36

Heterodox Probability Theory

PETER FORREST

The purpose of this chapter is to survey and assess the ways of departing from the Bayesian orthodoxy about probabilities as they apply to reasoning. Most of these departures are no longer fashionable, but they deserve reconsideration. For a more historical survey readers are referred to Hailperin (1990).

Perhaps the least controversial departure from orthodoxy is the way of adapting the standard theory to a non-classical logic. Next there is the Confirmation Theory program based upon Carnap's idea of a logical probability. This relies heavily on the standard Calculus of Probabilities but goes beyond Bayesianism. After looking at a more modest program concerning Proportional Syllogisms, Kyburg's continuation of Carnap's program will be considered. Kyburg, unlike Carnap, rejects the assumption of precise numerical probabilities in favor of ones which are fuzzy. Another approach that likewise rejects precise numerical probabilities is Levi's theory wherein probabilities are not so much fuzzy as indeterminate.

Again, there is a tradition going back to de Finetti of stating some axioms governing comparative probabilities, based upon the primitive relations: *p is more probable than q*; and *p and q are equiprobable*. The aim of this qualitative approach is to show that these comparisons may be faithfully represented by precise numerical probabilities.

Perhaps the central feature of the Bayesianism orthodoxy is the rule of Conditionalization, namely that the probability after discovering new evidence should equal the prior conditional probability on the supposition of that evidence. There are alternatives, notably Imaging.

Finally, there is the impact of quantum theory, including perhaps the most deviant theory of all, negative probabilities.

1 The Bayesian Orthodoxy

The Bayesian orthodoxy consists of four theses: (1) that probabilities are precise numerical representations of subjective degrees of confidence; (2) that the standard rules hold for the synchronic coherence of a system of subjective probabilities; (3) that the rule of Conditionalization holds for diachronic coherence; and (4) that the systematic constraints on the rationality of inferences are precisely the requirements of synchronic and diachronic coherence.[1]

The Dutch Book Argument for Bayesianism is based upon the assumption that subjective degrees of confidence have, for their behavioral manifestation, a willingness to take risks, which may be idealized as a public commitment to placing bets on the truth or falsity of any proposition. The probability is taken to equal the *betting quotient*, that is if the probability is p, the odds are in the ratio $p:1-p$. Your system of probabilities is then said to be *coherent* if and only if someone who knows the odds you are committed to cannot make a Dutch book, that is, cannot place a system of bets with you which is guaranteed not to make you anything and might result in your losing. As a consequence it is inferred that numerical probabilities may be represented to be real numbers in the interval 0 to 1 with the usual principles holding: If p entails q then $\text{Prob}(p) \leq \text{Prob}(q)$; $\text{Prob}(p\&\neg p) = 0$; $\text{Prob}(p) + \text{Prob}(q) = \text{Prob}(p \vee q) + \text{Prob}(p\&q)$; and $\text{Prob}(p|q) \times \text{Prob}(q) = \text{Prob}(p\&q)$. Here $\text{Prob}(p|q)$ is the conditional probability of p on the supposition that q, which is taken to be manifested by the preparedness to take a risk on p supposing q, itself idealized as the commitment to placing conditional bets on the truth or falsity of p, where in the event of $\neg q$ the bets are canceled. (See Kemeny (1955) for the Dutch Book Argument for these rules.)

The rule of Conditionalization may be stated thus: $\text{Prob}(p|q)$ is unchanged by the discovery (with certainty) that q.[2] The Dutch Book Argument for this, expounded by Teller (1976) but originally due to Lewis, depends on the rather strong assumption that if Conditionalization does not hold then before you discover that q you will have a public commitment to stating odds for non-conditional bets to be made when and if you discover that q. Hence this diachronic Dutch Book Argument seems less persuasive than the synchronic ones, which require only the commitment to bets here and now. There are, however, other arguments for Conditionalization (see Teller 1976).

The final Bayesian thesis is obviously defeasible but the Bayesian orthodoxy is that attempts, like Carnap's, to impose further systematic constraints have failed.[3]

2 Idealization

All systematic theories of probable reasoning are highly idealized. That is, various simplifications are assumed even though they are known to be false. The Bayesian assumption that betting should reflect degrees of confidence is one such idealization. Consider the bookies in the Dutch book argument who are out to make money from you no matter what eventuates. Some of their bets may seem to them much too generous when taken in isolation, but they make them out of a desire to profit come what may, which is perfectly reasonable. Hence Dutch bookies themselves provide examples of perfectly reasonable people whose degrees of confidence are not manifested by their betting behavior. None the less we might well accept the basic Bayesian idea of degrees of belief being represented by precise betting quotients as an idealization useful for modeling certain features of the belief system of a reasonable person.

Idealization can be avoided if we want to, but we usually do not. For instance, typically, the underlying language will be closed under various operations such as negation, conjunction, and disjunction and so contain propositions of arbitrary length. Now we can easily put some restriction on the complexity of propositions considered. Moreover, Bayesianism is usually thought of as having the consequence that everyone should be certain of all mathematical theorems even those they do not understand. But if we put a restriction on the language being considered and if we weaken the principle that if p entails q $\text{Prob}(p) \leq \text{Prob}(q)$ so it holds only for various stated rules of inference then if p is a theorem whose proof using these rules cannot be stated in the restricted language we no longer require $\text{Prob}(q) = 1$ even if all the axioms from which it is proved are certain. Such modifications to the system are, however, tedious and so for most purposes may be ignored.

Another idealization worth noting is that we are not merely concerned with ideally rational changes to a system of subjective probabilities, but that we are assuming the person concerned is ideally rational at all times. Often, however, we change our mind because we recognize that our previous state was not (ideally) rational. So for instance new evidence might force us to take seriously a hypothesis we should have assigned a non-negligible probability to, but had dismissed.

Yet again for some purposes we shall need to consider not just the ordinary real numbers but nonstandard ones, obtained by adjoining infinitesimals. For instance, suppose we have a continuum of hypotheses depending on a parameter λ which could be any positive real number, but is unlikely to be either very small or very large. (This is an actual example based upon Carnap's Confirmation Theory. See section 5.) Then we might judge that it is as probable that $\lambda < 1$ as that $\lambda > 1$ but that it is more probable that $\lambda \geq 1$ than that $\lambda > 1$, and likewise more probable that $\lambda \leq 1$ than that $\lambda < 1$. In that case we seem forced to assign not zero but infinitesimal value to the probability that $\lambda = 1$. This is analogous to saying that the open interval of real numbers $(0,1) = \{x: 0 < x < 1\}$ is infinitesimally smaller in size than the closed interval $[0,1] = \{x: 0 \leq x \leq 1\}$. Perhaps the most natural theory of infinitesimals in this context is geometric measure theory (Schanuel 1982). Another common idealization, then, is that we ignore infinitesimals.

Bayesianism can and has been queried, even given its various idealizations (see section 7). None the less the consequences of first three theses cohere well with our intuitions and all the heterodox positions defended by me in this article either extend it (thus violating the fourth thesis) or are less idealized.

3 Two Approaches to a Theory of Probability

The chief topic of this article is probability theory as it applies to reasoning. Here there are two different approaches. The dominant strategy, illustrated by the Bayesian orthodoxy, is to think of probability theory as putting constraints on subjective probabilities and on how they change with time, where a subjective probability, written Prob, is a measure of just how confidently a proposition is asserted (if the probability is over 0.5) or denied (if it is less than 0.5). The constraints are usually considered to be normative and to be necessary conditions for rationality. The other approach, championed by Carnap (1950) but going back to Keynes (1921) and ultimately to Johnson (1921–4), is to think of an inference as having a probability which is 1 just in case the inference is deductively valid and 0 just in case the conclusion is inconsistent with the premisses. This is called a logical probability and will be written $\mathrm{Prob}_{\mathrm{Log}}$. It is thought of as providing a degree of logical confirmation. In that case the conditional probabilities are more fundamental than the absolute ones which are defined thus: $\mathrm{Prob}_{\mathrm{Log}}(p) = \mathrm{Prob}_{\mathrm{Log}}(p|t)$ where t is any tautology.

We should anticipate a connection between logical and subjective probabilities. Consider the idealized situation in which all evidence is certain. Then someone – Carnap's 'Logically Omniscient Jones' – whose subjective probability equalled the logical probability on the evidence should not turn out to be irrational. That poses a problem if the probabilities are real numbers in the interval 0 to 1 inclusive. For then the two approaches differ in that for logical probabilities we have $\mathrm{Prob}_{\mathrm{Log}}(p|e) = 1$ only if p is entailed by e, whereas there is nothing irrational about someone who has evidence e being certain of some contingent truth p (e.g. that the whole universe did not come into existence 47,842 years ago) which is not strictly entailed by e and hence having $\mathrm{Prob}(p|e) = 1$. However, harmony can be restored if we allow the logical probabilities to take nonstandard values which include infinitesimals. In that case we could say that $\mathrm{Prob}_{\mathrm{log}}(p|e)$ differs from 1 by an infinitesimal and that the corresponding subjective probability equals the logical probability *modulo* infinitesimals.

4 Adjustment for Nonclassical Logics

To illustrate the adjustments required if we reject the classical sentential calculus, suppose we are considering the subjective probabilities. Then the standard calculus of probabilities contains either as axioms or derived theorems:

$$\mathrm{Prob}(p\&\neg p) = 0; \quad \mathrm{Prob}\ (p \vee \neg p) = 1; \quad \text{and } \mathrm{Prob}(p) + \mathrm{Prob}(\neg p) = 1.$$

These rules do not presuppose bivalence but they do presuppose the excluded middle and non-contradiction. Thus suppose p is the Liar and, as dialethic logicians hold all four of p, $\neg p$, $p \vee \neg p$ and $p\&\neg p$ are logical truths. Then we have $\mathrm{Prob}(p\&\neg p) = 1 \neq 0$. and $\mathrm{Prob}(p) + \mathrm{Prob}(\neg p) = 2 \neq 1$. Moreover, the standard calculus of probabilities implies a probabilistic version of the supposedly counter-intuitive rule of disjunctive syllogism, rejecting which is one of the motivations for relevance logic even when dialethic logic is not embraced. Thus we find that $\mathrm{Prob}(p) \geq \mathrm{Prob}(p \vee q) + \mathrm{Prob}(\neg q) - 1$. So for instance if $p \vee q$ is asserted with 99 percent confidence, and $\neg q$ asserted with confidence, it would be irrational to assert p with a confidence of less than 98 percent.

To accommodate heterodox logics we should adjust the calculus of probabilities. Instead of requiring merely that probabilities take values in the interval 0 to 1, we require in addition that the least upper bound of all probabilities is 1 and the greatest lower bound is 0. We have the addition rule: $\mathrm{Prob}(p \vee q) + \mathrm{Prob}(p\&q) = \mathrm{Prob}(p) + \mathrm{Prob}(q)$, and the usual multiplication rule: $\mathrm{Prob}(p|q), \mathrm{Prob}(q) = \mathrm{Prob}(p\&q)$. Moreover if p entails q then $\mathrm{Prob}(p) \leq \mathrm{Prob}(q)$. Given classical sentential calculus we then recover the standard calculus of probabilities.

While these adjustments are fairly obvious there is an important feature of any system in which it can happen that Prob(p) + Prob(¬p) differs significantly from 1. For in such a system there is a difference between confidently denying p, which corresponds to Prob(p) being near 0, and asserting ¬p which corresponds to Prob(¬p) being near 1.

5 Carnap's Confirmation Theory

Confirmation theory is based upon a rather strong version of Foundationalism, according to which given the total evidence any proposition p should have a unique probability assigned to it, namely the logical probability of the inference with the evidence as premises and p as conclusion. Here we are to idealize the situation by ignoring the very real possibility of evidence which is itself merely probable. In addition it is assumed that the evidence is consistent. As mentioned above we would require these probabilities to conform to the constraints on the subjective probabilities of someone who has that evidence and no other evidence, but the latter will typically underdetermine the probabilities, as in orthodox Bayesianism. Hence Carnap's theory of logical probability committed him to a stronger theory than that provided by the Standard Calculus of Probability. Moreover, it is an attractive idea that the structure of propositions, explicated by the calculus of predicates, should interact with probability theory. Hence he embarked upon a research program to discover the correct assignment of $\mathrm{Prob}_{\mathrm{Log}}(p)$ for every p in the calculus of predicates with suitable constraints on the interpretation of the names and predicates. For then, given any inference with consistent premises, we may take p as the conclusion and q as the conjunction of the premisses, in which case $\mathrm{Prob}_{\mathrm{Log}}(p|q) = \mathrm{Prob}_{\mathrm{Log}}(p\&q)/\mathrm{Prob}_{\mathrm{Log}}(q)$ is the logical probability of the inference in question. Carnap had several attempts at providing a satisfactory confirmation theory. All are based upon the extremely plausible principle of symmetry, namely that since the names are assumed to lack content beyond the fact of their naming particulars the logical probabilities must be invariant under permutation of names. Presumably that is an idealization of the actual situation regarding the referring expressions occurring in natural languages, but given the idealization the symmetry principle should be uncontroversial. Carnap's method was to seek the simplest confirmation theory which met various intuitive constraints, such as that we can learn by ordinary induction. His first choice (the c*-function of the appendix to Carnap (1950)) was unique but on further consideration he came up with a continuum of confirmation theories, depending on a parameter λ which was not fixed by intuition (Carnap 1952).

Perhaps because of Carnap's penchant for technical exposition this continuum of confirmation theories is not widely studied. This is a shame for the intuitive ideas are both simple and appealing. What Carnap does is treat the logical probability of p on q as having both an *a posteriori* and an *a priori* component. Suppose 10 Fs have been observed and 9 were Gs. Suppose also that the classification of Fs is into five possible kinds of equal status of which the Gs are one kind. We want to find the logical probability of an inference from those suppositions to the conclusion that b, some unobserved F, is a G. Then the *a posteriori* component is 0.9 and the *a priori* component is 0.2, and the probability we are seeking is somewhere in between 0.2 and 0.9. Where it is in the interval [0.2, 0.9] is determined by the weights assigned to the two components. Carnap took these weights to be the number of observed Fs, in this case 10 and the parameter λ, the weight of the a priori, so in this case the probability would be:

$$(10 \times 0.9 + \lambda \times 0.2)/(10 + \lambda) = (90 + 2\lambda)/(100 + 10\lambda).$$

Carnap's justification for the use of a linear weighting of the *a posteriori* and the *a priori* is an appeal to simplicity. Hence it should be treated as a hypothesis about probabilities which goes beyond our intuitions about them. If our aim is to find the best hypothesis then the appeal to simplicity is warranted. And as a hypothesis there is room for further empirical investigation of the most appropriate value for λ. For each value of λ we use the λ-confirmation theory to discover by induction the μ for which the μ-confirmation theory is most reliable. So for each λ there is a $\mu = f(\lambda)$, which is the value of the parameter implied by the original choice of the value λ. The only acceptable values of λ will be the fixed points, that is those for which $f(\lambda) = \lambda$. Following Carnap we should reject both very small and very large values for λ as neglecting either the *a priori* or the *a posteriori*. With good luck there might be jut one intermediate fixed point. And if that is very near an integer then simplicity would dictate that we round it off, obtaining the best hypothetical account of the unique logical probabilities. Perhaps we should have doubts as to whether such probabilities deserve to be called 'logical' but they would none the less provide a guide to reasoning.[4]

Carnap was criticized because his confirmation theory implied that induction does not justify universal generalizations such as 'All ravens are black' but only statistical generalizations such as 'At least 99.9 percent of ravens are black' and predictions such as 'The second raven I hear in 2005 will turn out to be black.' I urge readers to judge him to be right and his critics wrong, at least if we are ignoring such things as the purposes of an agent (human or divine) or laws of nature. None the less the Confirmation Theory research program was developed subsequently by Hintikka (1966) and more recently Zabell (1996), so as to arrive at a theory which allowed confirmation of universal generalizations. In Hintikka's theory there is a second parameter α, whose role may be illustrated by considering the simple case in which $\lambda = \infty$. Then if precisely n ravens have been observed, all of them black, the probability that all ravens are black is $1/(1 + 0.75^{1-\alpha})$. If n is 20 less than α, this probability is less than 0.4 percent. If n is 20 more than α then it is greater than 96.6 percent.

6 Proportional Syllogisms

Although Carnap's program was technically superb it suffered from the obvious defect that it was not applicable to anything other than a highly idealized language. Moreover his method was based upon the principle of selecting only the simplest out of a very many confirmation theories which were otherwise acceptable. But intuitively a slightly more complicated theory is only somewhat less probable than the simplest one. Hence even if there are precise logical probabilities even our best hypothesis about them will be too conjectural to command assent. This suggests two rather different ways of continuing something like the Carnapian program. One of these is to grant that we do not know enough completely to constrain subjective probabilities but to insist that we can go beyond the Bayesians by adopting what is intuitively the most secure part of Carnap's Program, namely that in the absence of either certain or probable evidence to the contrary any set of m Fs are as likely to contain precisely n Gs as any other set of m Fs. From this it follows that in many situations we may assign precise logical probabilities to proportional syllogisms such as: Precisely K percent of Fs are Gs. This is an F, so this is a G. In the appropriate circumstances the probability of the conclusion of that inference on the premises (together with background evidence) is K/100. And one of the defects of strict Bayesianism is that there can be quite coherent systems of subjective probabilities which capriciously assign higher or lower probability to some given F being a G.

Proportional syllogisms illustrate the difficulty of finding a systematic theory of probability. For unlike deductive logic the probability of an inference depends not just on the inference schema but on the choice of predicates to substitute for the schematic letters. For instance suppose we know that there are far more rabbits than bandicoots but are otherwise ignorant of bandicoots. Perhaps we know that all rabbits love carrots. Then we know the vast majority of rabbits-or-bandicoots love carrots and so, by a careless proportional syllogism, we might infer that very probably the first bandicoot we ever meet will love carrots. Obviously something has gone wrong, but it is not that you have relevant evidence about bandicoots. It is that 'rabbit-or-bandicoot' is the wrong sort of predicate. This well-known problem applies also to the Carnapian program if we attempt to apply it to natural languages. What it shows, I submit, is that any attempt at a theory which extends Bayesianism must also contain a theory of natural kinds and natural properties.

Granted that in suitable circumstances we know about the logical probabilities of proportional syllogisms we may then rely on the Williams–Stove justification of ordinary induction (Stove 1986). This is based upon the quite uncontroversial mathematical fact that the vast majority of large samples are, to a good approximation, representative of the population as a whole. This can be made quite precise as in Stove's example (Stove 1986: 67–71). Provided a fair proportion of the population have been observed, and the circumstances are appropriate for making the proportional syllogism, we may conclude, with high probability, that the observed members of the population are, to a good approximation, representative. Moreover having some special information about the sample, for example that it has all been observed prior to 2010 is, in most circumstances, intuitively irrelevant and so does not defeat the proportional syllogism. We may note that even if all the observed Fs have been Gs, the approximate nature of the representation prevents any conclusion stronger than that some very high percentage of Fs are Gs, which is in agreement with Carnap's confirmation theory.

Both Carnap's Confirmation Theory and the more general reliance on proportional syllogisms has been criticized for 'generating knowledge out of ignorance.' If knowledge is meant quite literally then this is not the case. For instance, knowing only that a coin has two sides and it is possible to toss a coin so that either heads or tails come up we might well be very confident of not getting a run of 20 heads. This would not count as knowledge even though actual experience, say with a biased coin, might, after very many tosses have resulted in about the same degree of confidence, and, if we are not being pedantic, count as knowledge. The difference lies in the sensitivity to further evidence. To count as knowledge even in a rather loose sense, a belief should not be too sensitive to further evidence. If, however, by 'knowledge' we just mean being almost certain then, far from seeing the generation of 'knowledge' out of ignorance as a defect, we should see this as a way of reconciling empiricism with the *a priori*.

7 Kyburg's Fuzzy Probabilities

Kyburg develops the Carnapian program by relying on proportional syllogisms.[5] That is, our knowledge of frequencies are taken as the sole determinant of the probabilities. (Ignoring for simplicity 'knowledge' of frequencies itself based upon probabilistic evidence.) Sometimes this results in precise numerical probabilities, but where there are rival proportional syllogisms they specify no precise probability but rather a (closed) interval. For instance, if we know that Tex is a Texan philosopher, that 30 percent of philosophers are vegetarians but only 10 percent of Texans are, then the resulting logical probability of Tex being a vegetarian is the closed interval $[0.1, 0.3]$ or 0.2 ± 0.1. (If $a \leq b$, the closed interval $[a,b] = \{a \leq x \leq b\}$.) We may think of these as fuzzy numbers, which may be added and multiplied, using the rules $[a,b] + [c,d] = [a + c, b + d]$, $[a,b] \times [c,d] = [a \times c, b \times d]$. Now the Addition Rule still holds, namely Prob(p&q) + Prob($p \vee q$) = Prob(p) + Prob(q), except that we are adding fuzzy numbers.

The most radical of Kyburg's theses is his rejection of the standard multiplication principle that Prob(p&q) = Prob(p|q), Prob(q), even in the special case in which the probabilities are precise. This is based upon the use of proportional syllogisms to specify probabilities. An example of Kyburg's (Harper 1982: 120) illustrates this nicely. Suppose in the whole population 50 percent are R (red) and 40 percent of Rs are also S (square). Now suppose a sample of 100 is taken out of this population. We are not told what the proportion of Rs is in the population but we are told that not 40 percent but 50 percent of the Rs in the sample are also Ss. Now consider one member of the sample. Because of Kyburg's reliance on proportional syllogisms he asserts that Prob(Sx $\vee$ Rx) = 0.5 and Prob(Rx) = 0.5, but Prob(Rx&Sx) = 0.2, not the 0.25 it should be according to the multiplication rule. Here I think Kyburg is mistaken. Intuitively our knowledge that in the sample 50 percent of Rs are Ss combined with what we already know about the whole population provides some information about the way in which the sample is unrepresentative. Quite how we use this information is not clear but it shows that the conditions are not appropriate for making the proportional syllogism. If so, Kyburg is mistaken about the consequences of his own reliance on proportional syllogisms, but the underlying theory of fuzzy logical probabilities is still tenable.

8 Levi's Indeterminate Systems

Kyburg's fuzzy probabilities are, as logical probabilities, assigned to inferences. As such they have their (fuzzy) values regardless of the probabilities of other propositions. By contrast Levi refines the theory of subjective probabilities by not requiring determinate numerical probabilities. Instead the actual system is represented by means of a set of *credence functions* assigning precise numbers to propositions, each one satisfying the standard diachronic and synchronic principles as accepted by Bayesians.[6] So if we follow van Fraassen's method of supervaluation we would say that all the Bayesian rules are definitely true. In addition Levi requires that the family of credence functions be convex. That is, if f and g are two credence functions in the set S representing the system of subjective probabilities, and if $0 < \alpha < 1$, then $\alpha f + (1 - \alpha)g$ is also a member of S.

As in the case of precise numerical probabilities there should be a connection between subjective probabilities and logical ones. We would expect that any rational indeterminate probability of the kind described by Levi should assign probabilities to propositions within the interval assigned as a logical probability by a Kyburg-type theory. In that case we could not, of course, follow both Kyburg, in his nonstandard rule for conditional probabilities, and Levi.

9 Qualitative Theories of Probability

Many have thought that it would be more fundamental to consider a purely qualitative system of subjective probabilities based upon a ranking of propositions as more or less probable.[7] Here we have two relations: the transitive and anti-reflexive relation of being more probable ($p > q$) and the equivalence relation of being equiprobable, ($p \sim q$). Moreover if $p \sim q$, $q > r$ and $r \sim s$ then $p > s$.

Because of the usual idealization we may assume (Axiom 1) that if p entails q, then $q \geq p$ (i.e. either $p \sim q$ or $q > p$). Write $p \perp q$ if p and q are inconsistent. Then we have:

AXIOM 2 the intuitive rule that if $p \perp q$, if $r \perp s$, if $p \geq r$, if $q \geq s$ and if $r \vee s \geq p \vee q$ then $p \sim r$, $q \sim s$, and $p \vee q \sim r \vee s$.

Idealizing the situation so as to assume logical omniscience, we also have a two-part Completability Principle (modeled on Ellis 1979: 9–16), as follows.

AXIOM 3 There must be an extension of the ranking to one satisfying both the principles governing qualitative probability and *trichotomy* (i.e. given any p, q either $p > q$ or $p \sim q$ or $q > p$). Moreover, if all such extensions agree that $p > q$, or agree that $p \sim q$, then we already have $p > q$, or $p \sim q$, respectively.

From these three axioms it is easy to show that if $p \geq q$, or $p \sim q$, then $\neg q > \neg p$, or $\neg q \sim \neg p$, respectively, which would otherwise have been assumed as an axiom.

We say a credence function d is *commensurate* with the qualitative system if whenever a proposition p is more probable than proposition q then $d(p) > d(q)$ and whenever p and q are equiprobable then $d(p) = d(q)$. Unfortunately Axioms 1 to 3 for qualitative probabilities stated thus far do not ensure the existence of commensurate credence functions even for a finite system of propositions. What is required is a strengthening of Axiom 2, as follows:

AXIOM 2* For any m,n, suppose r_j, $j = 1, \ldots, n$ are pairwise inconsistent, that is $r_j \perp r_k$ if $j \neq k$. And suppose that the $p_k = / \{r_j : j \in A_{kj}\}$, $q_k = / \{r_j : j \in B_{kj}\}$, where, for all k, A_{kj} and

B_{kj} are subsets of $\{1, \ldots, n\}$. Further suppose that $p_k \geq q_k$, for all k. Finally, suppose that $Vp_k = Vq_k$, then if $\#B_{kj} \geq \#A_{kj}$ for all k, j then $p_k \sim q_k$ for all k.

It is easy to see that Axiom 2* is a necessary condition for there being a commensurate credence function. For a finite system of propositions any system of qualitative probabilities satisfying Axiom 1, Axiom 2*, and Axiom 3 has a commensurate credence function.[8] Moreover the set C of all commensurate credence functions is convex and so forms a Levi system, which is easily seen to be a faithful representation in the sense that if for all $d \in C$ $d(p) > d(q)$, or $d(p) = d(q)$ then $p > q$, or $p \sim q$ respectively. That this does not extend to the infinite case could be shown by considering an example in which it would be more appropriate to consider probabilities which can take infinitesimal values than the ones actually being considered which are real valued only.[9]

If Axiom 2* is intuitive then this is a further way of justifying Levi's system. Otherwise it suggests that qualitative probabilities form an interestingly weaker kind of system of subjective probabilities.

10 The Dynamics of Subjective Probability

Carnap's theory of logical probability leaves no room for an interesting dynamics for probabilities. For if the total evidence changes from e^- to e^+ then the probability of the inference from the total evidence to some conclusion p changes from $Prob_{Log}(p|e^-)$ to $Prob_{Log}(p|e^+)$ without any change in $Prob_{Log}(p|q)$ for any p or q. In the theory of subjective probabilities the orthodoxy is the rule of conditionalization, according to which the new subjective probability $Prob^+(p)$ on coming to be certain of new evidence e equals the old conditional probability $Prob^-(p|e)$ provided $Prob^-(e) > 0$. Now the Dutch Book Argument for Conditionalization required commitment to the same *rule* by which subjective probabilities change. Any rule other than conditionalization results in the possibility of a Dutch Book. This does not exclude a position even more subjective than Bayesianism, namely resisting the suggestion that there be a rule governing the dynamics of belief. In spite of the Dutch Book Argument alternative rules such as *imaging* have been suggested. The difference between conditionalization and imaging is most easily seen by taking the probability distributions to be given by a probability measure on the set of possible worlds. The effect of coming to be certain of e is to excise all the ¬e-worlds, and redistribute their probability to the e-worlds. Conditionalization does this by preserving the relative probabilities of the e-worlds.

Imaging does this by re-assigning the probability previously assigned to an ¬e-world to the nearest e-world(s). In addition to the Dutch Book Argument for conditionalization there is Gärdenfors' argument based upon the plausible principle that if initially $Prob(q) > 0$ and $Prob(p) = 1$ then after discovering that q $Prob(p)$ should remain equal to 1. From this it is argued that no principle governing the change of subjective probabilities can contradict conditionalization. (See Gärdenfors (1988) for this and a more general discussion of imaging versus conditionalization. See also Teller (1976) for a defence of conditionalization.)

Gärdenfors' case for conditionalization illustrates once again the role of idealization. Suppose in fact Prob(e) is initially positive but rather small. Then the discovery that e might quite naturally prompt a reconsideration of the previous confidence that ¬e, resulting in a backdated change to the earlier subjective probabilities. It is assumed, however, that the person concerned is ideally rational at all times and so never has occasion to regret the earlier confidence.

11 Probability Theory and Quantum Theory

Quantum theory is open to many rival interpretations. But for a long time the most popular was to think of the state of a quantum theoretic system as specified by the probabilities of the affirmative answer if various 'questions' are asked, that is if various two-valued 'observations' are performed.[10] Often these 'observations' concern humanly unobservable entities such as quarks. So we are here considering an ideal observer. Obviously such two-valued questions correspond to propositions, so what started off as a physical theory is treated as if it were a theory of probability. However the underlying 'Sentential Calculus' is very far from classical. It is a Quantum Logic (qv) which in the most straightforward case may be represented as the lattice of closed subspaces of a Hilbert Space. The usual principles for the Calculus of Probability hold provided we replace conjunction by the intersection, disjunction by the sum, and negation by the orthogonal complement. In addition we require the probability distribution to be additive over the countable 'disjunction' of pairwise orthogonal observables. In these circumstances Gleason's Theorem tells us that, with the exception of the special case in which the Hilbert space has only two dimensions, any such probability distribution will be a mixture of pure states specified by vectors in the Hilbert space, as in the formalism of quantum mechanics.[11]

There is a rather commonsensical retort to this sophisticated but curious account. It is to insist that the propositions which correspond to the idealized observations be embedded in a larger system to include ones which are considered beyond even an ideal observer to observe. Then, it is hoped an ideal observer who was also an ideal reasoner could assign either precise numerical probabilities, or perhaps a Levi style family of credence functions, to all propositions so as to agree with the formalism in the case of the observables. The problem with this is that for many propositions the probabilities assigned would seem to be negative (Wigner 1932). In fact the currently popular Consistent Histories formulation (Omnès 1994: 122–43) restricts the propositions considered to just those which have probabilities no greater than 1 and no less than 0. I hold that this is all quite unnecessary because physicists have mistaken a mean value for a probability. There is some relevant quantity Q (mass, charge, or the number of particles of the kind considered minus the number in a quantum vacuum) which can be positive or negative, but whose mean value for the whole system is that of a single classical particle. Then the 'probability' assigned to the proposition that 'it' has such and such position, momentum, spin, etc. is in fact to be interpreted as the mean value of Q for all states such that p (see Forrest 1999).

If readers decline my kind offer to render quantum mechanics compatible with common sense, they might prefer a Kyburg-style alternative. We could assign to all propositions not allowed on the consistent histories approach, the default fuzzy probability [0,1].

Notes

1 For a defence of Bayesianism see Earman (1992).
2 The case in which the new evidence is merely probable is discussed by Jeffrey (1965).
3 Given the first three Bayesian theses, once the initial unconditional probabilities are specified, all future unconditional probabilities are then determined, provided none of the new evidence had previously had zero probability. So perhaps we should qualify the fourth thesis by allowing additional rules to govern the case in which the evidence did previously have zero probability. For instance there is the Levi–Gärdenfors method of Preservative Imaging (see Gärdenfors 1988: 117–18).
4 See also Carnap's remarks on what would, contingently, pick out one value of the parameter (Carnap 1952: chapter 3). For a more recent investigation of empirical constraints on logical probabilities, see Nolt (1990).
5 See Kyburg (1974), but perhaps the most accessible introduction to Kyburg's work is (Bogdan 1982), especially (Spielman 1982).
6 A good introduction to Levi's theory is Levi (1980). See also Bogdan (1982).
7 A first, inadequate, axiomatization of qualitative probabilities is found in De Finetti (1932). For a useful survey of some systems see the Appendix to Malmnäs (1981).
8 Axiom 2* is equivalent to strong coherence in the sense of (Malmnäs 1981: 17). Here we identify a proposition with the set of all the 'possible worlds' at which it is true, where the 'possible worlds' may in turn be thought of as maximal consistent sets of propositions. The existence of commensurate credence functions then follows from Theorem 1 (Malmnäs 1981: 31).
9 Consider a Boolean algebra with countably many atoms all of which are equiprobable. We may arrange for the axioms of qualitative probability to be satisfied yet there is no commensurate credence function.
10 See Mackey (1963) for one of the early expositions of this approach. See also Hooker (1973).
11 For a lucid exposition of the formalism of quantum theory, including Gleason's Theorem, see Hughes (1989).

References

Bogdan, Radu J. (1982) *Henry E. Kyburg, Jr. and Isaac Levi.* Dordrecht: Reidel.

Carnap, Rudolph (1950) *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Carnap, Rudolph (1952) *The Continuum of Inductive Methods.* Chicago: University of Chicago Press.

De Finetti, B (1932) La prévision: ses lois logiques, ses sources subjectives. *Annales de l'institut Henri Poincaré*, 7, 1–68.

Earman, John (1992) *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge, MA: MIT Press.

Ellis, Brian (1979) *Rational Belief Systems.* Oxford: Blackwell.

Forrest, Peter (1999) In defence of the phase space picture. *Synthese*, 119, 299–311.

Gärdenfors, Peter (1988) *Knowledge in Flux: Modelling the Dynamics of Epistemic States.* Cambridge, MA: Bradford/MIT.

Hailperin, Theodore (1990) Probability logic in the twentieth century. *History and Philosophy of Logic*, 71–110.

Harper, William L. (1982) Kyburg on direct inference. In Bogdan (1982), 97–128.

Hintikka, Jaakko (1966) A two-dimensional continuum of inductive methods. In Jaakko Hintikka and Patrick Suppes (eds.), *Aspects of Inductive Logic* (pp. 113–32). Amsterdam: North Holland.

Hooker, Cliff (ed.) (1973) *Contemporary Research in the Foundations and Philosophy of Quantum Theory.* Boston, MA: Reidel.

Hughes, R. I. G. (1989) *The Structure and Interpretation of Quantum Mechanics.* Cambridge, MA: Harvard University Press.

Jeffrey, R. C. (1965) *The Logic of Decision*. New York: McGraw-Hill.

Johnson, W. E. (1921–4) *Logic*, vols. 1–3. Cambridge: Cambridge University Press.

Kemeny J. (1955) Fair bets and inductive probabilities. *Journal of Symbolic Logic*, 20, 263–73.

Keynes, John Maynard (1921) *A Treatise on Probability*. London: Macmillan.

Kyburg, Henry E. (1974) *The Logical Foundations of Statistical Inference.* Dordrecht: Reidel.

Levi, Isaac (1980) *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge: MIT Press.

Mackey, George W. (1963) *The Mathematical Foundations of Quantum Mechanics.* Amsterdam: Benjamin.

Malmnäs, Per-Erik (1981) *From Qualitative to Quantitative Probability (Stockholm Studies in Philosophy 7).* Stockholm: University of Stockholm Press.

Nolt, John (1990) A fully logical inductive logic. *Notre Dame Journal of Formal Logic*, 415–36.

Omnès, Roland (1994) *The Interpretation of Quantum Mechanics.* Princeton, NJ: Princeton University Press.

Schanuel, Stephen H. (1982) What is the length of a potato? An introduction to geometric measure theory. In A. Dold and B. Eckmann (eds.), *Categories in Continuum Physics, Lecture Notes in Mathematics* (pp. 118–25). Berlin: Springer.

Spielman, Stephen (1982) Kyburg's system of probability. In (Bogdan 1982), 57–96.

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

Teller, Paul (1976) Conditionalization, observation, and change of preference. In W. L. Harper and C. L. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science.* Dordrecht: Reidel.

Wigner, E. (1932) On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40, 749–59; reprinted in Kim, Y. S. and Noz, M. E. (eds.), *Phase Space Picture of Quantum Mechanics: Group Theoretical Approach.* Singapore: World Scientific (1991) 219–31.

Zabell, S. L. (1996) Confirming universal generalisations. *Erkenntnis*, 45, 267–283.