latter. Juries are able to agree on verdicts, at least in clear-cut cases, which would happen extremely rarely if there were indeed the general disagreement Ramsey asserts. Again, his own theory requires that people should consistently act *as if* they believed that inductive evidence could be strong evidence for a particular conclusion. For example, he is confident that "we all" can be relied on to accept odds better than $\frac{35}{36}$ on the sun's rising tomorrow (p. 80). As for his assertion about his own inability to discern relations of partial entailment, one is at first loath to deny his privileged epistemic access to his own mental states. But one must rise to the challenge, when one finds him writing two pages later: "If anyone were to ask me what probability one [proposition] gave to the other, I should not try to answer by contemplating the propositions and trying to discern a logical relation between them, I should, rather, try to imagine that one of them was all that I knew, and to guess what degree of confidence I should then have in the other" (p. 59). Ramsey seems to have asserted here a distinction without a difference. Suppose that on performing the recommended imaginative act, I find that I have absolute confidence in the consequent. Have I not then perceived a strict logical implication between antecedent and consequent? According to "mental models" theories of implication, I have, since all there is to implication is the necessary holding of the consequent in situations where the antecedent holds (Johnson-Laird

1983). Theories of conditionals and implication are of course varied and controversial, but by and large they give little support to the distinction Ramsey is relying on. His remark seems to reinforce the similarity between deductive and partial implication more than it undermines it.

Ramsey goes on to make a further criticism of Keynes. "If", he says, "we take the simplest possible pairs of propositions such as 'This is red' and 'That is blue' or 'This is red' and 'That is red', whose logical relations should surely be easiest to see, no one, I think, pretends to be sure what is the probability relation which connects them. Or, perhaps, they may claim to see the relation but they will not be able to say anything about it with certainty, to state if it is more or less than $\frac{1}{3}$, or so on" (Ramsey 1926, p. 58). The answer to this objection lies in the approach developed above, according to which the relations between such propositions are of very little weight, and so cannot be expected to have a precise number attached to them meaningfully. But one could again ask whether the situation is really different in deductive logic, with which Ramsey obviously intends a contrast. What of the pair of propositions he has carefully avoided, 'This is red' and 'This is blue'? If these propositions were so easy to deal with, it ought to be clear whether their incompatibility is a matter of strict logic, or not. Which it isn't.

A last issue, which is not developed by Ramsey but arises from his work, is whether subjective Bayesians have an advantage from the Dutch Book justifications of the probability axioms. Are logical probabilists entitled to use these, and if not, can they give any better justification of the axioms?

In order to understand whether any school of probability has an advantage with respect to justification, it is desirable to recall the variety of approaches to axiom justification that have been offered. There are four distinct approaches:

1. It is argued that the axioms of "probability" are necessarily true of relative frequencies. So they naturally carry over to cases of partial belief where the evidence consists wholly of relative frequencies, and may as well be extended to other cases, if there is no reason to do anything else (Ramsey 1926, pp. 54–5).

   This style of argument is rather weak, especially for cases of probability where the evidence is not numerical – in such cases, it gives little reason even to think that the probabilities ought to be numerical. On the other hand, it is simple and intuitive. Also, there is an onus on anyone not agreeing with it to explain the remarkable coincidence between the laws of consistency of partial belief and those of relative frequency.

Plainly, this approach supports objective rather than subjective Bayesianism, since it derives the probability laws from the basic statements of probability: the law-instance

> P(this ball is white | 75% of balls are white) =
> 1 - P(this ball is not white | 75% of balls are white)

is derived *from* the pre-existing values of these two probabilities.

It is a little surprising that Ramsey actually approves this line of argument, up to a point. In material towards the end of his article, taken from C. S. Peirce, he writes, "granting that he is going to think always in the same way about all yellow toadstools, we can ask what degree of confidence it would be best for him to have that they are unwholesome. And the answer is that it will in general be best for his degree of belief that a yellow toadstool is unwholesome be equal to the proportion of yellow toadstools that are in fact unwholesome. (This follows from the meaning of degree of belief)" (p. 91). And again, "given a [predictive] habit of a certain form, we can praise or blame it accordingly as the degree of belief it produces is near or far from the actual proportion in which the habit leads to truth" (p. 92). This material was left somewhat undigested at the time of Ramsey's death, and he would no doubt have worked harder to make it consistent with his earlier doctrine of the free choice of basic degrees of belief. But as a obstacle to digestion, it would surely have proved a big pig even for a boa constrictor as powerful as Ramsey.

2. Dutch book arguments show that if the principles of probability are violated, it is possible to make a book against the violator, which ensures a sure loss if he bets in accordance with his probabilities. Ramsey calls the requirement of no sure loss "consistency", but it is not, of course, logical consistency (Discussion in Howson 1997). Whatever the outcome of debate about the strength of Dutch book arguments (Maher 1993, ch. 5; Davidson and Pargetter 1985), the immediate issue is whether they favour any interpretation of probability over others. They do not, strictly speaking, support the subjective Bayesian position that one can think what one likes about basic probabilities (subject only to consistency), since they say nothing about basic probabilities. Nevertheless, it is reasonable to take the possibility of justifying the "consistency" laws on their own as some support for the subjectivist view. Or it would be, if the other styles of justification did not exist.

3. Some approaches start with intuitively reasonable weak axioms for a comparative notion of probability, and work towards – but do not necessarily arrive at – a numerical notion (Koopman 1940). A view of this kind seems to have been held by Keynes, especially after he had

absorbed Ramsey's critique. He founds probability theory much more on simple judgements of preference, indifference, relevance and irrelevance than on numerical judgements (Runde 1994). Such approaches tend to favour a logical concept of probability, but are also consistent with subjectivism.

4. Cox's justification is interesting because it uses powerful mathematics and exceedingly weak assumptions. The assumptions are:

> The probability of an inference on given evidence determines the probability of its contradictory on the same evidence.
>
> The probability on given evidence that both of two inferences are true is determined by their separate probabilities, one on the given evidence, the other on this evidence with the additional assumption that the first inference is true.

(By "inference" he means, of course, "proposition inferred".) The second assumption means no more, in effect, than that the order in which evidence for a proposition arrives does not affect how it bears on the proposition in total. These axioms assume nothing about *how* the probability of a negation or conjunction depend on that of its components. Using only the standard Boolean algebra of propositions and some difficult but purely mathematical results on functional equations, Cox was able to derive the sum and product rules (Cox 1946; Cox 1961, pp. 3–5, 12–25; Paris 1994, pp. 24–33). (Or more exactly, since the assumptions do not determine a scale on which probability can be measured, he proved that the laws must be a rescaling of the usual sum and product rules on the normal zero-to-one interval.)

This derivation is in an abstract sense superior to the others, in that it uses the weakest assumptions. Its bearing on the subjective-versus-objective dispute is unclear. Cox was as it happened an objectivist (who denied that probabilities could always be assigned precise numbers (Cox 1961, p. 32)), but that does not follow from his derivation, which would make sense in the subjectivist framework. However, the existence of a derivation of the laws of probability which is superior to the Dutch Book approach, and is also free of that approach's apparatus of decisions and bets, does tend to undermine any support that Dutch Book arguments may be thought to give to subjectivism.

A further reason to doubt that the Dutch Book argument favours the subjective interpretation is that there exists a Dutch Book-style justification of the Principle of Indifference. Suppose your opponent tells you that a bean is in one of three boxes, and allows you to choose the odds of a bet against him. The opponent may choose which box has the bean, after you have made the choice of odds. Your aim is to minimize your total loss by choosing the odds. You should choose $\frac{1}{3}$ for each box, since

any deviation from equality of odds can be taken advantage of by your opponent, to cause a greater loss to you. This is essentially the reasoning of the Dutch Book argument, except that "avoiding sure loss" is replaced by "minimising sure loss" – which changes nothing from the point of view of justification (O'Neill 1996, p. 95).

Again, it is necessary to ask whether the situation regarding justification of the axioms in deductive logic is really better than it is in non-deductive logic. Ramsey was aware of the problem, writing, "I think it far better to seek an understanding of this 'necessity' [of logical connection] after the model of the work of Mr Wittgenstein, which enables us to see clearly in what precise sense and why logical propositions are necessary, and in a general way why the system of formal logic consists of the propositions it does consist of, and what is their common characteristic" (Ramsey 1926, p. 61). These optimistic hopes have not been realised, thanks among other things to the later works of Mr Wittgenstein. The justification of *modus ponens*, for example – that is, some kind of proof that it is truth-preserving – is a real but unsolved problem, as any suggested proof is in danger of being either too weak or circular (Haack 1976). The situation with, for example, justifying the axioms of modal logic is even worse. It seems that logical probability suffers no special disadvantages regarding the justification of its axioms.

One last line of objections to logical probability has come from the Popper-Miller arguments to the conclusion that there can be no such thing as an inductive logic. These have been adequately dealt with by many authors (Mura 1990).

That completes the defence of the notion of logical probability against the arguments that have been used against it. Now it is time for counterattack. It is argued that the adherents of rival conceptions of probability have not been able to avoid an implicit commitment to the essentials of logical probability.

6. SUBJECTIVE BAYESIANS ARE SECRET LOGICAL PROBABILISTS

The standard ("subjectivist") Bayesianism is "a theory of individuals' personal probability assignments subject to the sole constraint of consistency" (Howson 1992). That is, "anything goes" as regards the choice of prior probabilities. One can suspect that this apparently flexible and anarchic prescription is actually a recommendation of a uniform distribution in some higher space of possibilities. But the most serious question the Bayesian must answer is: May one assign a prior probability of zero to some initial possibilities (that is, to scenarios not ruled out on logical

grounds)? On the one hand, why not, if anything goes? On the other hand, the whole tenor of Bayesian rhetoric was along the lines of "Let a hundred flowers bloom". In contrast to the rigid prescriptions of logical probability, it was suggested, all possibilities should get their chance to have belief given to them, and later to have experience tell either for or against them (Howson and Urbach 1993, p. 419). But if a possibility is assigned zero prior probability, no amount of experience can give it any positive posterior probability: it has been ruled out of consideration from the start. No flower will bloom if it is bulldozed as soon as it pops above ground.

If now the Bayesian agrees to assign non-zero prior probability to all ("reasonable"?) possibilities, then there is still trouble. If the prior probability assigned to some hypothesis is non-zero, but still so small that experience as long as the expected lifetime of the universe will not make it appreciable, then it might as well have been zero: practically, one will never come to believe it, or even give it a probability as high as a half, no matter what happens.

So it appears that for Bayesianism to work in a real case, it must assign some "substantial" or "appreciable" prior probability to all "reasonable" hypotheses. This position differs only in name from a logical probability that is critical of its priors, as defended above. It is the logical probabilist, not the Bayesian, who leaves his belief state as open as possible to modification by data:

The [uniform, or more generally maximum entropy] distribution is the one which is, in a certain sense, spread out as uniformly as possible without contradicting the given information, i.e., it gives free rein to all possible variability of x allowed by the constraints. Thus it accomplishes, in at least one sense, the intuitive purpose of assigning a prior distribution; it agrees with what is known, but expresses a "maximum uncertainty" with respect to all other matters, and thus leaves a maximum possible freedom for our final decisions to be influenced by the subsequent sample data. (Jaynes 1968, p. 231)

The Bayesian faces the dilemma of either substantially agreeing with this, or allowing large diversions from it (that is, dogmatic priors) without reason.

The difficulty Bayesians have in escaping entirely from uniform priors is illustrated by Howson and Urbach's criticism of Carnap's assignment of prior probabilities:

In the first place, such a priori assignments necessarily exhibit a more or less strong bias against certain types of world, which we have no right, in advance of all empirical information, to indulge in. (Howson and Urbach 1993, p. 69)

That is a just criticism of Carnap, but is it not a case of the pot calling the kettle black? Who is the Bayesian to complain about *a priori* biases, and demand a more democratic distribution of prior probabilities? Another

example of the difficulty of avoiding uniformity and symmetry arguments comes from the work of the subjectivist de Finetti on "exchangeability". De Finetti observed that to make subjectivism work in cases like coin tosses, it was necessary to assume at least some weak symmetry principle that permitted certain probabilities to be the same as others (De Finetti 1964, p. 123; Kyburg pp. 111–2). He fails to explain why a symmetry argument should be acceptable in such cases but not more widely.

This is perhaps the natural place to discuss the vexed question of the "robustness" of posteriors to priors. As has been widely observed, the difference between priors tends to "wash out" with experience: experience will dominate differences in the prior probabilities, so that even if there were large initial differences in what two parties believed, after a moderate amount of observation they should come to believe much the same thing (Earman 1992, pp. 141–9; Barrett 1996). This fact has been used as a defence of Bayesianism against the charge that its priors are arbitrary. If the choice of prior is not important, then its arbitrariness is also unimportant (DeRobertis and Hartigan 1981; review in Wasserman 1992). To some extent, the motivation is close to that for the argument above for the weightlessness of priors. There is however an obvious difficulty with it, which is the reason why it has not had much success with frequentists, and has not been very popular among Bayesians themselves. The difficulty is that it works only for "reasonable" priors. If one takes a "crazy" prior, which assigns some real possibility a zero or near-zero prior probability, then no amount of experience will be able to dig one out of that commitment. The difference between a reasonable and a crazy prior will not wash out.

To resuscitate the defence, the Bayesian will need to drop the "anything goes" dogma about priors, and admit the possibility of and need for logical criticism to distinguish between reasonable and crazy priors (A similar argument in Maher 1996). But that is to become a logical probabilist. Then, the washing out of the difference between priors can be interpreted as it should be, as due to the low weight of the priors being easily dominated by the incoming evidence.

There is another problem the Bayesian must face, if the "anything goes" dogma is to be maintained. Intuitively, "the vast majority of As are Bs" is, just by itself and in the absence of other relevant evidence, a reason to believe "This A is a B." It is a principle taken for granted in, for example, statistical mechanics, where the fact that almost all possible states have high entropy is a reason for believing that the state at the present moment has high entropy (so that, for example, one does not expect the molecules in a jar of gas to bunch up in one corner, though this is possible) (Jaynes

1957, p. 627). Or again, the mathematical fact that the vast majority of large samples from a population resemble the population in composition is naturally thought to have some relevance to any sample-to-population inference, whether opinion-polling or induction (Stove 1986, ch. 6; Williams 1947). Yet to allow sheer weight of numbers to count, as giving the vast majority a strong vote does, is precisely to have a prior that is uniform across the possible individuals, or at least not grossly far from uniform. That is precisely what the genuinely subjectivist Bayesian cannot allow as inherently reasonable.

Finally, there is a hidden subtlety in Bayes' theorem itself where it is hard to avoid some "dogmatism" about probabilities. Bayes' theorem states that $P(h \mid e)$ is proportional to $P(e \mid h) \times P(h)$, that is, the probability of a hypothesis on some evidence depends on the prior probability of the hypothesis, $P(h)$, and also on how likely the evidence is, given the hypothesis. Bayesians speak as if subjectivity resides only in the choice of the prior $P(h)$. But how is one to know $P(e \mid h)$? Take, for concreteness, a typical inductive case where e = "Tex is rich", and h = "90% of all Texans are rich". In that case $P(e \mid h)$ is just

$$P(\text{Tex is rich} \mid \text{Tex is a Texan and 90\% of Texans are rich}).$$

What number does the Bayesian propose to assign to this? To agree to the natural choice of 0.9, or anything close to that, would be to concede to logical probabilists all they asked for, a uniform prior distribution across Texans of the probability of being rich. To do otherwise, and allow arbitrary divergence from 0.9, will be consistent, but threatens a regress of arbitrary choices before Bayesian inference can even get under way. The difficulty of avoiding the natural proportional syllogism is illustrated by its appearance in Earman's Bayesian analysis of the Ravens Paradox, already mentioned. There, it was proposed that the frequency of black ravens was much less than that of non-black non-ravens. True, but of what relevance is that, without a proportional syllogism? Again, Howson and Urbach license the Principal Principle, allowing inference from a proportion *p* in a suitable *infinite* series of throws to a degree of belief *p* in an individual throw; they justify this with a theorem to the effect that *with probability one* (that is, in the vast majority of cases), finite sequences have relative frequencies tending to *p* (Howson and Urbach 1993, p. 345). True, but again, what is the relevance of that, unless the proportional syllogism is already accepted?

7. FREQUENTISTS ARE SECRET LOGICAL PROBABILISTS

The "classical" theory of hypothesis testing in statistics, stemming from Neyman, is built on his frequentist theory of probability. Neyman was an explicit deductivist, asserting that "the term 'inductive reasoning' itself seems to involve a contradictory adjective. The word 'reasoning' generally seems to denote the mental process leading to knowledge. As such, it can only be deductive" (Neyman 1937, p. 379). Neyman developed a method of statistical estimation, concerning how observations bear on the estimation of parameters. For example, if measurements of the distance of a planet are corrupted by noise, one can take the mean of many observations in an attempt to get a better estimate of the true distance. If the characteristics of the noise are known, one can calculate, using Neyman's methods, a "95% confidence interval" for the parameter. The question is what this means, given that it cannot mean that one is 95% confident that the parameter lies in that interval (since that would be to admit logical or subjective probability). The standard answer, repeated almost word for word in a long succession of statistics textbooks, is:

The interval, not [the parameter] μ, is the random variable. If thousands of such intervals were calculated in the same way, each based upon a different random sample, then in the long run about 95 out of every 100 intervals would include or capture μ. The procedure is somewhat analogous to throwing intervals at a point target in such a way that 95 per cent of the time the interval will cover the point. (Guenther 1965, pp. 112–3; cf. Armore 1966, pp. 314–5; Kendall and Stuart 1979, vol. 2 p. 110; Steel and Torrie 1976, pp. 164–5; discussion in Howson and Urbach 1993, pp. 237–41)

That is the entire justification offered for the method used to estimate μ. But why is it a justification? We did not, and will not, make thousands of measurements. One interval was calculated, from one sample, and we wanted to know how reliable the resulting estimate of μ was. Clearly these authors believe that some fact about "thousands of such intervals" is relevant to "this interval", and that this relevance is too obvious to need stating. What can they mean?

Once the question is asked, the answer is clear. They are applying a straightforward proportional syllogism, arguing from

> Of thousands of confidence intervals calculated by the given method, 95% would cover μ,

to

> This confidence interval, calculated by the given method, covers μ.

That proportional syllogism, like any other, is part of logical probability. Therefore, the frequentists must be committed to a central tenet of logical probability. If they are not, then they must assert that

> Of thousands of confidence intervals calculated by the given method, 95% would cover μ,

is *no* reason to believe

> This confidence interval, calculated by the given method, covers μ.

In that case, there is no point in calculating what would happen if thousands of confidence intervals were considered, or in making any of the other calculations in the method, since any such calculation will leave us as ill-informed about μ as when we started.

Neyman was to some degree aware of the problem, and it is thus entertaining to watch him explain what he is doing, given that it is not reasoning. As is common among philosophers who have rejected some familiar form of justificatory reasoning, he seeks to permit the conclusion on some "pragmatic" grounds:

The statistician ... may be recommended ... to state that the value of the parameter θ is within ... (Neyman 1937, p. 288)

Doubtless he may be recommended to do so, or equally to do the opposite, but is one recommendation more reasonable than the other? Or is it just an act of the statistician's, or Neyman's, will? It is the latter:

To decide to 'assert' does not mean 'know' or even 'believe'. It is an act of will. (Neyman 1938, p. 352)

But is the will supported in its leap into being by some reasoning? No, it is not:

We may *decide* to behave as if we actually knew that the true value ϑ_1 of θ_1 were between $\underline{\theta}(E')$ and $\overline{\theta}(E')$. This is done as a result of our *decision* and has nothing to do with 'reasoning' and 'conclusion'. (Neyman 1941, p. 379; Neyman's italics)

It is difficult to argue with pure acts of the will, for the same reason that it is difficult to argue with tank columns – the attempt is a category mistake, since arguments can only bear on propositions or other arguments. Neyman's remarks are best regarded as a case of that "vacuum-activity" which Stove found in a number of other deductivist authors: the kind of activity observed in dogs confined indoors, who "dig" in corners and "bury" imaginary bones (Stove 1982, pp. 96–7). The philosopher who denies himself a chain of reasoning leading to some conclusion will need to

produce a series of something else, in this case acts of will, which mimics the structure of the reasoning and so leads to the same place.

Neyman is not the only frequentist, but was by far the most influential in the statistics community. In any case, all the other frequentists exhibit exactly the same problem. Von Mises and Reichenbach regarded probability as only properly definable in a infinite sequence of throws, or a "collective", or as a limit of relative frequencies in finite classes. As was always recognised, this created a difficulty in explaining how a probability so defined could be relevant to belief, decision or action in the single case. Reichenbach's answer was to advise looking at the narrowest feasible reference class containing the event of interest, and then making a "posit" on the basis of the relative frequency observed therein:

Assume that the frequency of an event B in a sequence is $= \frac{5}{6}$. Confronted by the question whether an individual event B will happen, we prefer to answer in the affirmative because, if we do so repeatedly, we shall be right in $\frac{5}{6}$ of the cases ... not as an assertion but as a *posit* ... it refers to the posit that will be the most successful when applied repeatedly. (Reichenbach 1949, pp. 372–3)

To say there is a hidden proportional syllogism here is surely too charitable. It could hardly be more overt.

8. POPPER WAS A SECRET LOGICAL PROBABILIST

For completeness, let us explain how the core doctrine that made Popper's opinions on the philosophy of science so widely attractive is a theorem of logical probability. Popper, it is true, himself claimed that scientific theories do not aim at probability. But his reason for saying so was merely the observation that if probability were the aim, it could best be achieved by making theories nearly vacuous. Obviously, that only shows that science does not pursue probability at all costs – a fact evident enough, as science does not aim to replicate indefinitely the result $1 + 1 = 2$, either, as it could if probability were all it sought. Popper makes this very point:

if you value high probability, you must say very little – or better still, nothing at all: tautologies will always retain the highest probability. (Popper 1959, p. 270 n. 3)

This is an argument parallel to "If you value peace, you must kill yourself." The reply is equally obvious: to value something is not to value it exclusively. Scientists and logical probabilists can and do value content as well as probability in theories, and strive to combine them.

Popper claimed instead that theories should be "contentful" or "testable" and aim to survive "rigorous" or "severe" tests. The natural interpretation of these claims, though one resisted by official Popperian ideology, is

that it is a good thing for a theory (in short, it increases the theory's probability substantially) if it is initially improbable, but predicts something that would be unlikely if it were false. For what is "content" (in the sense Popper uses it, in contrast to vacuousness) but initial improbability? ("a theory which asserts more, and thus takes greater risks, is better testable": Popper 1963, p. 966). And what is "rigour" or "severity" in a test but the high chance that the theory's falsity will be detected by the test, if it is in fact false? ("severe tests … the results of serious attempts to refute the theory, or of trying to fail it where failure would be expected in the light of all our knowledge" (Popper 1963, p. 964; cf. Popper 1962, pp. 112, 36, 390)). There is little hope of making sense of that in a purely deductivist framework (Grünbaum 1976). And once out of deductivism, one cannot pick and choose which theorems of probability one will accept. Indeed, the central Popperian claim is simply one theorem of logical probability:

> If $P(h \mid b)$ is low, and h & b implies e, while $P(e \mid \text{not-}h$ & $b)$ is low, then $P(h \mid e$ & $b)$ is much higher than $P(h \mid b)$

("If hypothesis h is initially unlikely – on background evidence b – then a consequence of h, which is unlikely without h, increases its probability greatly.")

It is certainly ironic that Popper, avowed opponent of probability, piggy-backed to fame on a single theorem of logical probability. Surely all philosophers who have discussed Popper with scientists have found that the scientists believe Popper's contribution was to show that one cannot be *certain* of universal generalizations on the basis of any finite amount of observation, and to urge that tests should be severe. If it is further pointed out to the scientists that everyone agrees with those theses, and that Popper's contribution was in fact to say that the theories are no more probable even after severe testing, that is treated as a minor technicality of little interest. The tendency of scientists to baptize principles of logical probability with Popper's name is well illustrated by a remark of Eccles, one of the most vociferously Popperian of scientists:

Often I have only a very vague horizon-of-expectations (to use POPPER'S felicitous phrase) at the beginning of a particular investigation, but of course sufficient to guide me in the design of experiments; and then I am able to maintain flexibility in concepts which are developed in the light of the observations, but always of a more general character so that the horizon-of-expectations is greatly advanced and developed and itself gives rise to much more rigorous and searching experimental testing … (Eccles 1970, p. 108)

And Peter Medawar, author of the much-quoted remark, "I think Popper is incomparably the greatest philosopher of science that has ever been" (Medawar 1972), undermines his adulation with his account of what he think Popper said:

Popper takes a scientifically realistic view: it is the daring, risky hypothesis, the hypothesis that might so easily not be true, that gives us special confidence if it stands up to critical examination. (Medawar 1969, p. 51; cf. Milne 1995)

It is true that all such general principles are simple consequences of Bayes' theorem, so subjective Bayesians as well as logical probabilists are entitled to them, but Popper's concern is with objectivity in science, so if he had accepted a probabilistic interpretation of his words, he would have been closer to logical probabilists. This is especially evident in his explanation of how his "corroboration" of hypotheses differs from logical probability. "Corroboration" is intended to be a measure of everything good about a theory, and thus is to increase "in inverse ratio to its logical probability" (Popper 1959, p. 270) (meaning, evidently, its *prior* logical probability; Popper does not object to the *concept* of logical probability, he merely believes it has no relevance to theory evaluation.) It thus appears at first that corroboration is something like the ratio of posterior to prior probability, which would indeed increase in inverse ratio to (prior) logical probability, and would, as intended, measure "the severity of tests to which a theory has been subjected" (Popper 1959, p. 387). It appears further, however, that corroboration is intended to include a measurement of certain other desirable features of theories, such as simplicity (Popper 1959, p. 270 n. 3). These dimensions are, however, also purely logical, and are the kind of property of theories admitted by logical probabilists as having a bearing on the probability of hypotheses. So Popper's position is indeed a hidden logical probabilism, rather than a hidden Bayesianism.

REFERENCES

Achinstein, P.: 1994, 'Stronger Evidence', *Philosophy of Science* **61**, 329–350.

Andersen, K. A. and J. N. Hooker: 1994, 'Bayesian Logic', *Decision Support Systems* **11**, 191–210.

Armore, S. J.: 1966, *Introduction to Statistical Analysis and Inference*, Wiley, New York.

Barrett, J. A.: 1996, 'Oracles, Aesthetics and Bayesian Consensus', *Philosophy of Science* **63**(3), supplement, S273–S280.

Berger, J. O. and J. M. Bernardo: 1992, 'On the Development of Reference Priors', in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, (eds), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, Clarendon, Oxford, pp. 35–60.

Carnap, R.: 1952, *The Continuum of Inductive Methods*, Chicago University Press, Chicago.

Castell, P.: 1998, 'A Consistent Restriction of the Principle of Indifference', *British Journal for the Philosophy of Science* **49**, 387–395.

Changizi, M. A. and T. P. Barber: 1998, 'A Paradigm-based Solution to the Riddle of Induction', *Synthese* **117**, 419–484.

Charniak, E.: 1991, 'Bayesian Networks without Tears', *AI Magazine*, **12**(4), 50–63.

Cohen, L. J.: 1986, 'Twelve Questions About Keynes' Concept of Weight', *British Journal for the Philosophy of Science* **37**, 263–278.

Cox, R. T.: 1946, 'Probability, Frequency and Reasonable Expectation', *American Journal of Physics* **14**, 1-13.

Cox, R. T.: 1961, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore.

Davidson, B. and R. Pargetter: 1985, 'In Defence of the Dutch Book Argument', *Canadian Journal of Philosophy* **15**, 405–424.

Davidson, B. and R. Pargetter.: 1987, 'Guilt Beyond Reasonable Doubt', *Australasian Journal of Philosophy* **65**, 182–187.

De Finetti, B.: 1964, 'Foresight: Its Logical Laws, Its Subjective Sources', in H. E. Kyburg and H. E. Smokler (eds), *Studies in Subjective Probability*, Wiley, New York, pp. 95–158.

Dennett, D.: 1987, *The Philosophical Lexicon*, http://www.blackwellpublishers.co.uk/lexicon/

DeRobertis, L. and J. A. Hartigan: 1981, 'Bayesian Inference Using Intervals of Measures', *Annals of Statistics* **9**, 235–244.

Diodorus Siculus [1933] *History*, Heinemann, London.

Earman, J.: 1992, *Bayes or Bust?*, MIT Press, Cambridge, MA.

Eccles, J. C.: 1970, *Facing Reality*, Springer-Verlag, New York.

Eggleston, R.: 1983, *Evidence, Proof and Probability*, 2nd edn, Weidenfeld and Nicolson, London.

Festa, R.: 1993, *Optimum Inductive Methods*, Kluwer, Dordrecht.

Forrest, P.: 1986, *The Dynamics of Belief*, Blackwell, Oxford.

Forster, M. R.: 1995, 'Bayes and Bust: Simplicity as a Problem for a Probabilist's Approach to Confirmation', *British Journal for the Philosophy of Science* **46**, 399–424.

Franklin, J.: 1987, 'Non-deductive Logic in Mathematics', *British Journal for the Philosophy of Science* **38**, 1–18.

Franklin, J.: 1991, 'Healthy Scepticism', *Philosophy* **66**, 305–324.

Ghosh, J. K. and R. Mukerjee: 1992, 'Non-informative Priors', in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, Clarendon, Oxford, pp. 195–210.

Greenland, S.: 1998, 'Probability Logic and Probabilistic Induction', *Epidemiology* **9**, 322–332.

Grünbaum, A.: 1976, 'Is the Method of Bold Conjectures and Attempted Refutations *Justifiably* the Method of Science?', *British Journal for the Philosophy of Science* **27**, 105–136.

Guenther, W. C.: 1965, *Concepts of Statistical Inference*, McGraw-Hill, New York.

Haack, S.: 1976, 'The Justification of Deduction', *Mind* **85**, 112–119, reprinted in Haack, S.: 1996, *Deviant Logic, Fuzzy Logic*, Chicago University Press, Chicago, pp. 183–191.

Hartigan, J.: 1964, 'Invariant Prior Distributions', *Annals of Mathematical Statistics* **35**, 836–845.

Heidbreder, G. R.: 1996, ed., *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht.

Hempel, C. G.: 1965, *Aspects of Scientific Explanation*, Free Press, New York.

Horgan, J.: 1991, 'In the Beginning …', *Scientific American* **64**(2), 100–109.

Howson, C.: 1992, Review of Earman, *Bayes or Bust, Nature* **358**, 552.

Howson, C.: 1995, 'Theories of Probability', *British Journal for the Philosophy of Science* **46**, 1–32.

Howson, C.: 1997, 'Logic and Probability', *British Journal for the Philosophy of Science* **48**, 517–531.

Howson, C. and P. Urbach: 1993, *Scientific Reasoning: The Bayesian Approach*, 2nd edn, Open Court, Chicago.

Jaynes, E. T.: 1957, 'Information Theory and Statistical Mechanics', *Physical Review* **106**, 620–660.

Jaynes, E. T.: 1968, 'Prior Probabilities', *IEEE Transactions on Systems Science and Cybernetics* **4**, 227–241.

Jaynes, E. T.: to appear, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge.

Johnson-Laird, P. N.: 1983, *Mental Models*, Cambridge University Press, Cambridge.

Kass, R. E. and L. Wasserman: 1996, 'The Selection of Prior Distributions by Formal Rules', *Journal of the American Statistical Association* **91**, 1343–1370.

Kendall, M. and A. Stuart: 1977, *The Advanced Theory of Statistics*, 4th edn, Griffin, London.

Keynes, J. M.: 1921, *Treatise on Probability*, Macmillan, London.

Koopman, B. O.: 1940, 'The Bases of Probability', *Bulletin of the American Mathematical Society* **46**, 763–774.

Kyburg, H. E.: 1974, *The Logical Foundations of Statistical Inference*, Reidel, Dordrecht.

Levi, I.: 1974, 'On Indeterminate Probabilities', *Journal of Philosophy* **71**, 391–418.

Maher, P.: 1993, *Betting on Theories*, Cambridge University Press, Cambridge.

Maher, P.: 1996, 'Subjective and Objective Confirmation', *Philosophy of Science* **63**, 149–174.

Medawar, P.: 1972, BBC Radio 3 broadcast, 28 July 1972, quoted in Magee, B.: 1973, *Popper*, at p. 9. Collins, London.

Medawar, P. B.: 1969, *Induction and Intuition in Scientific Thought*, Methuen, London.

Milne, P.: 1983, 'A Note on Scale Invariance', *British Journal for the Philosophy of Science* **34**, 49–55.

Milne, P.: 1995, 'A Bayesian Defence of Popperian Science', *Analysis* **55**, 213–215.

Mura, A.: 1990, 'When Probabilistic Support is Inductive', *Philosophy of Science* **57**, 278–289.

Neyman, J.: 1937, 'Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability', *Philosophical Transactions of the Royal Society of London* A **236**, 333–380 (reprinted in Neyman, 1967, pp. 250–290).

Neyman, J.: 1938, 'L'estimation Statistique Traitée Comme un Problème Classique de Probabilité', *Actualitées Scientifiques et Industrielles* **739**, 25–57 (reprinted in Neyman, 1967, pp. 332–353).

Neyman, J.: 1941, 'Fiducial Argument and the Theory of Confidence Intervals', *Biometrika* **32**, 128–150 (reprinted in Neyman, 1967, pp. 375–394).

Neyman, J.: 1967, *A Selection of Early Statistical Papers of J. Neyman*, University of California Press, Berkeley.

O'Donnell, R.: 1992, 'Keynes' Weight of Argument and Popper's Paradox of Ideal Evidence', *Philosophy of Science* **59**, 44–52.

O'Neill, L.: 1996, 'Indifference and Induction', in P. J. Riggs (ed.), *Natural Kinds, Laws of Nature and Scientific Methodology*, Kluwer, Dordrecht, pp. 93-102.

Pan, Y. and B. Yuan: 1997, 'Bayesian Inference of Fuzzy Probabilities', *International Journal of General Systems* **26**, 73–90.

Paris, J. B.: 1994, *The Uncertain Reasoner's Companion: A Mathematical Perspective*, Cambridge University Press, Cambridge.

Pearl, J.: 1993, 'Belief Networks Revisited', *Artificial Intelligence* **59**, 49–56.

Popper, K.: 1959, *The Logic of Scientific Discovery*, Hutchinson, London.

Popper, K.: 1962, *Conjectures and Refutations*, Basic Books, New York.

Popper, K.: 1963, 'Science: Problems, Aims, Responsibilities', *Federation Proceedings* **22**, 961–972.

Price, R.: 1763, Appendix to T. Bayes, 'Essay Towards Solving a Problem in the Doctrine of Chances', *Philosophical Transactions of the Royal Society* **53**, 370–418, reprinted in *Biometrika* **45** (1958), 293–315.

Ramsey, F. P.: 1926, 'Truth and Probability', in D. H. Mellor (ed.), *Philosophical Papers*, Cambridge University Press, Cambridge, 1990, pp. 52–94.

Reichenbach, H.: 1949, *The Theory of Probability*, University of California Press, Berkeley.

Rosenkrantz, R. D.: 1994, 'Bayesian Confirmation: Paradise Regained', *British Journal for the Philosophy of Science* **45**, 467–476.

Runde, J.: 1991, 'Keynesian Uncertainty and the Instability of Beliefs', *Review of Political Economy* **3**, 125–145.

Runde, J.: 1994, 'Keynes After Ramsey: in Defence of *A Treatise on Probability*', *Studies in History and Philosophy of Science* **25**, 97–121.

Salmon, W. C.: 1966, *The Foundations of Scientific Inference*, University of Pittsburgh Press, Pittsburgh.

Shafer, G.: 1976, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton.

Snare, F.: 1980, 'The Diversity of Morals', *Mind* **89**, 353–369.

Snare, F.: 1984, 'The Empirical Bases of Moral Scepticism', *American Philosophical Quarterly* **21**, 215–225.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, R. G.: 1993, 'Bayesian Analysis in Expert Systems', (with commentaries) *Statistical Science* **8**, 219–283.

Steel, R. G. D. and Torrie, J. H.: 1976, *Introduction to Statistics*, McGraw-Hill, New York.

Steenstrup, S.: 1984, 'Experiments, Prior Probabilities, and Experimental Prior Probabilities', *American Journal of Physics* **52**, 1146–1147.

Stove, D. C.: 1982, *Popper and After: Four Modern Irrationalists*, Pergamon, Oxford.

Stove, D.C.: 1986, *The Rationality of Induction*, Clarendon, Oxford.

Strevens, M.: 1998, 'Inferring Probabilities from Symmetries', *Nous* **32**, 231–246.

Swinburne, R. G.: 1971, 'The Paradoxes of Confirmation – A Survey', *American Philosophical Quarterly* **8**, 318–329.

Talmud, Yebamoth [1936], trans. I. W. Slotki, in *The Talmud*, translated under the editorship of I. Epstein, Soncino Press, London.

Uffink, J.: 1996, 'The Constraint Rule of the Maximum Entropy Principle', *Studies in History and Philosophy of Modern Physics* **27B**, 47–79.

Van Fraasen, B.: 1989, *Laws and Symmetry*, Oxford University Press, Oxford.

Verbraak, H. L. F.: 1990, *The Logic of Objective Bayesianism*, Verbraak, Amsterdam.

Wasserman, L.: 1992, 'Recent Methodological Advances in Robust Bayesian Inference', in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, Clarendon, Oxford, pp. 483–501.

Walley, P.: 1991, *Statistical Reasoning With Imprecise Probabilities*, Chapman and Hall, London.

Walley, P., L. Gurrin, and P. Burton: 1996, 'Analysis of Clinical Data Using Imprecise Prior Probabilities', *The Statistician* **45**, 457–485.

Williams, D. C.: 1947, *The Ground of Induction*, Harvard University Press, Cambridge, MA.

School of Mathematics
The University of New South Wales
Sydney 2052
Australia