



## A Logic of Induction

Colin Howson

*Philosophy of Science*, Vol. 64, No. 2 (Jun., 1997), 268-290.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8248%28199706%2964%3A2%3C268%3AAALOI%3E2.0.CO%3B2-V>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Philosophy of Science* is published by The University of Chicago Press. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

---

*Philosophy of Science*

©1997 Philosophy of Science Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

# A Logic of Induction\*

Colin Howson†

Department of Philosophy, Logic, and Scientific Method,  
London School of Economics and Political Science

**1. Probabilism.** Statistics is probably the last discipline the ordinary person would associate with ideological wars, but one has been raging there for the last thirty years and more. Until recently the Classical, also known as Frequentist, theory of statistical inference dominated. But gradually a quite different approach has attracted adherents. This, named the Bayesian theory after the eighteenth-century English clergyman, Thomas Bayes, is a phoenix, reborn from the theory of inductive inference dominant from the mid-eighteenth to the late nineteenth century, which said that the measure of confidence proper to employ in an uncertain proposition is its probability (the idea goes back well beyond Bayes; in his great *Ars Conjectandi*, published posthumously in 1715, James Bernoulli stated that probability is degree of certainty (Part IV, Chapter II)).

Such an identification seems entirely natural: probability is a measure of uncertainty in ordinary speech. From there it seems an obvious step to saying that *if* nothing is known that should discriminate among the possible values of a real-valued parameter  $Z$  lying in a closed interval, for example  $[1/3, 3]$ , *then* uncertainty should be spread evenly along that interval, in this case with a uniform density  $3/8$ . But if nothing is known about  $Z$  then, equally, nothing is known about  $W = Z^{-1}$ , whose range is also  $[1/3, 3]$ . However, it is not difficult to see that probability cannot be uniformly distributed over  $W$ 's values if it is uniformly distributed over  $Z$ 's (suppose it is and compute the probability of the equivalent propositions  $Z > 2$  and  $W < 1/2$ ).

Such 'paradoxes' started to be noticed in the nineteenth century. The one above is a version of the wine/water paradox, where one's infor-

\*Received December 1996.

†Send reprint requests to the author, Department of Philosophy, Logic, and Scientific Method, London School of Economics and Political Science, Houghton St., London WC2A 2AE.

Philosophy of Science, 64 (June 1997) pp. 268–290. 0031-8248/97/6402-0004\$2.00  
Copyright 1997 by the Philosophy of Science Association. All rights reserved.

mation is only that wine is mixed with water in some ratio between one part water to three parts wine and three parts water to one part wine; Z and W represent the inverse ratios of each other. A similar and equally famous 'paradox' is due to the French mathematician Bertrand, who was interested in the probability that the length of a 'randomly' chosen chord in a circle is greater than that of the side of the inscribed equilateral triangle. The position of the chord can be specified in (at least) the following three ways: (i) the distance between the center of the chord and that of the circle, (ii) the angle the chord makes with a tangent to the circle at its end point, and (iii) the position of the center of the chord within the circle. But as Bertrand showed, uniform probability densities cannot simultaneously be assigned each of these.

Such problems represented a serious conceptual threat, ultimately a fatal one, to the probabilistic theory developed by Bayes and Laplace, for that theory was squarely based on the use of uniform distributions to represent an initial lack of knowledge. In a profoundly original paper (1763), Bayes had derived the *posterior distribution* of a binomial parameter, i.e., he had derived, in modern terminology, the form of the density over the possible values in  $[0,1]$  of a physical probability parameter conditional on evidence stating how many times in  $n$  observations the event with that probability had been observed. Laplace gave a more modern, analytical derivation, in which he used the following version of what has become known as Bayes's Theorem:

$$f(p|e_{r,n}) = k \int_0^1 P(e_{r,n}|p)f(p)dp$$

where  $f(p)$  is the prior density function,  $k$  is a normalizing constant and  $e_{r,n}$  states that the event had been observed  $r$  times out of  $n$ . Laplace's model (an uncountably infinite urn with 'proportion'  $p$  of black tickets) allowed him to assume, as had Bayes, that the successive observations are independent with a fixed probability  $p$  of observing the event in question (in the model, drawing a black ticket).

Granted that assumption,  $k^{-1} = P(e_{r,n})$  is equal to  ${}^nC_r p^r (1-p)^{n-r}$ , leaving the prior density  $f(p)$  as the only unknown. Here Laplace followed Bayes and adopted the postulate, discussed in a Scholium in Bayes' paper, that  $f(p)$  is constant, and hence equal to 1. Bayes justified the uniform distribution of  $p$  precisely with the argument that if we assume that we know nothing initially about the value of  $p$ —and hence represent it by a uniform prior density—then we have a way, if the calculation can be performed, of systematically incorporating the empirical information supplied the observational data  $e_{r,n}$  in the posterior

distribution of  $p$ . And as we see, with the uniform prior that calculation can be performed.

Laplace was further able to derive the famous *Rule of Succession*: that the probability of observing the event at the next trial, conditional on the evidence  $e_{r,n}$ , is equal to  $(r + 1)/(n + 2)$ . The dependence of this probability on the sample parameters  $r$  and  $n$  made this an especially remarkable result in the late eighteenth century, for it seemed to refute Hume. Laplace seemed to have proved that induction, reasoning from past to future, was valid, and without assuming in any way what it set out to justify; indeed, insofar as it can be called an assumption, all that is assumed is an evenly distributed ignorance (a very good discussion of the mathematical and philosophical background can be found in Chapter 1 of Earman 1992, in which Bayes's posterior distribution for  $p$  is derived).

Yet, as we have seen, probabilistically-modelled ignorance appears to generate inconsistencies. The alternative to an 'ignorance' distribution is to represent one's non-null current state of information by the prior distribution, and update this distribution, by means of Bayes' theorem, on receipt of new information. The trouble with this idea is that the probabilistic representation of a state of information is supposed to be inferred, not assumed as a datum. If it is assumed, it cannot be held to be objectively determined, and hence any particular choice of prior must give that prior a subjective character, to be inherited by the posterior distribution. While this may be acceptable in the context of a personal decision problem, it certainly was not felt to be so in the context of scientific inference, in which the degree to which evidence supports a hypothesis is calculated. In R. A. Fisher's (highly influential) words, the reliance on uniform prior distributions

leads to apparent mathematical contradictions. In explaining these contradictions away, advocates of inverse probability seem forced to regard mathematical probability . . . as measuring merely psychological tendencies, theorems respecting which are useless for scientific purposes (1947, 6–7)

A desire to avoid the sort of subjectivism condemned by Fisher led the geophysicist Sir Harold Jeffreys in the 1920s and 1930s to consider ways of formulating more satisfactory objective criteria for determining prior distributions. One such criterion was that simpler hypotheses should receive larger prior probabilities, where a hypothesis is simpler the fewer independent adjustable parameters it contains. This he called the *Simplicity Postulate*. A problem with it is that it is difficult to implement consistently, and Jeffreys himself proposed other criteria actually inconsistent with it. One of these, which aroused and still arouses

a good deal of interest, is that of invariance. A rule for assigning a prior distribution to a parameter  $t$  is *invariant* if it assumes the same form under invertible differentiable transformations of  $t$ . In other words, the distribution is one which can be specified in a coordinate-invariant way. An idea pursued by Jeffreys was to combine the requirement of ignorance and that of invariance by constructing a rule for assigning priors which depends *only* on the probability model for the distribution of the observation variable  $X$ , via its class of conditional densities for  $X$  given a parameter  $t$ , and any smooth reparameterization of  $t$ .

One such invariant rule, now known as Jeffreys' rule, says that the prior for  $t$  should be the square root of the Fisher information, i.e. of the expected value, with respect to  $x$  given  $t$ , of

$$\partial^2 \ln p(x|t) / \partial t^2,$$

where  $p(x|t)$  is the density assigned by the model at the point  $x$ , given  $t$ , i.e., the likelihood of  $t$  given  $x$ . The priors that arise from this rule are often called Jeffreys priors. Where  $I$  is the information,  $t$  the parameter(s), and  $s$  is any differentiable transformation of  $t$ , we have that  $I(s)^{1/2} ds = I(t)^{1/2} dt$ . Consequently, Jeffreys's rule "could be stated for any law that is differentiable with respect to all parameters in it, and would have the property that the total probability in any region of the [the parameter  $t$ ] would be equal to the total probability in the corresponding region of [the new parameter  $s$ ]; in other words, it satisfies the rule that equivalent propositions have the same probability" (Jeffreys 1961, 181).

Invariant rules are not vulnerable to the transformational paradoxes. On the other hand any particular choice of one, like Jeffreys' rule, seems rather ad hoc. And Jeffreys' rule can give strange results: for example, the joint prior for the mean and standard deviation of a normally distributed variate is (constant,  $\sigma^{-2}$ ), whereas taken separately they are constant and  $\sigma^{-1}$  respectively. Jeffreys priors are also attended by more or less severe technical problems. First, they may not exist. Even where they do, they are often improper, i.e., their integrals diverge. The Jeffreys priors for the mean and standard deviation of a normal variate are clearly both improper. Improper priors are sometimes justified as approximations to proper distributions within the intended range, but when they are justified by some general principle this defense is hardly available. And there are other problems of consistency; see Dawid, Stone, and Zidek 1973. Formally and otherwise, the situation cannot really be called a satisfactory one.

A somewhat different approach to invariance was subsequently proposed by the physicist E. T. Jaynes. Jaynes claimed that an inference

problem will characteristically determine a group of transformations of relevant quantities under which the problem remains essentially unaltered, and hence under which the target distribution should be invariant. Jaynes called a problem well-posed if there is a unique distribution invariant under this group. He claims that Bertrand's problem is well-posed. "Neither Bertrand's original statement nor our restatement in terms of straws [thrown into a circle to represent randomly drawn chords] specifies the exact size of the circle, or its exact location" (1973, 480). The solution distribution should, therefore, be invariant under scale and location transformations, and Jaynes shows that these constraints determine it uniquely up to a multiplicative constant.

Employing the symmetries implicit in the statement of a problem sounds an impressive way to obtain an objectively determined prior. But on analysis it is not clear that it marks any advance. Suppose that the required prior distribution is of a parameter  $t$ , and the problem is stated in the form  $\Phi(t)$ . Let  $s(t)$  be any invertible transformation of  $t$ . Then, modulo some mathematics, there is a logically equivalent formulation  $\Phi'(s)$ ; it is the same problem, merely differently expressed. But then we should, it seems, require the prior distributions of  $s$  and  $t$  be the same, for any  $s$ , which is impossible.

Jaynes avoids this conclusion by restricting those transformations under which the problem remains unaltered to a proper subgroup of the group of all invertible continuous transformations. For example, in the Jaynesian literature, where  $t$  is a parameter representing a physical magnitude, the lack of any specification of scale for the measurement is usually taken to mean that  $f(t)$  should be invariant under transformations of the form  $at$ ,  $a > 0$ , determining  $f(t)$  as proportional to  $t^{-1}$  (but see Milne 1983, 52–53). But why consider only ratio scales? Clearly, it all depends what one means by 'equivalent formulations of the problem.' If, as I believe, it is arbitrary to restrict equivalence to some proper subspecies of logical equivalence, then no problem is well-posed.

Jaynes's other well-known proposal for determining priors, the *rule of maximum entropy*, founders ultimately on the same problem. This rule instructs choosing among all the possible distributions consistent with the available background information that which has maximum entropy subject to the constraints the information imposes, wherever a unique such distribution exists (Jaynes 1968). Problems of existence aside, the only entropy measure which consistently generalizes to continuous densities  $f(t)$  is the so-called cross-entropy, i.e., the expected value of  $-\log[f(t)/g(t)]$ , logarithm to any base, where  $g$  is a reference distribution which Jaynes suggests should be chosen according to his invariance theory above (ibid.). And so we are back again. It seems

fair to say that the problem of representing methodological ignorance has not been overcome.

**2. Fisher and Neyman-Pearson.** The problem of objective priors seems to be a difficult if not insoluble one. But no theory gets abandoned until an acceptable alternative is available to take its place. Such an alternative to the probabilistic account appeared, at any rate for statisticians, in the second and third decades of this century, when in a remarkable series of papers Fisher laid the foundations for what appeared to be an objective, non-probabilistic theory of statistical inference. Fisher substituted for the idea that a hypothesis asserting some systematic effect could be rendered probable by appropriate observations, the idea that a null hypothesis  $h_0$ , asserting that no such effect exists, could be safely rejected by a test of significance. An outcome  $x$  of such a test is said to be significant at the  $p\%$  level if the probability of the set of outcomes at least as discordant with  $h_0$  as  $x$  is at most  $p/100$  according to  $h_0$ . How discordant any observation  $y$  is with  $h_0$  is typically the likelihood of  $h_0$  relative to a the value  $t(y)$  of a suitable test statistic  $t$  (the likelihood of  $h_0$  relative to  $t(y)$  is the probability density of  $t(y)$  according to  $h_0$ ).

An outcome significant at a small enough significance level is a ground for rejecting  $h_0$  because such an outcome would almost never occur were  $h_0$  true. Thus a straightforward falsificationist theory of sound inference seemed to have been discovered by Fisher, and by contrast with the earlier probabilistic theory, wholly objective because framed in terms of objective statistical probabilities. Small wonder that Fisher's theory swept the earlier one aside. By one of those prescient strokes of history, at the same time and independently of Fisher, Karl Popper in Vienna was putting the finishing touches to his similarly falsificationist theory of test-evaluations of deterministic hypotheses, also designed to replace the old theory based on prior and posterior probabilities.

Fisher's theory was not without its own problems, however, and it was the analysis of these that produced the theory of statistical inference, the Frequentist theory, that in time absorbed it and then dominated the field after the Second World War. Jerzy Neyman who, together with E. S. Pearson, was responsible for this development of Fisher's original theory, pointed out that however desirable from the point of view of the various criteria defined by Fisher was a test statistic  $t$ , a suitable transformation of  $t$  determined another statistic  $t'$  with the same desirable test characteristics, except that a significant outcome with respect to  $t$  is not significant with respect to  $t'$ , and vice versa (Neyman 1952, 43–54).

Neyman and Pearson's solution of the problem is developed from a denial that hypotheses can be tested in isolation. It can be argued that the claim is anyway implicit in the rationale of significance tests of the null hypothesis: knocking out  $h_0$  is regarded as establishing the alternative hypothesis of systematic effect. At any rate, once an alternative  $h_1$  to the null  $h_0$  is introduced explicitly in a test of  $h_0$  against  $h_1$ , it becomes possible to design tests according to the desideratum of simultaneously keeping small the chances of two important types of error, called type 1 and type 2 errors respectively, of incorrectly rejecting  $h_0$  when it is true, and of incorrectly accepting it when it is false. One minus the chance of the type 2 error is known as the *power* of the test.

Suppose that  $h_0$  and  $h_1$  are regarded, at any rate given current knowledge, as exhausting the class of plausible alternatives. If we identify for the time being the chance of a type 2 error with the probability of accepting  $h_0$  when  $h_1$  is true, then if both  $h_1$  and  $h_0$  generate probability distributions over a test statistic  $t$ , it is possible to prove Neyman and Pearson's *Fundamental Lemma*: there exists a unique region, a 'best critical region', in the range of  $t$  for rejecting  $h_0$  which minimizes the probability of a type 2 error for any assigned probability of a type 1 error. The probability of the type 1 error is reckoned according to  $h_0$  and that of the type 2 error according to  $h_1$ . The Fundamental Lemma automatically solves the problem of the previous paragraph. The best critical region is always defined by an inequality of the form  $p_0(t)/p_1(t) \leq k$ , where  $k$  is determined by the preassigned probability of type 1 error and  $p_0$  and  $p_1$  are the probability densities determined by  $h_0$  and  $h_1$ . This inequality is invariant under all continuous invertible transformations of  $t$  (since the Jacobian appears in both numerator and denominator and consequently cancels), and hence under that which caused the problem for Fisher's own test criteria.

Suppose, however, that the hypothesis we want to test  $h_0$  against is not one which determines a unique probability distribution, but one instead which is equivalent to an entire family  $H$  of distributions alternative to the pure chance distribution specified by  $h_0$ . For example, the alternative to  $h_0$  may merely claim a systematic effect of a more or less extensive type. However, it may still be possible to show that there exists a region in the outcome space most powerful against all members of  $H$ . Failing that, textbooks of the Neyman-Pearson theory propose subsidiary criteria. It is not the details of the Neyman-Pearson theory that I am concerned with here, however, but the central thesis it and Fisher's theory oppose to probabilism: that sound inductive rules exist which do not appeal to the probabilities of the hypotheses to be evaluated empirically, i.e., to what are often called epistemic probabilities, or what Carnap called probabilities<sub>1</sub>, but only to the statistical prob-



abilities, probabilities<sub>2</sub> (Carnap), determined by the hypotheses themselves. We shall resume the discussion in Section 4.

**3. Probabilism Reborn.** It is now time to take up again the probabilist(,) story again. In the 1920s, two people, Frank Ramsey in Cambridge (England) and Bruno de Finetti in Rome, working independently, created a revolution within probabilism. They both started by trying to explain what all their predecessors had simply taken to be axiomatic, that the structure of degrees of certainty is that of mathematical probability. The upshot of their investigation was that the probabilistic structure is a condition of consistency in the individual's evaluations of uncertainty. This led to a completely novel probabilistic theory of inductive inference, called the subjective Bayesian theory.

Ramsey and de Finetti proceeded in different ways that have since generated two quite different lines of research, one utility-theoretic, the other based on rational betting strategies. The second, the one I shall outline here, stems from a classic paper of de Finetti (1937). This starts by characterizing an individual's uncertainty in terms of equilibrating betting quotients. Some terminology: a *bet* is a contract in which one individual X contracts to obtain from another Y a sum Q if some proposition a is true, and pay them a sum R if a is false. The odds accepted by X on a are R:Q, those accepted by Y against a are Q:R. The normalized odds  $p = R:(R + Q)$  are called the *betting quotient* on a, and  $S = R + Q$  is called the stake.

The bet can now be recast in the p,S system, as a contract in which X receives  $S(1 - p)$  if a is true, and loses  $pS$  if not. Suppose S is fixed and p allowed to vary. If you have very clear ideas about the likelihood of a, then there will be a value of p below which you see an advantage to X, and at that value and above you see an advantage to Y. We can call this cross-over value your *degree of belief* in a. It might be that you do not have such clear opinions that you can determine a unique cross-over value. It might be that there is merely an interval of values; if this is so, the lower endpoint is called your lower probability of a, and the upper endpoint your upper probability of a. Classical Bayesianism is an idealising theory which supposes that upper and lower probabilities coincide for every proposition (for what happens when this assumption is relaxed see Walley 1991).

Now let us see why degrees of belief so measured should be formally probabilities. Consider a system of bets on a set of specified propositions. De Finetti showed (though Ramsey also knew the result, which requires only elementary algebra to prove) that if the bookmaker has no control over the magnitude and sign of the stakes, then they can be made to lose a positive sum independently of the truth-values of the

propositions bet on if their betting quotients do not satisfy the (finitely or countably) additive probability axioms: the bookmaker can have a *Dutch Book* made against them, in betting jargon, and for this reason Ramsey's and de Finetti's result is usually called the *Dutch Book Argument*.

Of course, one is usually in no danger of being placed oneself in such a betting situation. Nor is it realistic to suppose that all the propositions whose truth-values one has opinions about will have their truth-values veraciously decided in one's lifetime if ever. But the scenario above should not be taken literally; the way to see what it tells us is to regard it as a thought-experiment (Howson and Urbach 1993, Ch. 5; Hellman, this volume). Thus, suppose the betting quotients involved are your cross-over ones. This means that they are fair in your eyes, in that you believe that they equilibrate the advantage between the two sides of the bet. But if they do not obey the probability axioms then it is possible to arrange the bets in such a way that you can calculate in advance that the bookmaker must lose. This seems to mean that your assessment of each as fair is erroneous.

Of course, we need an assumption that bets that are individually fair cannot in combination generate a positive net loss or gain in all circumstances, but this seems a reasonable assumption, and if we equate 'calculable advantage' with 'expected value' then it is a demonstrably true one (Howson and Urbach 1993; Hellman, this volume). Granted all this, de Finetti's result implies that a necessary condition for betting quotients to be fair is that they satisfy the probability axioms. Thus you are demonstrably inconsistent in your assessment of fair betting quotients if they are not formally probabilities.

The converse is also provable: if your fair betting quotients are formally probabilities then it is impossible to force a net gain or loss in all circumstances. De Finetti called a set of beliefs *coherent* if the betting quotients they represent is immune to a Dutch Book. Ramsey used the overtly logical language of consistency. I shall follow Ramsey because, just as did Ramsey himself, I shall claim later on that implicit in the view of the probability calculus as a set of consistency constraints on betting quotients is that of the calculus as a set of demonstrably sound logical axioms in an inductive logic.

Maher has objected that this way of measuring belief, and the Dutch Book argument based on it, rest on assumptions which are false: in particular, that the value of  $p$  which you think gives zero advantage to the bettor is independent of  $S$  if  $S$  is measured in non-utility units (1997). Maher supports this claim by reference to a dictionary of the English language, according to which 'advantage' can be glossed as 'benefit, profit or gain' (*ibid.*). He points out that a bet at even odds

on the outcome of a toss of a fair coin with a stake of \$10 would have zero advantage to him, while one with a stake of \$10,000 would definitely be disadvantageous, because there would be a 'benefit, profit or gain' in ridding himself of it; indeed, he would pay to do so.

But consider this. The definition of 'computable' given by the Oxford English Dictionary is 'calculable'. However, the function  $f$  from natural numbers to natural numbers defined by

$$\begin{aligned} f(x) &= 1 \text{ if Goldbach's Conjecture is true} \\ &= 0 \text{ if not} \end{aligned}$$

is not at present, and possibly may never be, calculable. Yet (assuming a mild Platonism)  $f$  is a constant function and so *is* computable according to Turing, Markov, Gödel/Herbrand and Church, all of whom explicated what they thought to be essential features of the preformal notion of computability, in ways that turned out to be equivalent and which are now the basis of the mathematical theory of computability.

Let us return to bets. Following a long tradition, I have understood 'advantage' as a *bias* in a bet, so that the advantage to you in a bet against me is equal to the advantage to me if and only if the advantage to both is zero, with zero advantages adding over bets. A *fair* bet is one in which there is no advantage. These properties of 'advantage' were traditionally explicated as *expected cash value*: Laplace explicitly defines 'advantage' in this way (1951, 20). That explication is not *presupposed* (see Maher 1987); but it is an illuminating way, once partial belief is represented by probabilities, of giving mathematical expression to the idea that a bet at even money on heads with a throw of a fair coin is fair, independently of the stake and the fortunes of the players. Maher, who sees a positive disadvantage to *either* side of that bet where the stake is \$10,000, is clearly appealing to an entirely different idea, one which as he tells us is explicated in terms of utility. To me, as to Hellman (1977), utility, even if one can make satisfactory sense of it (and the evidence is increasingly unfavorable), is conceptually quite distinct from judgments of the biasedness or otherwise of bets.

Maher is not alone, however, in (as I believe) conflating the two notions: many people, Ramsey included, are convinced that belief can be measured by the agent's fair betting quotients only if all payoffs are expressed in utility units. Ramsey explicitly developed his theory of utility for this purpose and his procedure has become classic: it was to state plausible axioms for consistent preferences among options, in such a way that it can be proved that those axioms determine (i) a unique probability function, (ii) a utility function unique up to positive affine transformations, and (iii) the ranking of options by expected utilities. Savage later provided a different set of axioms to Ramsey's,

which became canonical for later work in the field (Savage 1954; in Ramsey's theory an agent's probability is explicitly defined as their utility-valued betting quotients, whereas in Savage's they are obtained, like the utilities, by a representation theorem). The expected utility principle is, however, charged with yielding strongly counterintuitive evaluations, and its status remains controversial (the current debate is nicely represented in the collection of Gärdenfors and Sahlin 1988).

Yet another approach, also due to de Finetti, exhibits betting in the wider context of so-called scoring rules; subsequently Lindley showed that virtually any set of scoring rules leads to the probability axioms (Lindley 1982). The view implicit in all these results is that probabilism is the theory of consistent personal probability assignments, which has the consequence that those assignments are formally speaking no more than exogenously determined parameters. Why, in this case, is new probabilism any advance on old probabilism? Take for example the old problem of representing ignorance. Where uncertainty is modelled probabilistically, uniform uncertainty, across all transformations of a parameter  $t$ , is mathematically impossible. Since new probabilism sees the probability axioms as necessary and sufficient conditions of consistency, no prior distribution is endorsed by it. All it tells you is that if your uncertainty about  $t$  is given by a uniform density over  $t$ , then your uncertainty about  $s(t)$  should be given by a density proportional to  $|ds/dt|$ .

But it is precisely the refusal to legislate about prior distributions, of course, that has drawn the strongest criticism, and specifically the charge that virtually anything is permitted. Indeed, the theory does seem to represent just the sort of psychologism that Fisher condemned as useless to science in the quotation earlier. There is a response to this standard objection that I believe is quite satisfactory, however, and that is to deny that such a theory is useless to science. On the contrary, to possess a demonstrably sound inductive logic, one whose pronouncements are both non-empty and avoid the force of Hume's powerful sceptical arguments, is a great intellectual achievement, on a par in significance with the contemporary development of formal deductive logic, also of course 'merely' a logic of consistency.

For new probabilism is indeed a genuine *logic*. The probability axioms are a sound and complete syntax with respect to the semantic criterion of consistency-coherence. Second, it is a genuinely *inductive* logic. The relation between evidence  $e$  and a hypothesis  $h$  is expressed in the conditional probability  $P(h|e)$ . There are many theorems of the probability calculus which express the inductive character of the relation. Here are two of the most fundamental:

- (1) If  $0 < P(h)$ ,  $P(e) < 1$  and  $h$  entails  $e$  then  $P(h|e) > P(h)$ . In other words, if  $e$  is a prediction of  $h$  then, just so long as you do not have unalterably dogmatic opinions about  $h$  and  $e$ , the truth of  $e$  increases the probability of  $h$ .

$$(2) P(h|e) = \frac{P(h)}{P(h) + \frac{\sum P(e|H_i)P(h_i)}{P(e|h)}}$$

This formula, a form of Bayes's Theorem, tells us that  $P(h|e)$  is sensitive to the proportional degree to which  $e$  is explained by  $h$  as opposed to any other plausible alternative hypotheses. This expresses a basic principle of good experimental design: it should be very unlikely that the sought effect  $e$  can be attributed to any cause other than  $h$  itself.

This inductive logic is immune to Humean objections because it makes no categorical assertion about the probability of any contingent proposition. And yet it is far from useless to science, because scientific inference requires sound rules, and sound rules are certainly provided by the theorems above (and all the other consequences of the probability axioms). We shall see dramatic evidence in the next section of the perils of being guided by an unsound theory.

**4. Infirmities of Neyman-Pearson Theory.** The Neyman-Pearson theory is at first sight weaker than the Bayesian in scope if nothing else, because of its apparent restriction to hypotheses: the Bayesian theory is completely general. However, first Giere (1984) and then Mayo (1996) have generalized the Neyman-Pearson ideas beyond purely statistical hypotheses, and claim validity for those ideas for scientific data and hypotheses in general. I shall argue in this section that the exercise is to no avail, since acceptance and rejection rules based solely on the chances of type 1 and type 2 errors, or what Mayo calls error probabilities, are demonstrably unsound. To avoid clumsy constructions I shall henceforward talk simply about large and small type 1 and type 2 errors, rather than large and small chances of type 1 and type 2 errors.

Neyman's justification of those rules is well-known. He argued that by repeatedly making the decisions to accept and reject supplied by a test with small type 1 and type 2 errors you will make those errors only a small proportion of the time. This answer has been heavily criticized, not least by Fisher, on the grounds that while it might be relevant to quality control, i.e., to keeping the number of defective products within preassigned limits, but is quite irrelevant to the evaluation of a scientific hypothesis, which is essentially unique.

While Fisher's objection is perfectly correct in my opinion, I do not believe that it gets to the heart of the matter, which is that error prob-

abilities, however glossed in terms of frequencies or whatever, are the wrong probabilities to consult. I shall show this by means of a simple counterexample, and then discuss its lessons. The counterexample is of a well-known type. Consider a diagnostic test for a disease, with two outcomes ‘positive’ and ‘negative’, whose intended meaning is ‘has the disease’ and ‘does not have the disease’. The test is administered to randomly selected subjects in a population in which the incidence of the disease is very high, say 999 in 1000. Suppose the test has a false positive rate of 0, i.e., the chance of it erroneously registering the presence of the disease is 0, and a false negative rate of 0.5; i.e., the chance of it erroneously registering the absence of the disease is 0.5. If the null hypothesis is  $h_0$ : the subject does not have the disease, then the test has very small chances of committing both type 1 and type 2 errors, of 0 and .05 respectively. Yet it is easy to see that the chance of  $h_0$  being true when the test registers ‘negative’ is under 2% (.196 to be precise). In other words, the proportion of those testing disease-free who actually have the disease is almost 100%!

Thus a test with excellent error characteristics, i.e., very low type 1 and type 2 errors, has an extremely high chance of accepting a hypothesis when it is false. But how can that be? We are told in statistics textbooks of the NP persuasion that the chance of accepting a hypothesis when it is false is defined to be the chance of making a type 2 error. The solution to the apparent paradox is that the ordinary language functor ‘the probability of accepting a false hypothesis  $h$ ’ can be rendered as two quite different conditional probabilities: (i) the probability of  $h$  being false conditional on its being accepted, and (ii) the probability of  $h$  being accepted conditional on its being false. Only in exceptional circumstances are these conditional probabilities identical; indeed, we have the relation

$$(i) = (ii) \left( \frac{P(\neg h)}{(ii)P(\neg h) + (h \text{ is accepted}|h)P(h)} \right) \quad (3)$$

which is an easy rewriting of (2) above. (ii) is of course the chance of making a type 2 error, but what everyone really wants to know from the outcome of the test is (i), and (3) tells us that (ii) by itself does not provide this; indeed, (3) shows just how (ii) can be very small consistently with (i) being very large.

The message by (3) conveyed is radical: error probabilities of a test *are no guide to the correctness or reliability of the hypothesis tested*. Mayo rewords the NP criteria as the demand that tests which ‘reliably’ pass a hypothesis, or ‘indicate its correctness’, are just the severe ones, where severity is defined as having a very small type 2 error. But ‘cor-

rectness' is a (i)-quality, and (3) shows that it is definitely not reliably indicated by the smallness of (ii). Thus NP criteria are simply fallacious, or, in logicians' terminology, demonstrably unsound rules of (inductive) inference.

**5. Subjectivism.** It is not surprising, in view of the difficulty of distinguishing the conditional probabilities (i) and (ii) above in the informal expression of what a type 2 error is, that NP ideas have been so tenacious. It *is* surprising that the fallacious nature of NP inferences is taking so long to be recognized. Perhaps this is due to a reluctance to embrace probabilism, which explains precisely in equation (3) above why the NP-criteria do not suffice, with its apparent inability to deliver the desirable goods of inferences uncontaminated by the agent's personal prejudice. In this section, I shall try to mitigate this apparent disability by arguing that personal opinion is ineradicable.

That it is ineradicable is the corollary of Hume's sceptical arguments, arguments that nobody has yet been able to refute. And there is very good reason to suppose nobody ever will, since these arguments at bottom rest on a clear logical fact, the underdetermination of theory by observation. All our theoretical constructs, while we hope they are empirically based, are nevertheless not determined uniquely by experience. That is to say, they are not derivable from experience. This is hardly surprising: experience in its raw state not being propositional, derivability in any logical sense is clearly out of the question. Thus any inference to the correctness of theory, or to its reliability, is necessarily uncertain. You can put figures on the uncertainty if you want, but as Hume noted these too will be uncertain. In other words, subjectivism is ineliminable. Q.E.D.

Since it is ineliminable, an honest theory of uncertain inference will display the uncertainty explicitly. Probabilism does so, in the form of its exogenously determined priors. But nagging doubts are hard to quell. Surely (I have heard people say this) no adequate theory can make it anything other than rational to believe that Pan Am 103 (the flight that terminated in the Lockerbie disaster of 1987) was downed by a bomb, given the overwhelming evidence to that effect. Yet a clever—or perhaps a stupid—Bayesian can consistently, and hence in terms of that theory rationally, doubt it. Therefore the Bayesian theory is not an adequate theory. What can be said to this?

The answer the Bayesian should give is the same as the answer that should be given by someone who had never heard of Bayesianism, and it goes like this. The limits of belief and behavior beyond which we classify people as irrational are difficult to define. I think we would be strongly tempted to view as irrational someone who said, as Achilles

said to the tortoise in Lewis Carroll's fable, that they believed a proposition  $a$ , and they believed 'if  $a$  then  $b$ ', but they did not see any reason to believe  $b$ . But the Bayesian theory views them as irrational as well. Where there seems to be sound reasoning from bizarre premises is a different matter, however, and I myself do not believe that we would invariably cast this type of pathology as irrationality; 'mad' is the more likely epithet.

One (fortunately) does not have to appeal to madness to make the point. There are certainly non-fanatics who still refuse to allow that the facts say what most people think they say. Einstein famously, or perhaps notoriously, was unwilling to allow that God played dice with the world, in the teeth of evidence that, via a physical theory, strongly suggested otherwise. Was he irrational? It is not clear. It is a commonplace in current philosophy of science that evidence never bears an unambiguous interpretation; in other words, it is never unambiguous what exactly evidence is evidence for (or against). This is of course just the underdetermination thesis, also known now as Duhem's problem. And we have seen that an honest epistemological theory must concede an ineliminable uncertainty to the truth of the theoretical, explanatory constructs we erect on the basis of experience. This infects the diagnosis of Pan Am 103 just it does the evaluation of quantum chromodynamics. The response the Bayesian should give to the charge of not being able to prove irrational any but the bomb hypothesis is that they are indeed guilty as charged, but that since there *must be* room for doubt there *is* room for *reasonable* doubt, in that and any other contingent hypothesis.

That is not by any means all that can be said on Duhem's problem, as a succession of Bayesian analyses have shown: Hellman, etc. Generally speaking, the response of non-Bayesians to these is that they depend on a suitable distribution of prior probabilities (for example, Mayo 1996). But the answer to this charge is to separate it into its descriptive and normative components. Normatively speaking, it is true that the priors are not constrained by the theory, but we have seen that this is far from the telling objection it is often thought to be. As to the descriptive component, the response is that, as far as historical analyses are concerned, the priors are an attempt to represent the actual beliefs of the people involved.

It is sometimes claimed by Bayesians that objectivity is not sacrificed in having undetermined priors, because the posterior distribution  $P(h_i|e_1, \dots, e_n)$  typically becomes independent of the prior distribution in the limit. Where the posterior distribution is obtained from a prior density for a continuous parameter, that nowhere vanishes, and a likelihood function based on an independent identically distributed



sample this is indeed the case, since the the posterior distribution tends to the normal with the sample mean and standard deviation. The assumption of independence with common distribution is dispensed with in more general, 'with probability one' results. For example, if two prior distributions agree on their probability zero events, then with probability one the then the supremum distance between the posterior distributions, regarded as random variables depending on the sample, tends to zero (Blackwell and Dubins 1962). Another result is that if  $I_h$  is the indicator function of any proposition defined in the sample space of the outcomes of repeated observations, then with probability one the posterior probability of  $h$  tends to  $I_h$ , i.e. to one if  $h$  is true and 0 if not (Halmos 1950, 213, Theorem B). A similar result is proved in an explicitly logical setting by Gaifman and Snir (1980; for a longer discussion see Earman 1992 and Hellman, this volume).

These results should be interpreted with caution. Even were they not 'with probability one' results, they still talk about limiting distributions, and do not therefore say what will or will not happen in any finite time. And in addition as probability one results they say nothing about what will definitely happen now or at infinity, merely what you are a priori certain will happen, a quite different thing. In other words these 'merger of opinion' theorems, though remarkable in themselves, are of doubtful efficacy in restoring objectivity to the Bayesian enterprise. But to assume that the theory is in deficit on the objectively grounded true methodological judgments that can be made begs the question that it is possible to make objectively correct judgments where that theory fails to do so. This is a claim that is unsubstantiated and which there seems reason to doubt. We have seen that the claim of superiority on this score made by a rival account, Neyman-Pearson theory, is untenable: that theory's apparently greater power generates unsound fallacious inferences.

That there is an objectively determinate yes/no answer to every sensible theoretical question assumes attainable a degree of completeness which the limitative results of Gödel, Church, Tarski, Cohen and others in this century have shown to be not to be the case. There is no more reason to assume it true in the domain of methodology. Indeed, it is entirely possible that the Bayesian theory represents the limit of completeness, and that any increase in theoretical power will generate unsoundness. It is of course impossible to prove this, but just as we feel we have evidence for the correctness of Church's Thesis (that every computable number-theoretic partial function is partial recursive) falling short of an in-principle unattainable formal proof, so I believe that we have evidence for the correctness of this claim in the cluster of results that Lindley regards as showing 'the inevitability of probability'

(1982), together with the failure of alternative methodological accounts to deliver sound inferences where the Bayesian theory cannot.

**6. Updating Rules.** Up to the second decade of the twentieth century a posterior probability, intended to represent your new judgment about the probability of a hypothesis  $h$  on receipt of data  $e$ , was expressed in what is now called conditional probability form:  $P(h|e)$ . Kolmogorov's celebrated monograph (1950), and the work of Ramsey and de Finetti, conceptually unlinked the posterior probability conceptually from the conditional probability. Now the conditional probability  $P(h|e)$  is taken to be just as much a prior probability as the unconditional probability  $P(h)$  of  $h$ . Kolmogorov had simply defined  $P(h|e)$  as  $P(e|h)/P(e)$  where  $P(e) > 0$ . In the Dutch Book Argument for the probability axioms,  $P(h|e)$  and  $P(h)$  are both regarded as betting quotients for bets examined at the same time  $T$ .  $P(h)$  is just your fair betting quotient at  $T$ , and  $P(h|e)$  is your betting quotient at  $T$  in a bet on  $h$  that will go ahead if and only if  $e$  is true; such a bet is called a conditional bet on  $h$ .

So  $P(h|e)$  is your conditional fair betting quotient, part of your total system of fair betting quotients, conditional and otherwise, at  $T$ . So what should you do if you do learn  $e$  to be true? As Ian Hacking noted in an influential paper (1967), there is nothing in the contemporary Bayesian theory that tells you, though there certainly seemed an unwritten rule: If you learn  $e$  and no more, then your updated (posterior) probability function  $Q(\cdot)$  should be equal to  $P(\cdot|e)$ . What happened is that the rule became explicit and, under the name of *Bayesian Conditionalization*, added by most commentators to the *synchronous* probability axioms as a *diachronic*, or dynamic, rule. The synchronous axioms and the dynamic rule of conditionalization for most people constitute the modern Bayesian theory of inference.

The dynamic rule is clearly not a consequence of the probability axioms. On the other hand, if it is part of the core Bayesian theory, and the theory is a theory of coherent belief, then there presumably can be some coherence argument for the rule. In 1973 Paul Teller produced a Dutch Book argument for conditionalization, acknowledging that the idea for it had come from David Lewis. The idea was this. The way you respond to new factual information  $e$  should not be a haphazard affair, but controlled by some rule. Then anyone who knows your updating rule can, if you are prepared to bet at all your fair betting rates, force you to lose money if that rule is not conditionalization. The proof is simple. Let your fair betting quotient in a conditional bet on  $h$  given  $e$  be  $p$ , and on  $\bar{e}$  be  $r$ . Suppose your updating rule sets  $Q(h)$ , your updated probability of  $h$  after learning  $e$ , equal to  $r \neq p$ ; suppose in fact that  $r < p$ . Now suppose the following bets are made: (i) a

conditional bet on  $h$  given  $e$  with betting quotient  $p$  and stake 1; (ii) a bet on  $e$  at betting quotient  $q$  and stake  $p-r$ ; and (iii) if  $e$  is true, a bet against  $h$  at betting quotient  $r$  with stake 1. The net gain from these bets is the negative quantity  $q(r-p)$ ; i.e., whoever made these bets would lose whatever the truth-values of  $h$  and  $e$ , and so have a Dutch Book made against them.

**7. Jeffrey Conditionalization.** Thus we seem to have a coherence argument for the rule of conditionalization no different, *mutatis mutandis*, from those for the probability axioms themselves. It also turned out that the scope of coherence arguments was not exhausted by this result. Richard Jeffrey had considered the possibility of a shift in your belief function, not as here as the result of learning the truth of a proposition, but as the result of some sensory experience. In Jeffrey's example, you inspect a piece of material in poor light, and your initial probability  $p$  that it is green, say, before the inspection is thereby changed to some new value  $q$  (1983).

This example generalizes what happens in Bayesian conditionalization because any values at all are allowed for  $p$  and  $q$  apart from  $p = 0$ , whereas in Bayesian conditionalization the new probability  $Q(e)$  of the conditioning proposition  $e$  is 1 (since  $Q(e) = P(e|e) = 1$ ). How should we update the remainder of our probability function in the light of this exogenous shift? It is logically possible, of course, that there is simply no determinate answer to this question, but Jeffrey suggested the following rule, which is now called variously *Jeffrey conditionalization*, *probability kinematics*, or just *Jeffrey's rule*. I shall call it the last of these names. In the example it looks like a single proposition that has its probability shifted initially, the proposition  $g$  that the material is green. But if the agent is a coherent Bayesian, there will in fact be two propositions,  $g$  and  $\neg g$ , which have their probabilities shifted initially to  $q$ ,  $1-q$ . The pair  $\{g, \neg g\}$  is a partition, and Jeffrey's rule at its most general is stated for an arbitrary discrete partition  $\{e_1, e_2, \dots\}$  whose probabilities, initially all nonzero, shift to  $q_1, q_2, \dots$  as follows:

$$Q(\cdot) = \sum q_i P(\cdot | e_i),$$

where  $Q$  is the updated function. Where the partition is just the pair  $g, \neg g$  Jeffrey's rule clearly reduces to

$$Q(\cdot) = qP(\cdot | g) + (1 - q)P(\cdot | \neg g). \quad (4)$$

Jeffrey's rule has some interesting properties. For simplicity I shall state them for the pair case (4), but they are immediately generalizable. First, given that  $Q(g) = q$ , (4) is equivalent to the pair of identities

$$P(\cdot|g) = Q(\cdot|g), P(\cdot|\neg g) = Q(\cdot|\neg g). \quad (5)$$

Second, as  $q$  tends to 1, (2) tends continuously to

$$Q(\cdot) = P(\cdot|g)$$

i.e., to Bayesian conditionalization. Call a Jeffrey shift non-extreme if  $q < 1$ . Then, thirdly, if you change your mind about a shift that is non-extreme and wish to return to the status quo ante, you can reverse the updating (4) by another application of (4) to recover your original probability function. You can't do this with Bayesian conditionalization: once you shift to a new probability of 1 on any proposition your original probability function is not reversible by a further Bayesian conditionalization. Finally, again unlike Bayesian conditionalization, the order in which you update on sequential shifts may give distinct final updated functions.

What is important for the present discussion is that there is a Dutch Book argument for Jeffrey's rule. Armendt (1980) shows that anyone who has an updating rule which violates the identities (5) can have a Dutch Book made against them by someone who knows their rule. So Jeffrey's conjecture that his rule is also a sound one, in terms of the coherence criterion, seems to be correct. The following example shows that all is not as well as it seems, for Bayesian conditionalization and hence, because that is a special case of Jeffrey conditionalization, for Jeffrey's rule itself. Suppose someone knows that a drug will be administered to them which has a nonzero chance of making them uncertain about things they were, and currently are, certain about. Let  $a$  be some such factual proposition, about your identity, for example. So  $P(a) = 1$ . Let  $Q$  be a random variable whose possible values are your degrees of belief in  $a$  after taking the drug. You assign a non-vanishing chance to your coming to doubt  $a$ . Let  $b$  be this proposition. So  $P(b) > 0$ . It follows by coherence that  $P(a|b) = 1$ . Now suppose you do in fact, through introspection, learn the truth of  $b$  after taking the drug. But you cannot (Bayesian) conditionalize on  $b$  on learning its truth in this way, for if  $b$  is true it means that your new probability  $Q(a)$  of  $a$  is less than 1.

**8. Why Dynamic Dutch Book Arguments Fail to Provide a Standard of Consistency.** In other words, you cannot consistently conditionalize on  $b$  if you learn its truth directly. Long ago, Ramsey (1931) had pointed out that learning the truth of the conditioning proposition might well act as a shock forcing you, 'for psychological reasons', to violate Bayesian conditionalization. As we have just seen, there may also be purely logical reasons for violating it. This leaves the status of dynamic Dutch

Book arguments, allegedly showing that Bayesian and Jeffrey conditionalization are generally valid coherence constraints, unclear. To go further we must make a clear separation between coherence and consistency. In the discussion of the Dutch Book argument for the probability axioms these concepts were in effect identified, because they seemed to be equivalent. Indeed, for synchronic degrees of belief I believe they are equivalent. But not for diachronic. It is easy to see this. First, note that a synchronically coherent assignment can be dynamically incoherent. We have a ready-made example. Let  $a$  and  $b$  be as above, except that we shall assume that  $b$  now assigns a precise probability  $r$  less than 1 to  $a$ . By synchronic coherence,  $P(a|b) = 1$ . In the Lewis-Teller dynamic Dutch Book argument above let  $h$  be  $a$  and  $e$  be  $b$ ;  $p$  is now 1. On learning  $b$  your probability for  $a$  is now  $r$ . Suppose the same three bets are made. Then the bettor loses come what may. So the dynamic Dutch Book for conditionalization now shows that the assignment  $P(a|b) = 1$  is dynamically incoherent.

If there are good reasons to equate synchronic coherence with the consistency of the corresponding synchronic degrees of belief, then the simple counterexample above shows that dynamic incoherence cannot be equated inconsistency. The conclusion seems inescapable that dynamic incoherence indicates nothing of any logical or epistemological interest. There is also a quite independent and rather simple argument to the same conclusion. I am inconsistent if I accept both  $A$  and  $\neg A$  as true simultaneously. I am not inconsistent if I accept  $A$  today and  $\neg A$  tomorrow. I have merely changed my mind. My 'rule' for updating my probability for  $a$  on learning  $b$  must, if I appreciate the meanings of the words involved, give  $a$  a probability less than 1. I am not inconsistent in planning, in the appropriate circumstances, to entertain a degree of belief in  $a$  different from that which I have today; I merely know that those circumstances are a change of mind. It is quite beside the point that from my present perspective the change of mind is irrational. It is the entirely rational claim that I may be induced to act irrationally that the dynamic Dutch book argument, absurdly, would condemn as incoherent.

As a coda to this section I shall mention a discussion a decade ago of something called the Principle of Reflection. Suppose that  $C$  is a random variable whose values are your possible degrees of belief in some proposition  $c$  tomorrow, and that  $P$  is your current probability function. The equation  $P(c|C = x) = x$ , for all  $c$  in the domain of  $P$  and all  $x$  in  $[0,1]$ , has been called by van Fraassen the Principle of Reflection (1984). He argued its validity because it has, as we have seen, a dynamic Dutch Book argument for it. But as we now know, that is not a good reason.

**9. Where Does This Leave the Updating Rules?** Where indeed? I shall argue that the possibly surprising thesis that the Bayesian theory has no such rules, except in a derivative form in which they are essentially redundant. I shall start with Bayesian conditionalization. Suppose you make some observation as a result of which you become convinced of the truth of a proposition  $e$ ; i.e., your probability of  $e$  shifts exogenously from  $p < 1$  to 1. This determines the value of your new probability function on one point, namely  $e$ . So  $Q(e) = 1$ . Suppose also that learning  $e$  did nothing to affect the values of your conditional probabilities, on  $e$ , in every other proposition in the domain of  $P$ . This supplies further, conditional, values for  $Q$ ; namely,  $Q(.|e) = P(.|e)$ . But now  $Q$  is fully determined, if you are synchronically coherent, since by the probability calculus  $Q(.) = Q(.|e)$  [since  $Q(e) = 1$ ] =  $P(.|e)$ . Not only is  $Q$  fully determined, but it is determined to obey Bayesian conditionalization! But this is not a proof of an unconditional rule of Bayesian conditionalization, since it clearly depends on the identities  $P(.|e) = Q(.|e)$ . In the counterexample exhibited earlier, with the propositions  $a$  and  $b$ , these identities are violated, since  $P(a|b) = 1$  and  $Q(a|b) = Q(a) < 1$ .

Now recall from another earlier discussion that Jeffrey's rule (3), given  $Q(e) = q$ , is equivalent to all the pairs of identities  $P(.|e) = Q(.|e)$  and  $P(.|\neg e) = Q(.|\neg e)$ , which reduce to  $P(.|e) = Q(.|e)$  when  $q$  is 1. In other words, it is a theorem of the probability calculus that Bayesian and Jeffrey conditionalization are equivalent to essentially the same set of conditional probabilities remaining unchanged by the shift on  $e$ . And, given that there are counterexamples to Bayesian conditionalization (and hence to Jeffrey's rule too, since it subsumes Bayesian conditionalization as a special case), and these clearly entail a change in just those conditional probabilities, the conclusion seems very plausible *that Bayesian and Jeffrey conditionalization are derivative rules of the synchronic probability calculus, valid just in case the relevant conditional probabilities remain unchanged by the shift*. The situation is quite analogous to that with modus ponens, as Hellman has remarked (this volume): one is entitled to detach  $d$  from  $c$  and 'if  $c$  then  $d$ ' if and only if you accept both premises simultaneously; and you would not be so entitled had the learning of  $c$  caused you to reject the conditional 'if  $c$  then  $d$ '.

I myself think that this conclusion is correct. It is true that there have been attempts to derive these rules from other general principles, like a minimum information principle (Williams 1980), symmetry principles (van Fraassen 1989), and so on. These derivations are of considerable independent logical interest, but they hardly justify the updating rules themselves, if the foregoing arguments are correct, because these

rules possess no general validity, and to the extent that they are conditionally valid this is established by means of the probability axioms themselves.

And where does this conclusion leave the Bayesian theory? I have endeavored to describe the theory in such a way that it does not depend on any updating rule. This does not immediately cast doubt, or more than there was before, on the validity of such classical results as the convergence of opinion theorems, since these are framed in terms of your prior probability that your prior conditional probability will exhibit suitable convergence behavior. I claim that nothing valuable is lost by abandoning updating rules. The idea that the only updating policy sanctioned by the Bayesian theory is updating by conditionalization was untenable even on its own terms, since the learning of each conditioning proposition could not itself have been by conditionalization.

**10. Conclusion.** The Bayesian theory has been around a long time, longer than any other theory of inductive inference still with us. It provides a sounder basis for inductive inference than any developed from its principal rival, classical refutationist statistics, as Section 4 above demonstrates. The latter's vaunted objectivism is seen to be not a strength at all, but a generator of unsound inferences. Why it is taking the statistics community so long to recognize the essentially fallacious nature of NP logic is difficult to say, but I am reasonably confident in predicting that it will not last much longer. Indeed, the tide already seems strongly on the turn.

## REFERENCES

- Armendt, B. (1980), "Is there a Dutch Book argument for probability kinematics?", *Philosophy of Science* 47:583–589.
- Bayes, T. (1763), "An essay towards solving a problem in the doctrine of chances", *Philosophical Transactions of the Royal Society* 53:370–418.
- Blackwell, and L. Dubins (1962), "Merging of opinions with increasing information", *Annals of Mathematical Statistics* 33:882–887.
- Dawid, A. P., M. Stone, and J. V. Zidek (1973), "Marginalization paradoxes in Bayesian and structural inference", *Journal of the Royal Statistical Society B*: 189–223.
- Earman, J. (1992), *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.
- Fisher, R. A. (1947) *The design of experiments*, 4th edition. Edinburgh: Oliver and Boyd.
- Gaifman, H. and M. Snir (1980), "Probabilities over rich languages, testing and randomness", *Journal of Symbolic Logic* 47:495–548.
- Gärdenfors, P. and N-E Sahlin (1988), *Decision, probability and utility*. Cambridge: Cambridge University Press.
- Giere, R. (1984), *Understanding scientific reasoning*, 2nd edition. New York: Holt, Rinehart and Winston.
- Hacking, I. (1967), "Slightly more realistic personal probability", *Philosophy of Science* 34, 311–325.

- Halmos, P. (1950), *Measure theory*. New York: van Nostrand-Reinhold.
- Hellman, G. (1997), "Bayes and beyond", this volume.
- Howson, C. and P. Urbach (1993), *Scientific reasoning: the Bayesian approach*, 2nd edition. Chicago: Open Court.
- Jaynes, E. T. (1968), "Prior probabilities", *IEEE Transactions on Systems Science and Cybernetics* SSC-4: 227–241.
- . (1973), "The well-posed problem", *Foundations of Physics* 3:477–493.
- Jeffrey, R. C. (1983), *The logic of decision*. Chicago: University of Chicago Press.
- Jeffreys, H. (1961), *Theory of Probability*, 3rd edition. Oxford: Clarendon Press.
- Kolmogorov, A. N. (1950), *Foundations of the theory of probability* (translation for the German of 1933 by N. Morrison). New York: Chelsea.
- Laplace, P. S. de (1951), *Philosophical essay on probabilities*. New York: Dover. (English translation of *Essai philosophique sur les probabilités*, 1820).
- Lindley, D. (1982), "Scoring rules and the inevitability of probability", *International Statistical Review* 50: 1–26.
- Maher, P. (1997), "Depragmatized Dutch Book Arguments", this volume.
- Mayo, D. (1996), *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.
- Milne, P. M. (1983), "A note on scale invariance", *British Journal for the Philosophy of Science* 34: 49–55.
- Neyman, J. (1952), *Lectures and conferences on mathematical statistics*, 2nd edition. Washington, D.C.: Graduate School of U.S. Dept. of Agriculture.
- Ramsey, F. P. (1931), "Truth and probability", in *The foundations of mathematics and other logical essays*. London: Routledge and Kegan Paul.
- Savage, L. J. (1954), *The foundations of statistics*. New York: John Wiley.
- Teller, P. (1973), "Conditionalisation and observation", *Synthese* 26:218–258.
- van Fraassen, B. C. (1984), "Belief and the will", *Journal of Philosophy* LXXXI: 235–256.
- . (1989), *Laws and symmetry*. Oxford: Oxford University Press.
- Walley, P. (1991), *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.
- Williams, P. M. (1980), "Bayesian conditionalisation and the principle of minimum information", *British Journal for the Philosophy of Science* 31: 131–144.