Bayesianism with a Human Face

What's a Bayesian?

Well, I'm one, for example. But not according to Clark Glymour (1980, pp. 68-69) and some other definers of Bayesianism and personalism, such as Ian Hacking (1967, p. 314) and Isaac Levi (1980, p. xiv). Thus it behooves me to give an explicit account of the species of Bayesianism I espouse (sections 1 and 2) before adding my bit (section 3, with lots of help from my friends) to Daniel Garber's treatment in this volume of the problem of new explanation of common knowledge: the so-called problem of old evidence.

With Clark Glymour, I take there to be identifiable canons of good thinking that get used on a large scale in scientific inquiry at its best; but unlike him, I take Bayesianism (what *I* call "Bayesianism") to do a splendid job of validating the valid ones and appropriately restricting the invalid ones among the commonly cited methodological rules. With Daniel Garber, I think that bootstrapping does well, too—when applied with a tact of which Bayesianism can give an account. But my aim here is to elaborate and defend Bayesianism (of a certain sort), not to attack bootstrapping. Perhaps the main novelty is the further rounding-out in section 3 (by John Etchemendy, David Lewis, Calvin Normore, and me) of Daniel Garber's treatment of what I have always seen as the really troubling one of Clark Glymour's strictures against Bayesianism. After that there is a coda (section 4) in which I try to display and explain how probability logic does so much more than truth-value logic.

1. Response to New Evidence

In Clark Glymour's book, you aren't a Bayesian unless you update your personal probabilities by conditioning (a.k.a. "conditionalization"), i.e., like this:

> As new evidence accumulates, the probability of a proposition changes according to Bayes' rule: the posterior probability of a hypothesis on the new evidence is equal to the prior conditional probability of the hypothesis on the evidence. (p. 69)

That's one way to use the term "Bayesian," but on that usage I'm no Bayesian. My sort of Bayesianism gets its name from another sense of the term "Bayes's rule," equally apt, but stemming from decision theory, not probability theory proper. Whereas Bayes's rule in Glymour's sense prescribes conditioning as the way to update personal probabilities, Bayes's rule in my sense prescribes what Wald (1950) called "Bayes solutions" to decision problems, i.e., solutions that maximize expected utility relative to some underlying probability assignment to the states of nature. (No Bayesian himself, Wald contributed to the credentials of decision-theoretic Bayesianism by proving that the Bayes solutions form a complete class.) The Reverend Thomas Bayes was both kinds of Bayesian. And of course, he was a third kind of Bayesian, too: a believer in a third sort of Bayes's rule, according to which the right probability function to start with is m* (as Carnap (1945) was to call it).

Why am I not a Bayesian in Glymour's sense? This question is best answered by way of another: What is the "new evidence" on which we are to condition? (Remember: the senses are not telegraph lines on which the external world sends observation sentences for us to condition upon.) Not just any proposition that newly has probability one will do, for there may well be many of these, relative to which conditioning will yield various posterior probability distributions when applied to the prior.

All right, then: what about the conjunction of all propositions that newly have probability one? That will be the total new evidence, won't it? Why not take the kinematical version of Bayes's rule to prescribe conditioning on that total?

I answer this question in Chapter 11 of my book (1965, 1983), and in a few other places (1968, 1970, 1975). In a nutshell, the answer is that much of the time we are unable to formulate any sentence upon which we are prepared to condition, and in particular, the conjunction of all the sentences that newly have probability one will be found to leave too much out for it to serve as the Archimedean point about which we can move our probabilities in a satisfactory way. Some of the cases in which conditioning won't do are characterized by Ramsey (1931, "Truth and Probability," end of section 5) as follows:

> I think I perceive or remember something but am not sure; this would seem to give me some ground for believing it, contrary to Mr. Keynes' theory, by which the degree of belief in it which it would be rational for me to have is that given by the probability relation between the

proposition in question and the things I know for certain.

Another sort of example is suggested by Diaconis and Zabell (1982): a record of someone reading Shakespeare is about to be played. Since you are sure that the reader is either Olivier or Gielgud, but uncertain which, your prior probabilities for the two hypotheses are nearly equal. But now comes fresh evidence, i.e., the sound of the reader's voice when the record is played. As soon as your hear that, you are pretty sure it's Gielgud, and the prior value ca. .5 is replaced by a posterior value ca. .9, say. But, although it was definite features of what you heard that rightly made you think it very likely to have been Gielgud, you cannot describe those features in an observation sentence in which you now have full belief, nor would you be able to recognize such a sentence (immensely long) if someone else were to produce it.

Perhaps it is the fact that there surely is definite evidence that prompts and justifies the probability shift in the Olivier/Gielgud case, that makes some people think there must be an evidence *sentence* (observation sentence) that will yield the new belief function via conditionalization. Surely it is all to the good to be able to say just what it was about what you heard, that made you pretty sure it was Gielgud. But few would be able to do that; nor is such inability a mark of irrationality; nor need one be able to do that in order to count as having had good reason to be pretty sure it was Gielgud. The Olivier/Gielgud case is typical of our most familiar sorts of updating, as when we recognize friends' faces or voices or handwritings pretty surely, and when we recognize familiar foods pretty surely by their look, smell, taste, feel, and heft.

Of course conditioning is sometimes appropriate. When? I mean, if your old and new belief functions are p and q, respectively, when is q of form p_E for some E to which p assigns a positive value? (Definition: $p_E(H)$ is the conditional probability of H on E, i.e., $p(H/E)$, i.e., $p(HE)/p(E)$.)

Here is an answer to the question:

(C) If $p(E)$ and $q(E)$ are both positive, then the conditions (a) $q_E = p_E$ and (b) $q(E) = 1$ are jointly necessary and sufficient for (c) $q = p_E$.

You can prove that assertion on the back of an envelope, via the Kolmogorov axioms and the defition of conditional probability. Here is a rough-and-ready verbal summary of (C):

> Conditioning is the right way to update your probability judgments iff the proposition conditioned upon is not only (b) one you now fully

believe, but is also (a) one whose relevance to each proposition is unchanged by the updating.

The point of view is one in which we take as given the old and new probability functions, p and q, and then ask whether the condition (c) $q = p_E$ is consistent with static coherence, i.e., the Kolmogorov axioms together with the definition of conditional probability applied to p and q separately. In (C), (a) is the ghost of the defunct condition of total evidence.

In the Olivier/Gielgud example, and others of that ilk, fresh evidence justifies a change from p to q even though $q \neq p_E$ for all E in the domain of p. What is the change, and when is it justified? Here is the answer, which you can verify on the back of the same envelope you used for (C):

(K) If "E" ranges over some partitioning of a proposition of p-measure 1 into propositions of positive p-measure, then the ("rigidity") condition

(r) $q_E = p_E$ for all E in the partitioning

is necessary and sufficient for q to be related to p by the following ("kinematical") formula:

(k) $q = \Sigma_E \, q(E)p_E$.

There is no more question of justifying (k) in (K) than there was of justifying (c) in (C): neither is always right. But just as (C) gives necessary and sufficient conditions (a) and (b) for (c) to be right, so (K) gives (r), i.e., the holding of (a) for each E in the partitioning, as necessary and sufficient for (k) to be correct—where in each case, correctness is just a matter of static coherence of p and q separately. We know when (k) is right:

The kinematical scheme (k) yields the correct updating iff the relevance of each member E of the partitioning to each proposition H is the same after the updating as it was before.

It is an important discovery (see May and Harper 1976; Williams 1980; Diaconis and Zabell 1982) that in one or another sense of "close," (k) yields a measure q that is closest to p among those that satisfy the rigidity condition (r) and assign the new probabilities q(E) to the Es, and that (c) yields a measure that is closest to p among those that satisfy the conditions (a) and (b) in (C). But what we thereby discover is that (so far, anyway) we have adequate concepts of closeness: we already knew that (k) was equivalent to (r), and that (c) was equivalent to (a) and (b) in (C). This is not to deny the interest of such minimum-change principles, but rather to

emphasize that their importance lies not in their serving to justify (c) and (k)—for they don't—but in the further kinematical principles they suggest in cases where (k) holds for no interesting partitioning. To repeat: (c) and (k) are justified by considerations of mere coherence, where their proper conditions of applicability are met, i.e., (a) and (b) for (c), and (r) for (k). And where those conditions fail, the corresponding rules are unjustifiable.

Observe that in a purely formal sense, condition (r) is very weak; e.g., it holds whenever the Boolean algebra on which p and q are defined has atoms whose p-values sum to 1. (Proof: with "E" in (r) ranging over the atoms that have positive p-measure, p(H/E) and q(H/E) will both be 1 or both be 0, depending on whether E implies H or −H.) Then in particular, (k) is always applicable in a finite probability space, formally. But if (k) is to be useful to a human probability assessor, the E partitioning must be coarser than the atomistic one. To use the atomistic partitioning is simply to start over from scratch.

The Olivier/Gielgud example is one in which the partitioning is quite manageable: {O, G}, say, with O as the proposition that the reader is Olivier, and G for Gielgud. The hypothesis H that the reader (whoever he may be) loved Vivien Leigh serves to illustrate the rigidity conditions. Applied to H, (r) yields

$$q(H/O) = p(H/O), \quad q(H/G) = p(H/G).$$

Presumably these conditions both hold: before hearing the reader's voice you attributed certain subjective probabilities to Olivier's having loved Leigh (high), and to Gielgud's having done so (low). Nothing in what you heard tended to change those judgments: your judgment about H changed only incidentally to the change in your judgment about O and G. Thus, by (k),

$$q(H) = q(O)p(H/O) + q(G)p(H/G).$$

q(H) is low because it is a weighted average of p(H/O), which was high, and p(H/G), which was low, with the low value getting the lion's share of the weight: $q(G) = .9$.

2. Representation of Belief

In Clark Glymour's book, Bayesianism is identical with personalism, and requires not only updating by conditioning, but also a certain superhuman completeness:

There is a class of sentences that express all hypotheses and all actual or possible evidence of interest; the class is closed under Boolean operations. For each ideally rational agent, there is a function defined on all sentences such that, under the relation of logical equivalence, the function is a probability measure on the collection of equivalence classes. (pp. 68-69)

The thought is that Bayesian personalism must represent one's state of belief at any time by a definite probability measure on some rather rich language. And indeed the two most prominent personalists seem to espouse just that doctrine: de Finetti (1937) was at pains to deny the very meaningfulness of the notion of unknown probabilities, and Savage (1954) presented an axiomatization of preference according to which the agent's beliefs must be represented by a unique probability.

But de Finetti was far from saying that personal probabilities cannot fail to *exist*. (It is a separate question, whether one can be unaware of one's existent partial beliefs. I don't see why not. See Mellor (1980) and Skyrms (1980) for extensive discussions of the matter.) And Savage was far from regarding his 1954 axiomatization as the last word on the matter. In particular, he viewed as a live alternative the system of Bolker (1965) and Jeffrey (1965), in which even a (humanly unattainable) complete preference ranking of the propositions expressible in a rich language normally determines no unique probability function, but rather an infinite set of them. The various members of the set will assign various values throughout intervals of positive length to propositions about which the agent is not indifferent: see Jeffrey (1965, section 6.6) for details.

Surely the Bolker-Jeffrey system is not the last word, either. But it does give one clear version of Bayesianism in which belief states—even superhumanly definite ones—are naturally identified with infinite sets of probability functions, so that degrees of belief in particular propositions will normally be determined only up to an appropriate quantization, i.e., they will be interval-valued (so to speak). Put it in terms of the thesis of *the primacy of practical reason*, i.e., a certain sort of pragmatism, according to which belief states that correspond to identical preference rankings of propositions are in fact one and the same. (I do not insist on that thesis, but I suggest that it is an intelligible one, and a clearly Bayesian one; e.g., it conforms to Frank Ramsey's (1931) dictum (in "Truth and Probability," section 3): "the kind of measurement of belief with which probability is concerned . . . is a measurement of belief *qua* basis of action.") Applied to

the Bolker-Jeffrey theory of preference, the thesis of the primacy of practical reason yields the characterization of belief states as sets of probability functions (Jeffrey 1965, section 6.6). Isaac Levi (1974) adopts what looks to me like the same characterization, but labels it "un-Bayesian."

But of course I do not take belief states to be determined by full preference rankings of rich Boolean algebras of propositions, for our actual preference rankings are fragmentary, i.e., they are rankings of various subsets of the full algebras. Then even if my theory were like Savage's in that full rankings of whole algebras always determine unique probability functions, the actual, partial rankings that characterize real people would determine belief states that are infinite sets of probability functions on the full algebras. Here is the sort of thing I have in mind, where higher means better:

A, B	C
	D
W	W
	−C
−A, −B	−D
(1)	(2)

This is a miniature model of the situation in which the full Boolean algebra is infinite. Here the full algebra may be thought of as consisting of the propositions A, B, C, D and their truth-functional compounds. W is the necessary proposition, i.e., $W = Av - A = Cv - C$, etc. Here is a case in which the agent is indifferent between A and B, which he prefers to W, which in turn he prefers to −A and to −B, between which he is indifferent. But he has no idea where AB, $Av - B$, etc. come in this ranking: his preferences about them remain indeterminate. That is what ranking (1) tells us. And ranking (2) gives similar sorts of information about C, D, and their denials: the agent's preferences regarding CD, $Cv - D$, etc. are also indeterminate. But the two rankings are related only by their common member, W. Thus C and D are preferred to −A and −B, but there is no information given about preference between (say) A and C.

That is the sort of thing that can happen. According to (1), we must have $p(A) = p(B)$ for any probability function p in the belief state determined by preferences (1) and (2): see Example 3 in chapter 7 of Jeffrey (1965). And according to (2), we must have $p(C) < p(D)$ for any such p: see problem 1 in

section 7.7. Then the belief state that corresponds to this mini-ranking (or this pair of connecting mini-rankings) would correspond to the set $\{p:p(A) = p(B)$ and $p(C) < p(D)\}$.

The role of definite probability measures in probability logic as I see it is the same as the role of maximal consistent sets of sentences in deductive logic. Where deductive logic is applied to belief states conceived unprobabilistically as *holdings true* of sets of sentences, maximal consistent sets of sentences play the role of unattainable completions of consistent human belief states. The relevant fact in deductive logic is

> *Lindenbaum's Lemma:* a truth value assignment to a set of sentences is consistent iff consistently extendible to the full set of sentences of the language.

(There is a one-to-one correspondence between consistent truth-value assignments to the full set of sentences of the language and maximal consistent sets of sentences of the language: the truth value assigned is t or f depending on whether the sentence is or is not a member of the maximal consistent set.) The corresponding fact about probability logic is what one might call

> *De Finetti's Lemma:* an assignment of real numbers to a set of sentences is coherent (= immune to Dutch books) iff extendible to a probability function on the full set of sentences of the language.

(See de Finetti 1972, section 5.9; 1974, section 3.10.)

It is a mistake to suppose that someone who assigns definite probabilities to *A* and to *B* (say, .3 and .8 respectively) is thereby committed in Bayesian eyes to some definite probability assignment to the conjunction *AB*, if de Finetti (1975, (2) on p. 343, and pp. 368-370) is to be counted as a Bayesian. On the other hand, probability logic in the form of the Kolmogorov axioms, say, requires that any assignment to that conjunction lie in the interval from .1 to .3 if the assignments $p(A) = .3$ and $p(B) = .8$ are to be maintained: see Boole (1854, Chapter 19), Hailperin (1965), or Figure 1 here. Thus probability logic requires that one or both of the latter assignments be abandoned in case it is discovered that *A* and *B* are logically incompatible, since then $p(AB) = 0 < .1$.

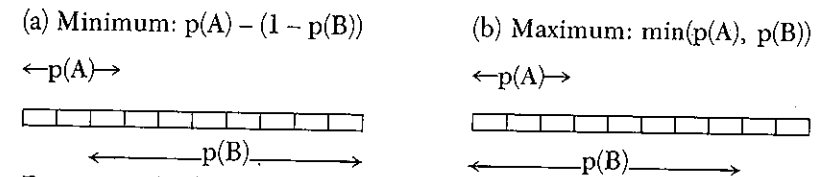Clearly indeterminacies need not arise as that of $p(AB)$ did in the

(a) Minimum: $p(A) - (1 - p(B))$

$\leftarrow p(A) \rightarrow$

(b) Maximum: $\min(p(A), p(B))$

$\leftarrow p(A) \rightarrow$

$\longleftarrow p(B) \longrightarrow$ $\longleftarrow p(B) \longrightarrow$

Figure 1. $p(AB)$ is the length of overlap between the two segments.

foregoing example, i.e., out of an underlying determinate assignment to the separate components of the conjunction. See Williams (1976) for an extension of de Finetti's lemma to the case where the initial assignment of real numbers $p(S_i) = r_i$ is replaced by a set of conditions of form $r_i \leq p(S_i) \leq s_i$. And note that indeterminacies need not be defined by such inequalities as these. They might equally well be defined by conditions (perhaps inequalities) on the mathematical expectations $E(X_i)$ of random variables—conditions that impose conditions on the underlying probability measures p via the relation $E(X_i) = \int_W X_i \, dp$. More complex special cases arise when "X_i" is replaced by "$(X_i - EX_i)^2$", etc., so that belief states are defined by conditions on the variances etc. of random variables.

Such definitions might be thought of as generalizing the old identification of an all-or-none belief state with the proposition believed. For propositions can be identified with sets of two-valued probability measures. Each such measure, in which the two values must be 0 and 1, can be identified with the possible world in which the true statements are the ones of probability 1. Then a set of such measures works like a set of possible worlds: a proposition. Now Levi and I take belief states to be sets of probability measures, omitting the requirement that they be two-valued. Call such sets "probasitions." The necessary probasition is the set P of all probability measures on the big Boolean algebra in question. P is the *logical space* of probability logic. My current belief state is to be represented by a probasition: a region R in this space. If I now condition upon a proposition E, my belief state changes from R to

$$R/E = _{Df} \{p_E : p \ \varepsilon \ R \text{ and } p(E) \neq 0\}.$$

Perhaps R/E is a proper subset of R, and perhaps it is disjoint from R, but for the most part one would expect the change from R to R/E to represent a new belief state that *partly* overlaps the old one. And in some cases one

would expect the operation of conditioning to *shrink* probasitions, e.g., perhaps in the sense that the diameter of R/E is less than that of R when diameter is defined

$$\text{diam}(R) = \sup_{p,q \in R} \| p - q \|$$

and the norm $\| p - q \|$ is defined

$$\| p - q \| = \sup_A | p(A) - q(A) |$$

where "A" ranges over all propositions in the Boolean algebra.

That is how I would ride that hobby-horse. But I would also concede I. J. Good's (1952, 1962) point, that probasitions are only rough characterizations of belief states, in which boundaries are drawn with artificial sharpness, and variations in the acceptability of different members of probasitions and in the unacceptability of various nonmembers go unmarked. In place of probasitions, one might represent belief states by probability measures μ on suitable Boolean algebras of subsets of the space P. Good himself rejects that move because he thinks that μ would then be equivalent to some point μ^* in P, i.e., the point that assigns to each proposition A the definite probability

$$\mu^*(A) = \int_{p \in P} p(A) \, d\mu(p).$$

In our current terminology, the thought is that a definite probability measure μ on P must correspond to a sharp belief state, viz., the probasition $\{\mu^*\}$. To avoid this reduction, Good proposes that μ be replaced by a nondegenerate probasition of type 2, i.e., a nonunit set of probability measures on P; that in principle, anyway, that probasition of type 2 be replaced by a nondegenerate probasition of type 3; and so on. "It may be objected that the higher the type the woolier the probabilities. It will be found, however, that the higher the type the less wooliness matters, provided the calculations do not become too complicated." (Good 1952, p. 114)

But I do not see the need for all that. It strikes me that here, Good is being misled by a false analogy with de Finetti's way of avoiding talk of unknown probabilities (i.e., the easy converse of his representation theorem for symmetric probability functions). De Finetti's point was that where objectivists would speak of (say) coin-tossing as a binominal process with unknown probability of success on each toss, and might allow that their subjective probability distribution for the unknown objective probability x of success is uniform throughout the unit interval, an uncompromis-

ing subjectivist can simply have as his belief function the subjectively weighted average of the various putatively objective possibilities, so that, e.g., his subjective probability for heads on the first n tosses would be

$$p(H_1 H_2 \ldots H_n) = \int_0^1 x^n dx = \frac{1}{n + 1}.$$

In the analogy that Good is drawing, the probasition R is the set all binomial probability functions p_x where $p_x(H_i) = x$, and μ is the probability measure on P that assigns measure 1 to R and assigns measure $b - a$ to any nonempty subset $\{p_x : a \leq x < b\}$ of R. But whereas for de Finetti the members of R play the role of (to him, unintelligible) hypotheses about what the objective probability function might be, for Good the members of R play the role of hypotheses about what might be satisfactory as a subjective probability function. But if only the members of R are candidates for the role of satisfactory belief function, their subjectively weighted average, i.e., p as above, is not a candidate. (*That* p is not binomial: $p(H_1/H_2) = 2/3 \neq p(H_1) = 1/2$, whereas for each p_x in R, $p_x(H_1/H_2) = p(H_1) = x$.)

The point is that the normalized measure μ over P is not being used as a subjective probability distribution that indicates one's degrees of belief in such propositions as that the true value of x lies between .1 and .3. On the contrary, the uniformity of the μ distribution within R is meant to indicate that one would be indifferent between having to behave in accordance with p_x and having to behave in accordance with p_y for any x and y in the unit interval (where such behavior is determined as well by his utility function); and the fact that $\mu(R) = 1$ is meant to indicate that one would prefer having to behave in accordance with any member of R to having to behave in accordance with any member of P – R. (These are rough characterizations because μ assigns measure 0 to each unit subset of P. A precise formulation would have to talk about having to behave in accordance with randomly selected members of intervals, $\{p_x : a \leq x \leq b\}$.)

Then I think Good's apprehension unfounded: I think one can replace probasitional belief states R by probability distributions μ over P that assign most of their mass to R, without thereby committing oneself to a belief state that is in effect a singleton probasition, $\{\mu^*\}$. But this is not to say that one must always have a sharp probability distribution over P: perhaps Good's probasitions of types 2 and higher are needed in order to do justice to the complexities of our belief states.

On the other hand, I think that in practice, even the relatively simple transition from probasitional belief states to belief states that are sharp probability measures on P is an idle complexity: the probasitional representation suffices, anyway, for the applications of probability logic that are considered in the remainder of this paper.

An important class of such examples is treated in Chapter 2 of de Finetti (1937), i.e., applications of what I shall call

> de Finetti's Law of Small Numbers: the estimated number of truths among the propositions $A_1, \ldots, A_n$ must equal the sum of their probabilities.

That follows from the additivity of the expectation operator and the fact that the probability you attribute to A is always equal to your estimate of the number of truths in the set {A}: as de Finetti insists, the thing is as trivial as Bayes's theorem. (He scrupulously avoids applying any such grand term as "law" to it.) Dividing both sides of the equation by n, the law of small numbers takes this form:

> The estimated relative frequency of truths among the propositions is the average $(p(A_1) + \ldots + p(A_n))/n$ of their probabilities.

Suppose, then, that you regard the A's as equiprobable but have no view about what their common probability is (i.e., you have no definite degree of belief in the A's), and suppose that tomorrow you expect to learn the relative frequency of truths among them, without learning anything that will disturb your sense of their equiprobability. Thus you might represent your belief state tomorrow by the probasition {p:p(A_1) = ... = p(A_n)}, or by a measure on P that assigns a value near 1 to that probasition. But what's the point? If you don't need to do anything on which tomorrow's belief state bears until tomorrow, you may as well wait until you learn the relative frequency of truths among the A's, say, r. At that point, your estimate of the relative frequency of truths will be r (with variance o), and by mere coherence your degree of belief in each of the A's will also be r. You know all that today.

Note that in the law of small numbers, the A's need not be independent, or exchangeable, or even distinct! The "law" is quite general: as general and as trivial as Bayes's theorem, and as useful.

A mistake that is easy to make about subjectivism is that anything goes, according to that doctrine: any weird belief function will do, as long as it is coherent.

The corresponding mistake about dress would go like this: any weird getup will do, if there are no sumptuary laws, or other laws prohibiting inappropriate dress. That's wrong, because in the absence of legislation about the matter, people will generally dress as they see fit, i.e., largely in a manner that they think appropriate to the occasion and comfortable for them on that occasion. The fact that it is legal to wear chain mail in city buses has not filled them with clanking multitudes.

Then have no fear: the fact that subjectivism does not prohibit people from having two-valued belief functions cannot be expected to produce excessive opinionation in people who are not so inclined, any more than the fact that belief functions of high entropy are equally allowable need be expected to have just the opposite effect. For the most part we make the judgments we make because it would be unthinkable not to. Example: the foregoing application of de Finetti's law of small numbers, which explains to the Bayesian why knowledge of frequencies can have such powerful effects on our belief states.

The other side of the coin is that we generally suspend judgment when it is eminently thinkable to do so. For example, if I expect to learn the frequency tomorrow, and I have no need for probabilistic belief about the A's today, then I am not likely to spend my time on the pointless project of eliciting my current degrees of belief in the A's. The thought is that we humans are not capable of adopting opinions gratuitously, even if we cared to do so: we are generally at pains to come to opinions that strike us as right, or reasonable for us to have under the circumstances. The laws of probability logic are not designed to prevent people from yielding to luscious doxastic temptations—running riot through the truth values. They are designed to help us explore the ramifications of various actual and potential states of belief—our own or other people's, now or in the past or the future. And they are meant to provide a Bayesian basis for methodology. Let us now turn to that—focussing especially on the problem ("of old evidence") that Clark Glymour (1980, Chapter 3) identifies as a great Bayesian sore point.

3. The Problem of New Explanation

Probability logic is typically used to reason in terms of partially specified probability measures meant to represent states of opinion that it would be fairly reasonable for people to be in, who have the sort of information we take ourselves to have, i.e., we who are trying to decide how to proceed in

some practical or (as here) theoretical inquiry. Reasonableness is assessed by us, the inquirers, so that what none of us is inclined to believe can be summarily ruled out, but wherever there is a real issue between two of us, or whenever one of us is of two minds, both sides are ruled reasonable. Of course, if our opinions are too richly varied, we shall get nowhere; but such radical incommensurability is less common in real inquiry, even in revolutionary times, than romantics would have us think.

It is natural to speak of "the unknown" probability measure p that it would be reasonable for us to have. This is just a substitute for more intelligible speech in terms of a variable "p" that ranges over the probasition (dimly specified, no doubt) comprising the probability measures that we count as reasonable. Suppose now that in the light of evidence that has come to our attention, we agree that p should be modified in a certain way: replaced by another probability measure, p'. If p'(H) exceeds p(H) for each allowable value of "p," we regard the evidence as supporting or confirming H, or as positive for H. The degree of support or confirmation is

$$p'(H) - p(H)$$

In the simplest cases, where $p' = p_E$, this amounts to

$$p(H/E) - p(H),$$

and in somewhat more complicated cases, where p' comes from p by kinematics relative to the partitioning $\{E_1, \ldots, E_n\}$, it amounts to

$$\Sigma_i p(H/E_i) \, (p'(E_i) - p(E_i)).$$

But what if the evidence is the fresh demonstration that H implies some known fact E? In his contribution to this volume, Daniel Garber shows that—contrary to what one might have thought—it is not out of the question to represent the effect of such evidence in the simplest way, i.e., by conditioning on the proposition H ⊢ E that H implies E, so that H's implying E supports H if and only if

(1) $p(H/H \vdash E) > p(H)$.

And as he points out in his footnote 28, this inequality is equivalent to either of the following two (by Bayes' theorem, etc.):

(2) $p(H \vdash E/H) > p(H \vdash E)$

(3) $p(H \vdash E/H) > p(H \vdash E/ -H)$

This equivalence can be put in words as follows:

(I) A hypothesis is supported by its ability to explain facts in its *explanatory domain*, i.e., facts that it was antecedently thought likelier to be able to explain if true than if false.

(This idea was suggested by Garber some years ago, and got more play in an early version of his paper than in the one in this volume.) This makes sense intuitively. Example: Newton saw the tidal phenomena as the sorts of things that ought to be explicable in terms of the hypothesis H of universal gravitation (with his laws of motion and suitable background data) if H was true, but quite probably not if H was false. That is why explanation of those phenomena by H was counted as support for H. On the other hand, a purported theory of acupuncture that implies the true value of the gravitational red shift would be undermined thereby: its implying *that* is likely testimony to its implying everything, i.e., to its inconsistency.

But something is missing here, namely the supportive effect of belief in E. Nothing in the equivalence of (1) with (2) and (3) depends on the supposition that E is a "known fact," or on the supposition that p(E) is 1, or close to 1. It is such suppositions that make it appropriate to speak of "explanation" of E by H instead of mere implication of E by H. And it is exactly here that the peculiar problem arises, of old knowledge newly explained. As E is common knowledge, its probability for all of us is 1, or close to it, and therefore the probability of H cannot be increased much by conditioning on E before conditioning on H ⊢ E (see 4a)—or after (see 4b), unless somehow the information that H ⊢ E robs E of its status as "knowledge."

(4a) $p(H/E) \approx p(H)$ if $p(E) \approx 1$

(4b) $p_{H\vdash E}(H/E) \approx p_{H\vdash E}(H)$ if $p_{H\vdash E}(E) \approx 1$

As (4b) is what (4a) becomes when "p" is replaced by "$p_{H\vdash E}$" throughout, we shall have proved (4b) as soon as we have proved (4a) for arbitrary probability functions p. Observe that (4b) comes to the same thing as this:

$$p(H/E \ \& \ H \vdash E) \approx p(H/H \vdash E) \text{ if } p(E/H \vdash E) \approx 1.$$

There and in (4), statements of form $x \approx y$ are to be interpreted as saying that x and y differ by less than a ("small") unspecified positive quantity, say ε.

Proof of (4a). The claim is that for all positive ε,

if $p(-E) < \varepsilon$ then $-\varepsilon < p(H/E) - p(H) < \varepsilon$.

To prove the "$< \varepsilon$" part of the consequent, observe that

$$p(H/E) - p(H) \leq p(H/E) - p(HE) \text{ since } p(HE) \leq p(H)$$

$$= \frac{p(HE)}{p(E)} - p(HE) = \frac{p(HE)p(-E)}{p(E)}$$

$$\leq p(-E) \text{ since } p(HE) \leq p(E).$$

Then $p(H/E) - p(H) < \varepsilon$. To prove the "$-\varepsilon <$" part, note that it is true iff

$$p(H) - \frac{p(HE)}{p(E)} < \varepsilon, \text{ i.e., iff}$$

$$p(H - E) + \left(\frac{p(HE)}{1} - \frac{p(HE)}{p(E)}\right) < \varepsilon, \text{ i.e., iff}$$

$$\frac{p(HE)}{p(E)}(p(E) - 1) < \varepsilon - p(H - E)$$

where the left-hand side is 0 or negative since $p(E) \leq 1$, and where the right-hand side is positive since $p(H-E) \leq p(-E) < \varepsilon$. Then the "$-\varepsilon <$" part of the consequent is also proved.

Yet, in spite of (4), where E reports the facts about the tides that Newton explained, it seems correct to say that his explanation gave them the status of evidence supporting his explanatory hypotheses, H—a status they are not deprived of by the very fact of being antecedently known.

But what does it mean to say that Newton regarded H as the sort of hypothesis that, if true, ought to imply the truth about the tides? I conjecture that Newton thought his theory ought to explain the truth about the tides, *whatever that might be*. I mean that I doubt whether Newton knew such facts as these (explained in *The System of the World*) at the time he formulated his theory:

[39.] The tide is greatest in the syzygies of the luminaries and least in their quadratures, and at the third hour after the moon reaches the meridian; outside of the syzygies and quadratures the tide deviates somewhat from that third hour towards the third hour after the solar culmination.

Rather, I suppose he hoped to be able to show that

(T)   H implies the true member of $\mathscr{E}$

where H was his theory (together with auxiliary data) and $\mathscr{E}$ was a set of mutually exclusive propositions, the members of which make various

claims about the tides, and one of which is true. I don't mean that he was able to specify $\mathscr{E}$ by writing out sentences that express its various members. Still less do I mean that he was able to identify the true member of $\mathscr{E}$ by way of such a sentence, to begin with. But he knew where to go to find people who could do that to his satisfaction: people who could assure him of such facts as [39.] above, and the others that he explains at the end of his *Principia* and in *The System of the World*. Thus you can believe T (or doubt T, or hope that T, etc.) without having any views about which member of $\mathscr{E}$ is the true one, and, indeed, without being able to give an account of the makeup of $\mathscr{E}$ of the sort you would need in order to start trying to deduce members of $\mathscr{E}$ from H. (Nor do I suppose it was clear, to begin with, what auxiliary hypotheses would be needed as conjuncts of H to make that possible, until the true member of $\mathscr{E}$ was identified.)

David Lewis points out that in these terms, Garber's equivalence between (1) and (2) gives way to this:

(5) $p(H/T) > p(H)$ iff $p(T/H) > p(T)$.

Lewis's thought is that someone in the position I take Newton to have been in, i.e., setting out to see whether T is true, is in a position of being pretty sure that

(S)   H implies *some* member of $\mathscr{E}$

without knowing which, and without being sure or pretty sure that (T) *the member of $\mathscr{E}$ that H implies is the true one*. But in exactly these circumstances, one will take truth of T to support H. Here I put it weakly (with "sure" instead of "pretty sure," to make it easy to prove):

(II)   If you are sure that H implies *some* member of $\mathscr{E}$, then you take H to be supported by implying the *true* member of $\mathscr{E}$ unless you were already sure it did.

Proof. The claim is that

if $p(S) = 1 \neq p(T)$ then $p(H/T) > p(H)$.

Now if $p(S) = 1$ then $p(S/H) = 1$ and therefore $p(T/H) = 1$ since if H is true it cannot imply any falsehoods. Thus, if $1 \neq p(T)$, i.e., if $1 > p(T)$, then $p(T/H) > p(T)$, and the claim follows via (5).

Notice one way in which you could be sure that H implies the true member of $\mathscr{E}$: you could have known which member that was, and cooked H up to imply it, e.g., by setting $H = EG$ where E is the true member of $\mathscr{E}$

and G is some hypothesis you hope to make look good by association with a known truth.

Now (II) is fine as far as it goes, but (John Etchemendy points out) it fails to bear on the case in which it comes as a surprise that H implies *anything* about (say) the tides. The requirement in (II) that $p(S)$ be 1 is not then satisfied, but H may still be supported by implying the true member of $\mathscr{C}$. It needn't be, as the acupuncture example shows, but it may be. For example, if Newton had not realized that H ought to imply the truth about the tides, but had stumbled on the fact that $H \vdash E$ where E was in $\mathscr{C}$ and known to be true, then H would have been supported by its ability to explain E.

Etchemendy's idea involves the propositions S, T, and

(F)   H implies some false member of $\mathscr{C}$.

Evidently $F = S - T$, so that $-F$ is the material conditional, $-F = -S \lor T$ ("If H implies any member of $\mathscr{C}$ then it implies the true one"), and so the condition $p(F) = 0$ indicates full belief in that conditional. Etchemendy points out that Lewis's conditions in (II) can be weakened to $p(F) \neq 0$ and $p(HS) = p(H)p(S)$; i.e., you are not antecedently sure that H implies nothing false about X (about the tides, say), and you take truth of H to be independent of implying anything about X. Now Calvin Normore points out that Etchemendy's second condition can be weakened by replacing "$=$" by "$\geq$", so that it becomes: your confidence in H would not be weakened by discovering that it implies something about X. Then the explanation theorem takes the following form:

(III) Unless you are antecedently sure that H implies nothing false about X, you will regard H as supported by implying the truth about X if learning that H implies something about X would not make you more doubtful of H.

The proof uses Garber's principle

(K*) $p(A \,\&\, A \vdash B) = p(A \,\&\, B \,\&\, A \vdash B)$.

This principle will hold if "$\vdash$" represents (say) truth-functional entailment and if the person whose belief function is p is alive to the validity of *modus ponens*; but it will also hold under other readings of "$\vdash$," as Garber points out. Thus it will also hold if $A \vdash B$ means that $p(A - B) = 0$, on any adequate interpretation of probabilities of probabilities. The proof also uses the following clarifications of the definitions of T and S:

(T)   For some E, $E \in \mathscr{C}$ and $H \vdash E$ and E is true
(S)   For some E, $E \in \mathscr{C}$ and $H \vdash E$.

Proof of (III). The claim is this:

If $p(S - T) \neq 0$ and $p(HS) \geq p(H)p(S)$ then $p(HT) > p(H)p(T)$.

By (K*), $p(HS) = p(HT)$, so that the second conjunct becomes $p(HT) \geq p(H)p(S)$. With the first conjunct, that implies $p(HT) > p(H)p(T)$ because (since T implies S) $p(S - T) \neq 0$ implies $p(S) > p(T)$.

Note that (III) implies (II), for they have the same conclusion, and the hypotheses of (II) imply those of (III):

(6)   If $p(S) = 1 \neq p(T)$ then $p(S - T) \neq 0$ and $p(HS) \geq p(H)p(S)$

Proof: $p(S - T) \neq 0$ follows from $p(S) = 1 \neq p(T)$ since T implies S, and $p(S) = 1$ implies that $p(HS) = p(H) = p(H)p(S)$.

The explanation theorem (III) goes part way toward addressing the original question, "How are we to explain the supportive effect of belief in E, over and above belief in $H \vdash E$, where H is a hypothesis initially thought especially likely to imply E if true?" Here is a way of getting a bit closer:

(IV) Unless you are antecedently sure that H implies nothing false about X, you take H to be supported more strongly by implying the truth about X than by simply implying *something* about X.

Proof: the claim is that

if $p(S - T) \neq 0$ then $p(H/T) > p(H/S)$,

i.e., since T implies S, that

if $p(S) > p(T)$ then $p(HT)p(S) > p(HS)p(T)$,

i.e., by (K*), that

if $p(S) > p(T)$ then $p(HT)p(S) > p(HT)p(T)$.

But the original question was addressed to belief in a particular member E of $\mathscr{C}$: a particular truth about X, identified (say) by writing out a sentence that expresses it. The remaining gap is easy to close (as David Lewis points out), e.g., as follows.

(7)   For any E, if you are sure that E is about X, implied by H, and true, then you are sure that T is true.

Proof. The claim has the form

and G is some hypothesis you hope to make look good by association with a known truth.

Now (II) is fine as far as it goes, but (John Etchemendy points out) it fails to bear on the case in which it comes as a surprise that H implies *anything* about (say) the tides. The requirement in (II) that p(S) be 1 is not then satisfied, but H may still be supported by implying the true member of $\mathscr{C}$. It needn't be, as the acupuncture example shows, but it may be. For example, if Newton had not realized that H ought to imply the truth about the tides, but had stumbled on the fact that H ⊢ E where E was in $\mathscr{C}$ and known to be true, then H would have been supported by its ability to explain E.

Etchemendy's idea involves the propositions S, T, and

(F)   H implies some false member of $\mathscr{C}$.

Evidently F = S − T, so that −F is the material conditional, −F = −SvT ("If H implies any member of $\mathscr{C}$ then it implies the true one"), and so the condition p(F) = 0 indicates full belief in that conditional. Etchemendy points out that Lewis's conditions in (II) can be weakened to p(F) ≠ 0 and p(HS) = p(H)p(S); i.e., you are not antecedently sure that H implies nothing false about X (about the tides, say), and you take truth of H to be independent of implying anything about X. Now Calvin Normore points out that Etchemendy's second condition can be weakened by replacing "=" by "≥", so that it becomes: your confidence in H would not be weakened by discovering that it implies something about X. Then the explanation theorem takes the following form:

(III)   Unless you are antecedently sure that H implies nothing false about X, you will regard H as supported by implying the truth about X if learning that H implies something about X would not make you more doubtful of H.

The proof uses Garber's principle

(K*)   p(A & A ⊢ B) = p(A & B & A ⊢ B).

This principle will hold if "⊢" represents (say) truth-functional entailment and if the person whose belief function is p is alive to the validity of *modus ponens*; but it will also hold under other readings of "⊢," as Garber points out. Thus it will also hold if A ⊢ B means that p(A − B) = 0, on any adequate interpretation of probabilities of probabilities. The proof also uses the following clarifications of the definitions of T and S:

(T)   For some E, E ε $\mathscr{C}$ and H ⊢ E and E is true
(S)   For some E, E ε $\mathscr{C}$ and H ⊢ E.

Proof of (III). The claim is this:

If p(S − T) ≠ 0 and p(HS) ≥ p(H)p(S) then p(HT) > p(H)p(T).

By (K*), p(HS) = p(HT), so that the second conjunct becomes p(HT) ≥ p(H)p(S). With the first conjunct, that implies p(HT) > p(H)p(T) because (since T implies S) p(S − T) ≠ 0 implies p(S) > p(T).

Note that (III) implies (II), for they have the same conclusion, and the hypotheses of (II) imply those of (III):

(6)   If p(S) = 1 ≠ p(T) then p(S − T) ≠ 0 and p(HS) ≥ p(H)p(S)

Proof: p(S − T) ≠ 0 follows from p(S) = 1 ≠ p(T) since T implies S, and p(S) = 1 implies that p(HS) = p(H) = p(H)p(S).

The explanation theorem (III) goes part way toward addressing the original question, "How are we to explain the supportive effect of belief in E, over and above belief in H ⊢ E, where H is a hypothesis initially thought especially likely to imply E if true?" Here is a way of getting a bit closer:

(IV)   Unless you are antecedently sure that H implies nothing false about X, you take H to be supported more strongly by implying the truth about X than by simply implying *something* about X.

Proof: the claim is that

if p(S − T) ≠ 0 then p(H/T) > p(H/S),

i.e., since T implies S, that

if p(S) > p(T) then p(HT)p(S) > p(HS)p(T),

i.e., by (K*), that

if p(S) > p(T) then p(HT)p(S) > p(HT)p(T).

But the original question was addressed to belief in a particular member E of $\mathscr{C}$: a particular truth about X, identified (say) by writing out a sentence that expresses it. The remaining gap is easy to close (as David Lewis points out), e.g., as follows.

(7)   For any E, if you are sure that E is about X, implied by H, and true, then you are sure that T is true.

Proof. The claim has the form

For any E, if p(Φ) = 1 then p(for some E, Φ) = 1 where Φ is this:

E ε $\mathscr{C}$ and H $\vdash$ E and E is true.

Now the claim follows from this law of the probability calculus

p(X) $\leq$ p(Y) if X implies Y

in view of the fact that Φ implies its existential generalization.

Here is an application of (III):

Since Newton was not antecedently sure that H implied no falsehoods about the tides, and since its implying anything about the tides would not have made it more doubtful in his eyes, he took it to be supported by implying the truth about the tides.

And here is a corresponding application of (7):

Newton came to believe that H implied the truth about the tides when he came to believe that H implied E, for he already regarded E as a truth about the tides.

To couple this with (III), we need not suppose that Newton was antecedently *sure* that H implied something or other about the tides, as in (II). In (III), the condition p(S) = 1 is weakened to p(S) > p(T), which is equivalent to p(S - T) $\neq$ 0, i.e., to p(F) $\neq$ 0.

Observe that in coming to believe T, one also comes to believe S. But if it is appropriate to conditionalize on T in such circumstances, it is not thereby appropriate to conditionalize on S, unless p(S) = p(T), contrary to the hypotheses of (III).

Observe also that although we have been reading "H $\vdash$ E" as "H implies E," we could equally well have read it as "p(E/H) = 1" or as "p(H − E) = 0": (K*) would still hold, and so (III) would still be provable.

4. Probability Logic

Let us focus on the probabilistic counterpart of truth-functional logic. (See Gaifman 1964 and Gaifman and Snir 1982 for the first-order case.)

With de Finetti (1970, 1974) I take expectation to be the basic notion, and I identify propositions with their indicator functions, i.e., instead of taking propositions to be subsets of the set W of all possible "worlds," I take them to be functions that assign the value 1 to worlds where the propositions are true, and 0 where they are false.

Axioms: the expectation operator is
  linear: E(af + bg) = aEf + bEg
  positive: Ef $\geq$ 0 if f $\geq$ 0
  normalized; E1 = 1

("f > 0" means that f(w) > 0 for all w in W, and 1 is the constant function that assigns the value 1 to all w in W.)

Definition: the probability of a proposition A is its expectation, EA, which is also written more familiarly as p(A). De Finetti (1974, section 3.10) proves what he calls "The Fundamental Theorem of Probability":

Given a coherent assignment of probabilities to a finite number of propositions, the probability of any proposition is either determined or can coherently be assigned any value in some closed interval.

(Cf. de Finetti's Lemma, in section 2 above.)

A remarkably tight connection between probability and frequency has already been remarked upon. It is provided by the law of small numbers, i.e., in the present notation,

$$E(A_1 + \ldots + A_n) = p(A_1) + \ldots + p(A_n).$$

That is an immediate consequence of the linearity of E and the definition of "$p(A_i)$" as another name for EA_i. But what has not yet been remarked is the connection between observed and expected frequencies that the law of small numbers provides.

Example: "Singular Predictive Inference," so to speak. You know that there have been s successes in n past trials that you regard as like each other and the upcoming trial in all relevant respects, but you have no information about which particular trials produced the successes. In this textbook case, you are likely to be of a mind to set $p(A_1) = \ldots = p(A_n) = p(A_{n+1}) = x$, say. As $E(A_1 + \ldots + A_n) = s$ because you *know* there were s successes, the law of small numbers yields s = nx. Thus your degree of belief in success on the next trial will equal the observed relative frequency of successes on the past n trials: $p(A_{n+1}) = s/n$.

In the foregoing example, no particular prior probability function was posited. Rather, what was posited was a condition $p(A_i) = x$ for i = 1, . . . , n + 1, on the posterior probability function p: what was posited was a certain probasition, i.e., the domain of the variable "p." The law of small numbers then showed us that for all p in that domain, $p(A_i) = s/n$ for all i = 1, . . . ,n + 1. But of course, p is otherwise undetermined by the condition

of the problem, e.g., there is no telling whether the A_i are independent, or exchangeable, etc., relative to p, if all we know is that p belongs to the probasition $\{p : p(A_1) = \ldots = p(A_n) = p(A_{n+1})\}$.

A further example: *your expectation of the relative frequency of success on the next m trials will equal the observed relative frequency s/n of success on the past n trials in case*

$$(8) \quad p(A_1) = \ldots = p(A_n) = x = p(A_{n+1}) = \ldots = p(A_{n+m}).$$

Proof: as we have just seen, the first part of (8) assures us that $x = s/n$, and by the second part of (8), the law of small numbers yields an expected number of successes on the next m trials of $E(A_{n+1} + \ldots + A_{n+m}) = mx$. Then by linearity of E, the expected relative frequency of success on the next m trials is

$$E(\frac{A_{n+1} + \ldots + A_{n+m}}{m}) = \frac{ms/n}{m} = \frac{s}{n},$$

i.e., the observed relative frequency of success in the first n trials.

What if you happen to have noticed which particular s of the first n trials yielded success? Then the first part of (8) will not hold: $p(A_i)$ will be 0 or 1 for each $i = 1, \ldots, n$. Still, your judgment *might* be that

$$(9) \quad \frac{s}{n} = p(A_{n+1}) = \ldots = p(A_{n+m}),$$

in which case the expected relative frequency of success on the next m trials will again be s/n, the observed relative frequency on the first n. But maybe the pattern of successes on the first n trials rules (9) out, e.g., perhaps your observations have been that $p(A_1) = \ldots = p(A_s) = 1$ but $p(A_{s+1}) = \ldots = p(A_n) = 0$, so that you guess there will be no more successes, or that successes will be rarer now, etc. The cases in which (9) will seem reasonable are likely to be ones in which the pattern of successes on the first n trials exhibits no obvious order.

These applications of the law of small numbers are strikingly un-Bayesian in Clark Glymour's sense of "Bayesian": the results $p(A_{n+1}) = s/n = E(A_{n+1} + \ldots + A_{n+m})/m$ are not arrived at via conditioning (via "Bayes's theorem"), but by other theorems of the calculus of probabilities and expectations, no less Bayesian in my sense of the term.

The emergence of probability in the mid-seventeenth century was part of a general emergence of concepts and theories that made essential use of

(what came to be recognized as) real variables. These theories and concepts were quite alien to ancient thought, in a way in which two-valued logic was not: witness Stoic logic. And today that sort of mathematical probabilistic thinking remains less homely and natural than realistic reasoning from definite hypotheses ("about the outside world") to conclusions that must hold if the hypotheses do. Perhaps "Bayesian" is a misnomer—perhaps one should simply speak of *probability logic* instead. (Certainly "Bayesian *inference*" is a misnomer from my point of view, no less than from de Finetti's and from Carnap's.) But whatever you call it, it is a matter of thinking in terms of estimates (means, expectations) as well as, or often instead of, the items estimated. Thus one reasons about estimates of truth values, i.e., probabilities, in many situations in which the obvious reasoning, in terms of truth values themselves, is unproductive. The steps from two-valued functions (= 0 or 1) to probability functions, and thence to estimates of functions that need not be two-valued brings with it an absurd increase in range and subtlety. To take full advantage of that scope, I think, one must resist the temptation to suppose that a probasition that is not a unit set must be a blurry representation of a sharp state of belief, i.e., one of the probability measures that make up the probasition: an imprecise measurement (specified only within a certain interval) of some precise psychological state. On the contrary, I take the examples of "prevision" via the law of small numbers to illustrate clearly the benefits of the probasitional point of view, in which we reason in terms of a variable "p" that ranges over a probasition R without imagining that there is an unknown true answer to the question, "Which member of R *is* p?"

References

Bolker, Ethan. 1965. *Functions Resembling Quotients of Measures*. Harvard University Ph.D. dissertation (April).

Boole, George. 1854. *The Laws of Thought*. London: Walton and Maberley. Cambridge: Macmillan. Reprinted Open Court, 1940.

Carnap, Rudolf. 1945. On inductive logic. *Philosophy of Science* 12: 72-97.

de Finetti, Bruno. 1937. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7. Translated in Kyburg and Smokler (1980).

——. 1972. *Probability, Induction, and Statistics*. New York: Wiley.

——. 1970. *Teoria delle Probabilità*, Torino: Giulio Einaudi editore s.p.a. Translated: *Theory of Probability*. New York: Wiley, 1974 (vol. 1), 1975 (vol. 2).

Diaconis, Persi and Sandy Zabell. 1982. Updating Subjective Probability. *Journal of the American Statistical Association* 77: 822-30.

Gaifman, Haim. 1964. Concerning Measures on First Order Calculi. *Israel Journal of Mathematics* 2: 1-18.

—, and Snir, Mark. 1982. Probabilities over Rich Languages, Testing, and Randomness. *The Journal of Symbolic Logic* 47: 495-548.

Glymour, Clark. 1980. *Theory and Evidence,* Princeton, Princeton University Press.

Good, I.J. 1952. Rational decisions. *Journal of the Royal Statistical Assn.,* Series B, 14: 107-114.

—. 1962. Subjective Probability as the Measure of a Non-Measurable Set. In *Logic, Methodology and Philosophy of Science,* Ernest Nagel, Patrick Suppes, and Alfred Tarski, Stanford: Stanford University Press. Reprinted in Kyburg and Smokler, 1980, pp. 133-146.

Hacking, Ian. 1967. Slightly More Realistic Personal Probability. *Philosophy of Science* 34: 311-325.

Hailperin, Theodore. 1965. Best Possible Inequalities for the Probability of a Logical Function of Events. *The American Mathematical Monthly* 72: 343-359.

Jeffrey, Richard. 1965. *The Logic of Decision.* New York: McGraw-Hill. University of Chicago Press, 1983.

—. 1968. Probable knowledge. In *The Problem of Inductive Logic,* ed. Imre Lakatos. Amsterdam: North-Holland, 1968, pp. 166-180. Reprinted in Kyburg and Smokler (1980), pp. 225-238.

—. 1970. Dracula Meets Wolfman: Acceptance vs. Partial Belief. In *Induction, Acceptance, and Rational Belief,* ed. Marshall Swain, pp. 157-185. Dordrecht: Reidel.

—. 1975. Carnap's empiricism. In *Induction, Probability, and Confirmation,* ed. Grover Maxwell and Robert M. Anderson, pp. 37-49. Minneapolis: University of Minnesota Press.

Keynes, J.M. 1921. *A Treatise on Probability,* Macmillan.

Koopman, B.O. 1940. The Axioms and Algebra of Intuitive Probability. *Annals of Mathematics* 41: 269-292.

Kyburg, Henry E., Jr. and Smokler, Howard E., eds. *Studies in Subjective Probability,* 2nd edition, Huntington, N.Y.: Krieger.

Levi, Isaac. 1974. On Indeterminate Probabilities. *The Journal of Philosophy* 71: 391-418.

—. 1980. *The Enterpirse of Knowledge.* Cambridge, Mass.: MIT Press.

May, Sherry and William Harper. 1976. Toward an Optimization Procedure for Applying Minimum Change Principles in Probability Kinematics. In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science,* ed. W.L. Harper and C. A. Hooker, volume 1. Dordrecht: Reidel.

Mellor, D.H., ed. 1980. *Prospects for Pragmatism.* Cambridge: Cambridge University Press.

Newton, Isaac. 1934. *Principia* (Volume 2, including *The System of the World*). Motte/Cajori translation. Berkeley: University of California Press.

Ramsey, Frank. 1931. *The Foundations of Mathematics and Other Logical Essays.* London: Kegan Paul. Also: *Foundations,* Cambridge U.P., 1978.

Savage, Leonard J. 1954. *The Foundations of Statistics,* New York: Wiley. Dover reprint, 1972.

Skyrms, Brian. 1980. Higher Order Degrees Of Belief. In Mellor (1980).

Wald, Abraham. 1950. *Statistical Decision Functions,* New York: Wiley.

Williams, P.M. 1980. Bayesian Conditionalization and the Principle of Minimum Information. *British Journal for the Philosophy of Science* 31.

—. 1976. Indeterminate Probabilities. In *Formal Methods in the Methodology of Empirical Sciences,* Proceedings of the Conference for Formal Methods in the Methodology of Empirical Sciences, Warsaw, June 17-21, 1974, ed. Marian Przetecki, Klemena Szaniawski, and Ryszand Wojcick.

— Brian Skyrms —

Three Ways to Give a Probability Assignment a Memory

Consider a model of learning in which we update our probability assignments by conditionalization; i.e., upon learning S, the probability of not-S is set at zero and the probabilities of statements entailing S are increased by a factor of one over the initial probability of S. In such a model, there is a certain peculiar sense in which we lose information every time we learn something. That is, we lose information concerning the initial relative probabilities of statements not entailing S.

The loss makes itself felt in various ways. Suppose that learning is meant to be corrigible. After conditionalizing on S, one might wish to be able to decide that this was an error and "deconditionalize." This is impossible if the requisite information has been lost. The missing information may also have other theoretical uses; e.g., in giving an account of the warranted assertability of subjunctive conditionals (Adams 1975, 1976; Skyrms 1980, 1981) or in giving an explication of "evidence E supports hypothesis H" (see the "paradox of old evidence" in Glymour 1980).

It is therefore of some interest to consider the ways in which probability assignments can be given a memory. Here are three of them.

I. *Make Like an Ordinal* (Tait's Suggestion)[1]: A probability assignment will now assign each proposition (measurable set) an ordered pair instead of a single number. The second member of the ordered pair will be the probability; the first member will be the memory. To make the memory work properly, we augment the rule of conditionalization. Upon learning P, we put the current assignment into memory, and put the result of conditionalizing on P as the second component of the ordered pairs in the new distribution. That is, if the pair assigned to a proposition by the initial distribution is (x, y), then the pair assigned by the final distribution is ((x, y), z), where z is the final probability of that proposition gotten by conditionalizing on P. (If P has initial probability zero, we go to the closest state in