THE JOURNAL OF PHILOSOPHY

VOLUME LXXXI, NO. 5, MAY 1984

BELIEF AND THE WILL*

CAN we rationally come to believe a proposition that is entailed neither by those we have believed heretofore nor by our previous opinions conjoined to the evidence before us? Discussing this question, William James quoted W. K. Clifford's statement (in "Ethics of Belief") that it is wrong always, everywhere, and for everyone to believe anything on insufficient evidence.[1] Arguing against this, James claimed that, in forming beliefs, we pursue two aims: to believe truth and to avoid error, and argued that the extent to which we pursue either at the cost of the other is a matter of choice: "he who says 'Better go without belief forever than believe a lie!' merely shows his own preponderant private horror of becoming a dupe. He may be critical of many of his desires and fears, but this fear he slavishly obeys . . . a certain lightness of heart seems healthier than this excessive nervousness [about error]. At any rate, it seems the fittest thing for the empiricist philosopher."[2]

In philosophy of science, until recently, something of this sort was regarded as part of the received view: general theories, such as Darwin's, Einstein's, or Bohr's, cannot be established on the basis

* The author wishes to thank the National Science Foundation and Princeton University for support of his research and sabbatical leave, and the participants of Richard Jeffrey's seminar (especially David Lewis) for much helpful discussion of a preliminary draft of sections II and III, circulated under the title "A Puzzle for Both True and Partial Believers" in November, 1982. Thanks is also due for help to improve this paper, to Nancy Cartwright, Roger Cooke, Paul Fitzgerald, William Harper, and Zeno Swijtinga, and especially to Isaac Levi and Brian Skyrms who prepared detailed commentaries.

[1] "The Will to Believe." Page references will be to his *Essays in Pragmatism* (New York: Hafner, 1948).

[2] *Op. cit.* p. 100. Note that on the next page James grants that scientists doing science proceed as Clifford has it. This concession may have been for the sake of argument (for compare the skepticism about the reach of science on pages 23, 25, and 38), rather than a genuine subscription to the objectivity of strict induction from the evidence. Recent philosophy of science has in any case not been so sanguine.

0022-362X/84/8105/0235$02.20

of the evidence, but we may rationally come to believe that they are true. In addition, what we take as evidence itself is not indubitable, and we may later come to regard it as having been false. We regard ourselves as infallible neither with respect to what we take as evidence nor with respect to our extrapolation beyond the evidence, but neither do we think ourselves irrational for engaging in this cognitive enterprise.

The situation is prima facie not affected by the replacement of undogmatic full belief by gradations of partial belief. Perhaps when I profess belief or acceptance, I merely indicate that the proposition seems highly likely to me. But the evidence at hand, especially if itself not fully believed, plus our opinions heretofore, generally do not entail a high probability of truth for general hypotheses or theories—especially not for the sort studied by scientists, which have empirical consequences for all past and future. Only recently have these views come under attack, by writers inspired by Bayesian foundations of statistics.

My strategy in this paper will be first to submit the traditional epistemological views to a critique along Bayesian lines (without claiming to be a Bayesian of any sort). Then I shall show the implications of that critique for those ways of changing one's opinions which Bayesians have generally admitted as rational. The result will be, I think, a puzzle for all concerned. Indeed, this puzzle suggests that we must obey a principle (which I shall call *Reflection*), going beyond the simple laws of probability, which looks prima facie quite unacceptable. I selected James's essay to introduce the topic because I wish to propose a solution to the puzzle along the broadly voluntarist lines of the views he defended. I hope that by consistently carrying through the voluntarist point of view we can, without sacrificing the theory of personal probability as a logic of epistemic judgment, nevertheless maintain the traditional epistemology.[3]

I. TO BELIEVE A THEORY

Imagine that today I do not profess total certainty about whether the basic theory of evolution is true nor about whether I shall be sure of its truth next year. It does seem quite possible to me that I

[3] James's view may be attacked on the flank by arguing that belief is not a matter of the will at all, not under voluntary control. Voluntarism with respect to belief is usually attacked in its naive versions and defended in more sophisticated formulation; I will of course not suggest that we can believe just any proposition at will. Cf. James, *op. cit.* p. 90; Barbara Winters, "Believing at Will," this JOURNAL, LXXVI, 5 (May 1979): 243-256; and Robert Holyer, "Belief and Will Revisited," *Dialogue*, XXII, 2 (June 1983): 273-290.

shall become sure of its truth, but also, unfortunately, somewhat possible that I shall form this belief although it is in fact false. Does the state of opinion I have just described seem totally absurd or irrational to you? If not, this section may convince you otherwise.

The critique I am about to offer is along Bayesian lines, though not exactly standard ones, nor perhaps uncontroversial. I request the reader to bear with my rather informal and naive presentation here; in the next section I shall make the argument at once more general and more precise. As described, my present state of opinion is one of uncertainty. The degrees of uncertainty about the different propositions are not the same; it is common today to describe them in terms of subjective or personal probability.[4] In Bayesian eyes, personal probability is the guide to life. The simplest cases we find are in buying contracts, insurance policies, and wagers. Without going into the details, I shall take the following as paradigm: if a contract is worth 1 to me if A be the case, and nothing otherwise, then its present value for me equals my personal probability for A. More generally,

> if it is worth z to me if A and nothing otherwise and if my personal probability that A is the case equals $P(A)$, then the value of this contract for me (fair in buying or selling) equals $zP(A)$.

That is all we shall need for our discussion.

So let H be the hypothesis under discussion—say, the theory of evolution—and let E be the proposition that Bas van Fraassen will fully believe that H (say, one year from today). For definiteness, suppose that $P(E)$—my degree of belief that E will be the case—equals 0.4 and $P(\sim H \& E)$—my degree of belief that I will mistakenly come to bestow full belief on H—equals 0.2. For now I shall assume that full belief entails personal probability equal to 1. The argument would go through for a degree very close to 1 as well, but I shall in any case consider more explicitly the case of non-full-belief formation below.

At this point we may introduce into the story a Dutch bookie.[5]

[4] Some common objections, such as that we do not have numerically precise degrees of certainty and uncertainty, are, I think, easily met (see, further, section v below). But if the reader is willing to conclude that it is the idea of subjective probability that is at fault, he does not need my present defense of traditional epistemology.

[5] This term is a reference to the so-called "Dutch book theorem." The usual or synchronic Dutch book argument establishes the obedience of degrees of belief to the probability calculus as a criterion of rationality ("coherence") for one's state of opinion at a single time. The betting scheme I am about to describe is part of David

He elicits all the above information from me, and he decides on a secret strategy for betting with me. As a first step, he offers me three bets. I call him Dutch, because what he has offered me is what is called a *Dutch book*, a set of bets such that, no matter what happens, I will lose money. And the unfortunate fact is that each of the bets is fair, according to my own state of opinion.

Because I will describe his betting scheme in full generality in the next section, I ask the reader to consider the present figures only cursorily. The trick up his sleeve is that (a) if I do not come to fully believe H, I win only the second bet, and (b) if I do come to fully believe H, then I lose the second bet, but I also tell the bookie myself that I have lost the first bet as well. At this point he takes the second step in his strategy, which is in effect to buy back the ticket for the first bet, for a pittance. (He can do this by formally offering to buy from me a bet that H is false; since I am sure that H is true, any price at all for that new bet will be more than fair in my opinion.) In either case I will have a net loss.

Here are the bets: the first pays 1 if I come to believe H and H is really false—he asks 0.2 for it. The second will pay 0.5 if I do not come to believe H, and he asks me 0.3 for that one. The third pays 0.5 if I really do come to believe H; that one costs 0.2. All these prices are fair, given my state of opinion. (I leave out units of value; so they can be adjusted for inflation and the like.) None of the bets pay anything if they are not won. My total cost is 0.7 for all three.

On one scenario I do not embrace H; I win the second bet and lose the other two. On the other scenario I do embrace the hypothesis; now I lose the second bet, tell him myself that H is true, so I get nothing for the first bet (though I receive a pittance when I sell him back a bet on $\sim H$ for next to nothing), and I win the third. On either scenario I get at most a little more than 0.5, and I have a net loss. This bookie had a strategy which he knew beforehand would allow him to offer me only bets that would be fair by my lights, and yet necessarily give him a net profit. He devised this strategy without any special knowledge either of whether Darwin was right or of whether I would come to believe that hypothesis.

All this may look like so much *léger-de-main* at this point. Sup-

---

Lewis's diachronic Dutch book argument to justify conditionalization as the correct rule for transforming prior into posterior degrees of belief [see P. Teller, "Conditionalization, Observation, and Change of Preference" in W. L. Harper and C. A. Hooker, *Foundations of Probability Theory*, vol. 1 (Boston: Reidel, 1976)]. Bayes himself had given a similar argument, and a more sophisticated theorem has been proved by Glen Shafer; see his "Bayes' Two Arguments for Conditioning," *Annals of Statistics*, x (1982): 1075-1089, and "A Subjective Approach to Conditional Probability" *Journal of Philosophical Logic*, xii, 3 (November 1983): 453-466.

pose for a moment, however, that I have not pulled any tricks. In that case whoever is as I described myself, hypothetically, at the beginning of this section, is in a state of opinion which the Bayesian calls incoherent (a polite word for irrational). Whether or not I actually bought the bets does not matter, of course: my incoherence consists in regarding them as fair.

Unhappy mortal! I found myself in this incoherence merely by contemplating that I could do what James said I could—without even actually deciding to believe Darwin's theories, or anything like it. Not only people so rash as actually to come to believe theories on less than totally compelling evidence, but anyone who does not, with Clifford, reject such a new belief as utterly irrational, is caught in the trap. Who, upon seeing this Bayesian refutation, does not immediately find himself in full flight from voluntarism and pragmatism, toward the imitation of Carnap's robot?

II. TO RAISE ONE'S OPINION OF A MATTER OF FACT

The preceding argument gives rise to three initial suspicions. The first is that bets cannot sensibly be made on propositions, like Darwin's hypothesis, which cannot be verified or falsified in a finite amount of time. The second is that it is irrational to become fully certain of any propositions except tautologies. The third suspicion one may have is that it is irrational to change one's mind in any way except by what the Bayesians call "conditionalization on one's evidence." [Roughly speaking, this means that one becomes fully certain of the proposition(s) one takes as evidence, and makes only the minimal adjustments to the rest of one's opinions needed to accommodate this new certainty. We may think of this as Clifford's position, updated to accommodate degrees of belief.] Note well that the second and third suspicion cannot be jointly entertained unless evidence is always tautological. So we must confront the second and third separately, but I think we can show the irrelevance of the first along the way. Later on we shall turn to still further suspicions, for example, about the suitability of one's own future opinions as a subject for prevision.

Before going on to examples, we should look at what exactly is involved in Dutch book arguments. In the simple or *synchronic* case, the bookie is able (without having knowledge superior to the agent's) to offer the agent several bets, which demonstrably have the following features: (a) each bet taken individually looks fair to the agent at this time, and (b) taken together the bets are such that, no matter what happens, the agent will suffer a net loss. The expression 'looks fair' is explicated by the Bayesians in terms of the agent's personal probability P and utility evaluations, following the paradigm that $z\mathrm{P}(A)$ is the exact value of a bet on proposition

A with payoff z. In the case described, the bets in question constitute a Dutch book, and the agent's vulnerability brands his state of opinion as *incoherent* (and indeed, it can be deduced that P violates the probability calculus).

In the *diachronic* case we should speak of a *Dutch strategy* rather than a Dutch book. The bookie is able (without superior knowledge of present or later circumstances) to devise a strategy for offering bets to the agent which is demonstrably to the agent's disadvantage. This strategy is demonstrably such that, under all eventualities, the agent will be offered bets with two features: (a) individually, each bet will look fair to the agent at the time of the offer, and (b) taken together, the bets offered will be such that, whatever happens, the agent will suffer a net loss. Let us emphasize especially that these features are demonstrable *beforehand,* without appeal to any but logical considerations, and the strategy's implementation requires no information inaccessible to the agent himself. The general conclusion must be that an agent vulnerable to such a Dutch strategy has an initial state of opinion or practice of changing his opinion, which together constitute a demonstrably bad guide to life. In this paper, success of the strategies discussed will be independent of the agent's practices for changing opinion, and hence any blame must attach to his initial state of opinion—his vulnerability reveals an initial incoherence.

It is now time to describe the exact betting strategy used by our Dutch bookie. We have two propositions, H (the hypothesis) and E, a proposition about the customer's future attitude to the hypothesis. The customer has degrees of belief $\mathrm{P}(E)$ and $\mathrm{P}(\sim H\&E)$, neither of which is 0 or 1. The three bets are:

(I) The bet which pays 1 if $(\sim H\&E)$ and which costs $\mathrm{P}(\sim H\&E)$
(II) The bet which pays x if $\sim E$ and which costs $x\mathrm{P}(\sim E)$
(III) The bet which pays y if E and which costs $y\mathrm{P}(E)$

Here the probability of $\sim E$ equals 1 minus the probability of E. The number x is the usual conditional probability of $\sim H$ given E; that is, $\mathrm{P}(\sim H\&E) \div \mathrm{P}(E)$. And finally y is x minus the subjective probability the customer will have for the hypothesis, when and if E becomes true. It helps to observe that I and II together form in effect a *conditional bet* on $\sim H$ on the supposition that E, which bears the cost x and has prize 1, with the guarantee of your money back should the supposition turn out to be false.[6] So the total cost of all the bets together must equal $x + y\mathrm{P}(E)$.

[6] To see what the total cost is of I and II together, calculate

$$\mathrm{P}(\sim H\&E) + x\mathrm{P}(\sim E) = \mathrm{P}(\sim H \mid E)\mathrm{P}(E) + x\mathrm{P}(\sim E) = x(\mathrm{P}(E) + \mathrm{P}(\sim E)) = x$$

Let us now consider an example in which all propositions will have their truth value settled by a certain definite time and in which it is not strictly implied that anyone is fully certain of the truth of any nontautology. Since we are now on the attack, the example should be made as simple and hygienic as possible. Let it be a race, at Hollywood Park, tomorrow at noon. The proposition H is that the horse Table Hands will run in that race and win it. The bookie now asks me seriously to consider the possibility that tomorrow morning, at 8 A.M., I shall consider fair a bet on this proposition at odds 2 to 1. I say I do not know if that will happen—my personal probability for that eventuality, call it E, is $P(E) = 0.4$. Next he elicits my opinion about how reliable I think I am as a handicapper of horses. What is my subjective probability that E will indeed be true but that the hypothesis that Table Hands will win, is false? Suppose I answer that this degree of belief of mine, $P(\sim H \& E)$, equals 0.3. The exact numbers do not matter here too much, except that they indicate a certain lack of confidence in my own handicapping skill. In this case they entail that my present conditional probability for Table Hands' winning, on the supposition that tomorrow morning I will have subjective probability 1/3 for it, is only 1/4. The calculation is simple.[7]

What the bookie does now, if I buy the bets, is also simple. He approaches me at 8 A.M. the next morning. If I do not consider odds of 2 to 1 on Table Hands fair, he pays me off on the second bet, but he has won (I) and (III). On the other hand, if I do call those odds fair, he first of all pays me for bet (III). But then he buys from me a bet, with prize 1, against Table Hands' winning, at my newly announced odds. The result of this is, of course, that whether or not Table Hands wins at noon, no money need change hands between us—he has, so to say, bought (I) back from me. So we can now tally up our prospective losses and gains, and again it turns out that I shall have been the loser come what may.[8]

I chose this example to disarm both the first two initial suppositions at once. For there is no implication, in the description of this case, that anyone ever raises the probability of any nontautology to *one* (though in that case the bookie is being quite agreeable about

[7] E implies that my probability for H tomorrow morning will be 1/3, and so my probability for $\sim H$ then is 2/3. We have $x = 0.3 \div 0.4 = 3/4$ and $y = x - 2/3 = 1/12$. The costs of the bets are 0.3 for (I), $x(1 - P(E)) = (3/4)(0.6) = 0.45$ for bet (II), and $yP(E) = (1/12)(0.4) = (1/30)$ for (III), for a total cost of $(3/4) + (1/30)$.

[8] From footnote 7 we know that the initial total cost was $(3/4) + (1/30)$. If E is false, I collect only $x = (3/4)$. If E is true, I collect 1/12 on the third bet, but then I receive in addition only what I then consider a fair price for the bet against Table Hands' winning, namely 2/3; so my total return equals 3/4 again.

paying me off before he is totally certain that he has heard me correctly). On the other hand, every proposition becomes settled in a certain finite amount of time. The disaster—which consists of course in my present vulnerability to his strategy, not in any actual bets made or lost—happened again because I profess some doubts today about my judgment of tomorrow.

Let us therefore not think about gambling anymore, and turn to the scientist in his lair, Clifford's ideal who (according to James's quotation) "will guard the purity of his belief with a very fanaticism of jealous care, lest at any time it should rest on an unworthy object, and catch a stain which can never be wiped away" and who, therefore, never believes anything upon insufficient evidence (James, *op. cit.*, p. 92). He is then just like Carnap's robot: his senses bring him propositions that he takes as evidence, and his total response to this consists in *conditionalizing* his present state of opinion on these propositions.[9] To conditionalize on a proposition X taken as evidence means this: your odds for various eventualities on the supposition that X are still the same, but that supposition you now regard as certainly true.

Well suppose that e is the sort of proposition that I typically do take as evidence. We need not decide here exactly what sort that is. Perhaps it is the sort of report that comes from Mount Wilson observatory, after having been checked and verified numerous times. Or perhaps it is simple everyday propositions like "That rose is red" or "That is a rose." In any of these cases, the example is decided on the basis of perception. Now let me give the reins over to you, reader: do you think that I am infallible when it comes to perception? Do you think that I shall certainly not take a rose to be red if it is not? Or that a needle will never turn out to have been to the left of the number 7 on a dial, when I said it was to the right? All right, you have convinced me: my subjective probability that e is false, on the supposition that I shall take it as evidence, is not zero.

It is not difficult to see that, formally speaking, I am now in exactly the same position as I was when I thought that I might come to believe a false hypothesis of Darwin's. (Let E be the proposition that I shall take e as evidence, and H the hypothesis that e is true.) Merely by contemplating this eventuality and admitting that I am not sure it cannot happen, I imply that I regard as fair each of three bets which together form the basis for a Dutch strategy. Even if I insist that my epistemic life is lived in the Imitation of Car-

[9] When conditionalized on A, the function P becomes the function P′ such that $P'(X) = P(X \mid A) = P(X \& A) \div P(A)$ for all propositions X. This can be done only if $P(A)$ is not zero.

nap's Robot, mere admission of my fallibility, it seems, makes me diachronically incoherent.

III. PREVISION OF OUR OWN PREVISIONS

When we begin to think about the laws and sources of our own epistemic judgments and states of opinion, we are automatically led to deal with them as facts in the world and to consider them in general: that is, with no regard to persons, treating others' no differently from our own. Yet a closer reading of the preceding arguments, once the initial suspicions have been disarmed, presents us with only two possible ways out. The first is that we should have no opinion at all concerning the reliability of our own future judgments; the second, to form as a matter of principle an exceptionally high opinion of their reliability in our own case.[10]

The first may claim precedent in the discussions of de Finetti and Savage themselves, rejecting the intelligibility of higher-order degrees of belief. Their reasons have been incisively criticized by Brian Skyrms.[11] As I shall explain later, I think there is something to the view that the statement that my opinion is such and such "is not a proposition." But we can, I think, quickly dismiss the simpler objections along this line. First of all, whatever is done by the person who says "It seems as likely to me as not that today will be rainy," we do have a proposition that is true if and only if he is at the moment in the psychological state of considering rain as likely as not, being as willing to bet on rain as on the toss of a coin, and so forth. Psychological studies of this subject are well known and we do not think them, surely, to be of an illusory or nonexistent phenomenon.

More important is the worry that, in asking us to consider our own states of opinion, we may be led into the vagaries and paradoxes of self-reference. It would be no surprise if the attempt to assign degrees of credence or credibility to self-referential statements generally were as beset with paradox as the attempt to assign them all truth values. But actually the puzzles or arguments I have presented do not presuppose that degrees of belief are accorded to self-referential statements at all. Suppose that "Cicero" and "D-Day"

[10] A third possibility was advocated in discussion by David Lewis: that the standard of rationality exemplified by Dutch-book Invulnerability applies to a certain sort of ideally rational agent, who not only believes himself to be, but is, infallible with respect to perception, and which we explicitly realize ourselves not to be. But this leaves us still with the task of constructing an epistemological theory that does apply to our own case.

[11] "Higher Order Degrees of Belief" in D. H. Mellor, ed. *Prospects for Pragmatism: Essays in Honour of F. P. Ramsey* (New York: Cambridge, 1980), pp. 109–137, and Appendix 2 of his *Causal Necessity* (New Haven, Conn.: Yale, 1980).

are context-independent rigid designators referring to a person and a time, respectively, and that p is a function defined on some set of propositions such that $p(A) = r$ if and only if Cicero has on D-Day subjective probability r for proposition A. For definiteness, suppose that the domain of p contains only propositions of an extremely simple sort, such as that Table Hands wins the race or that a certain coin lands heads up or that a certain rose is red. There can surely be no difficulty in anyone's having at any time a degree of belief for the proposition that $p(A) = r$. Hence Cicero may have exactly that the day before D-Day. In addition, there is (independent of these considerations) surely no problem about Cicero's being able to know that he is Cicero or to know that the day in question is in fact the day before D-Day. If there are difficulties with any of these suppositions, they must be deep skeptical problems concerning the very coherence (in the nontechnical sense) of the concept of subjective probability and the concept of knowledge about who we are and what time it is. This coherence is all our arguments required. At no point did we need to assume that anyone's degrees of belief were accorded to any but time- and context-independent propositions.

We come therefore finally to the last way out, which is to say that all three examples were cases in which I made the agent out to be genuinely irrational. This could only be because in each case his degree of belief about what would happen, on the supposition that he would have a certain opinion about that in the future, differed from *that* opinion. The principle we are thereby led to postulate as a new requirement of rationality, in addition to the usual laws of probability calculation is this:

$$\text{(Reflection)}\quad \mathrm{P}_t^a(A \mid p_{t+x}^a(A) = r) = r$$

Here P_t^a is the agent a's credence function at time t, x is any nonnegative number, and $(p_{t+x}^a(A) = r)$ is the proposition that at time $t + x$, the agent a will bestow degree r of credence on the proposition A. To satisfy the principle, the agent's present subjective probability for proposition A, on the supposition that his subjective probability for this proposition will equal r at some later time, must equal this same number r. It is tempting to call this principle of reflection by some more memorable name, such as 'Self-confidence', 'Optimism', or perhaps 'EST', or even 'Self-deception', but I have chosen a more neutral name because I propose to examine, and indeed advocate, serious attempts to defend the principle.[12]

[12] In Skyrms's article the synchronic form ($x = 0$) is advocated; the discussion contains diachronic examples as well, but they concern the supposition that the agent *learns* his posterior credence, whereupon the synchronic form applies.

Since none of us is willing to adopt a similar principle governing our own opinion concerning the reliability of others' opinions—or the corollary that they will never take as evidence something that is in fact false—justification of this principle can follow no ordinary route! Indeed, it would seem that we already believe that most people whose credence function obeys this principle of Reflection are by that very fact mistaken about themselves.

At the same time we can give independent or indirect reasons to think that criteria of coherence, concerning degrees of belief that are guides for action, will require this Reflection principle for their satisfaction. To show this I must first briefly outline another justification for the additivity of synchronic degrees of belief, a sort of dual to the Dutch book argument.[13]

To explain the idea of *calibration,* consider a weather forecaster who says in the morning that the probability of rain equals 0.8. That day it either rains or does not. How good a forecaster is he? Clearly to evaluate him we must look at his performance over a longer period of time. Calibration is a measure of agreement between judgments and actual frequencies. Without going into detail, it is still easy to explain perfect calibration. This forecaster was perfectly calibrated over the past year, for example, if, for every number r, the proportion of rainy days among those days on which he announced probability r for rain, equalled r.

Although perfect calibration may not be a reasonable aim by itself, and hardly to be expected at the best of times, it certainly looks like a virtue. It would seem to be irrational to organize your degrees of belief in such a way as to ruin, a priori, the possibility of perfect calibration. A few qualifications must at once be introduced: this forecaster would not have been perfectly calibrated over the past year if he had announced irrational numbers, or even numbers not equal to some fraction of form $x/365$. So the only possibility that we should require him not to ruin beforehand is that of arbitrarily close approximation to perfect calibration if he were asked sufficiently often about events that he considers exactly similar to those he was actually asked about during the evaluation period. It can now be proved that satisfaction of this criterion of potential perfect calibration is exactly equivalent to satisfaction of the probability calculus (in exactly the same sense that this equiv-

[13] See my "Calibration: A Frequentist Justification of Personal Probability," in L. Laudan and R. Cohen, eds., *Philosophy, Physics, Psychoanalysis* (Boston: Reidel, 1983). Please note well that calibration by itself is not a good scoring rule, and the criterion explained below does not entail that better calibration is always better *tout court.*

alence can be claimed for the criterion of invulnerability to Dutch books).

But it is easy to see what will happen if the evaluation is extended to the forecaster's opinions concerning the calibration of his own judgments. For suppose that he is actually perfectly calibrated in his judgments concerning rain over the next year. Then if he has made judgments to the effect that there will be a discrepancy between the actual frequencies and his announced probabilities, *those* judgements will not be perfectly calibrated. Hence by adding such a judgment as "The probability of rain on days on which I announce the probability of rain to be 0.8, equals 0.7" he would automatically ensure that the class of all his judgments was not perfectly calibrated on any possible scenario. Our criterion accordingly appears to require him to express perfect confidence in the calibration of his own judgments.

Dutch book considerations are of course more familiar; it is interesting to see that the principle of Reflection follows as an immediate corollary to this equivalent, less familiar criterion of coherence. It helps to dispel as vain the small hope that criteria of rationality of this general sort could be satisfied by anyone with doubts that violate Reflection. Yet—and here is the puzzle—we all begin with the intuition that such doubts are not of the radically skeptical kind, but reasonable and rightly common.

IV. CIRCUMVENTING MOORE'S PARADOX

The main purpose of this section will be to show that certain attempts to defend the principle of Reflection do not work. But at the same time I will attempt to show that even an agent adhering to that principle may have some way to express doubt about the reliability of his own future opinions. Hence the discussion will at least undermine one objection to the principle, even if it does not yet issue in a good defense.

The first proposal to defend (Reflection) is this: to announce my subjective conditional probability for X, on the supposition that Y, is simply to announce what my opinion concerning X would be, should I learn that Y. This thesis implies (Reflection) at once, but the thesis is quite untenable. Richmond Thomason once objected to a similar theory of what it was to believe a conditional, that he believed to be true the proposition that, if his wife were not faithful to him (she being so clever), he would believe that she was. If I go on to reflect on other examples, it is only because I wish to do more than defeat the proposal.[14]

[14] It would not help to say that $P(A \mid B)$ is the probability that A would have for me if B were to become my total new evidence, just because that would tell us noth-

Are there propositions that we must admit to be possibly true but could never believe? Hilary Putnam has argued this status for the proposition that we are brains in a vat, and Donald Davidson for the proposition that most of our beliefs are false. These are forms of general and radical skepticism. An older and simpler case is Moore's paradox: "There is a goldfinch in the garden and I do not believe that there is." This statement could of course be true (at the moment I do not believe that there is, yet there might be one) but I could not very well assert it, for this is not a proposition that I can believe. Note, however, that I have just stated parenthetically that, for all I know or believe, it may be true; so I clearly do not *dis*believe it. It is also to be remarked that Moore's paradox does not presuppose that belief is a propositional attitude that we can have toward self-referential propositions. For if Cicero knows himself to be Cicero, he cannot believe that (there is a goldfinch and Cicero does not believe that there is). To consider a somewhat more general version, we must introduce the distinction between probabilities as gradations of belief and as degrees of objective chance.

This distinction is now commonly made, and several recent papers have been devoted to the principles governing their combination.[15] The minimal such principle looks formally similar to the synchronic version of (Reflection):

$$\text{(Miller)}\quad P_t(A \mid \mathrm{ch}_t(A) = r) = r$$

so called because of its role in the (famous but fallacious) argument known as Miller's paradox.[16] To satisfy this principle, the agent's subjective probability for a proposition A, on the supposition that the objective chance that A equals r, must be equal to that same number r. Justification of this principle certainly rests on nontrivial assumptions about what we are like—namely, that we are temporal and finite beings, aware of our temporality and finitude. To see this we deduce that, for an agent whose epistemic history satisfies (Miller), perfect foreknowledge is incompatible with indeter-

---

ing about what $P(A \mid B)$ is when B is not the sort of proposition that could be one's total new evidence. Instead I interpret conditional probability in a way that has no logical connection with learning. To say that $P(A) = 2/3$ is to say that, to me, A is twice as likely to be the case as not—this re-expresses the opinion in terms of personal odds for A as against $\sim A$. Similarly, $P(A \mid B) = 2/3$ expresses my personal odds for $(A \& B)$ as against $(\sim A \& B)$.

[15] See my "A Temporal Framework for Conditionals and Chance," *Philosophical Review*, LXXXIX, 1 (January 1980), 91-108, and reprinted in W. L. Harper, *Ifs* (Boston: Reidel, 1981); and David Lewis, "A Subjectivist's Guide to Objective Chance," *ibid.*, pp. 267-298.

[16] See Richard Jeffrey's review of articles by David Miller *et al.*, *Journal of Symbolic Logic*, XXXV, 1 (March 1970): 124-127.

minism. For suppose that such an agent had subjective probability P equalling 1 or 0 for every factual proposition, and indeed, 1 exactly if the proposition is true. Then there is, for each factual proposition A, a number r such that $\mathrm{ch}_t(A) = r$ and $\mathrm{P}_t(\mathrm{ch}_t(A) = r) = 1$. Hence also $\mathrm{P}_t(A) = r$, by (Miller); but then it follows that r is 0 or 1; so whether or not A will be the case is already determined with certainty by the facts at this time.[17]

If we add to (Miller) the synchronic—I should think, uncontroversial—part of (Reflection) we can now find a proposition which I can admit to be quite possibly true but which I know I could never fully believe. Suppose I have a coin in my hand which I am about to toss and I have picked it at random from a box that contained one fair coin and one magician's coin, the latter having a two-to-one chance of landing heads up. My present subjective probability for the coin in my hand to land heads up is, accordingly, the average of the two objective chances, $1/2(1/2 + 2/3) = 7/12$. So my present subjective probability for the proposition (the chance of heads equals $1/2$ and my personal probability for it equals $7/12$) equals $1/2$. But of course I could never fully believe that conjunction; for, by (Miller), if I fully believed the first conjunct, my personal probability would automatically equal $1/2$ too. [More rigorously: (Miller) and the synchronic ($x = 0$) part of (Reflection) together entail that if $\mathrm{P}_t(\mathrm{ch}_t(A) = r \,\&\, p_t(A) = s) = 1$ then $r = s$.]

So now we have found a proposition Y to which we can indeed assign a positive subjective probability, but which we cannot conditionalize on. Hence it is clear that $\mathrm{P}(X \mid Y)$ is not to be thought of as the probability we would accord X should we learn that Y. The proposal for defending (Reflection) made at the beginning of this section has failed. But we have learned something useful. Even while adhering to (Reflection) we can to some extent express doubts about the correctness or reliability of our future opinions. For example, without violating (Reflection) I can say: "It does not seem unlikely to me that Table Hands' objective chance of winning tomorrow will be considerably less than my subjective probability for that event tomorrow morning."

Those who believe that we conditionalize on—hence raise to subjective certainty—propositions that we take as evidence, do not

[17] If we generalize (Miller) to $\mathrm{P}_t(A \mid \mathrm{ch}_{t+x}(A) = r) = r$, then we can derive the stronger result that if the truth value of A becomes settled at time $t + x$ [this truth value then equals $\mathrm{ch}_{t+x}(A)$, and must be 0 or 1], the agent cannot at t believe with certainty that the present chance of A is something different from 0 or 1 if he also believes that A will be true (respectively false) at its settling time ("there are no crystal balls").

have this sort of consolation. For presumably we mean to take as evidence at $t + x$ only propositions A whose truth value becomes settled at or by that time, which implies that A is equivalent to $\mathrm{ch}_{t+x}(A) = 1$. To say, therefore, that it is not totally unlikely that tomorrow morning I shall take A as evidence even though its chance is less than 1 is to violate (Reflection) by implication. A simple one-place probability function will never allow us to characterize the epistemic state of someone who says that he may become certain of a proposition but will not reject as absurd the possibility that future evidence will prove him wrong.[18] But it remains that in the preceding paragraphs we have seen considerable leeway for the person who wishes to be diachronically coherent and yet express doubt about the reliability of his future opinions considered as indicators of what will happen.

Leaving this (at least somewhat) happy digression, let us turn to another proposal to defend (Reflection). Could it not be entailed by some more general principle about conveyance of factual information? Perhaps it would not be rational to have a state of opinion that it was not rational to convey, in so many words, to a suitable audience. But suppose I were to tell you: "If I say tomorrow morning that it will rain, there will still be a 50/50 chance that it will not." You would certainly look at me askance and reply that, in that case, you might as well not listen to me tomorrow morning. But then my assertion just now has taken away all value from my words of tomorrow morning about rain. We can see this as pathological if we take the following point of view: my expressions of opinion make statements about my mental state and, more particularly, about the aspect of my mental state which is meant to be a reliable indicator of relevant facts outside it. The value of these descriptions of my mental state—whether in the terminology of belief or of subjective probability—to my audience lies exactly in the information thus conveyed indirectly about what it is meant to be a reliable indicator of. Hence I have made a statement that cancels the normal conversational force of my statements of that sort.

I do not think that these reflections are entirely without force or relevance to (Reflection), but, as they stand, the rationale is quite wrong, and they do not constitute a defense. There is some ambi-

[18] This is not meant as an argument against conditionalization as a rational procedure; more sophisticated machinery than single one-place probability functions can be explored. This problem of how to represent certainty without dogmatism, which I shall not go into further here, is broached in Isaac Levi, *The Enterprise of Knowledge* (Cambridge, Mass.: MIT Press, 1980). It is not a problem if full certainty is not rational.

guity in the common use of both 'say' and 'there is a chance'. The first can be used to mean "assert" in a sense that implies belief, or requires in some other way that the assertor believes what he asserts; and the terminology of chance is sometimes used simply to express degrees of credence. If we adopt these interpretations when reading the example, it certainly has something putatively wrong with it, but that something is exactly that it implies a violation of (Reflection). Hence it does not manage to point to a more general principle to help us. If on the other hand we understand 'say' as "utter the words" or 'chance' as "objective chance," we have merely a statement that expresses doubt about the reliability of either my mental states or feelings or my words as indicators of rain. Although it is true that the audience is thereafter well advised not to take my words or opinion into account when deciding about the need for umbrellas, no principles or conversational maxims have been contravened. Such statements about my reliability as indicator of rain, need no more be logically odd or conversationally pathological than similar statements about the reliability of my watch. The audience is simply, in strict accordance with our conventions of conversational cooperation, advised to listen to the radio weather report (respectively, time signal) rather than to my guesses about this particular topic.

V. VOLUNTARISM AS SOLUTION

> "for what else is it to believe but to assent to the truth of what is propounded? Consent being a matter of the will . . ." (St. Augustine, *On the Spirit and the Letter*, 54)

The problem raised by the apparent need for principle of Reflection is, it seems to me, one of interpretation. A tenable interpretation of personal probability must either sever the link between rationality and coherence or else entail that Reflection is a form of epistemic judgment to which we must assent. It seems to me that among the debris in the preceding section there are some usable materials for the construction of an interpretation of the latter sort. The interpretation will first of all consider how the probability calculus can be viewed as a logic of epistemic judgments, and then consider exactly what such judgments are.

Let us begin with two challenges, one very familiar, the other due to Gilbert Harman. The first is that we simply do not have such a finely graded state of opinion as numerically precise subjective probabilities require. This challenge is answered by the admission that our personal probabilities are to some extent vague. Rain

tomorrow seems no less likely to me than a tossed coin's coming up heads four times in a row, no more likely than at least one of four tossed coins' coming up heads. My state of opinion is no more precise than this. Harman's challenge goes deeper. Since probabilities, unlike truth values, are not functional—$P(A\&B)$ is not a function of $P(A)$ and $P(B)$—, storing the information contained in an assignment of probabilities to sentences of even a "small" simple language quickly gets beyond the storage capacity of the mind. With vague probabilities the information storage problem gets worse, because each sentence now has two numbers assigned—a lower and upper probability. To circumvent this information explosion we must characterize a person's opinions as consisting of some which are more or less directly accessible plus all those to which the former commit him, on pain of violation of some higher criteria of rationality to which he subscribes.

No one, we say, has numerically precise degrees of belief. But at a given time I may, more or less consciously or overtly, make or be committed to a number of judgments of such forms as: it seems likely to rain, it seems as likely as not to snow, it seems likely to me—supposing it rains—that it will be cold, and so forth. These judgments express my opinions on various matters of fact; let us call them *epistemic judgments.* A certain family of these, accordingly, characterize my present state of opinion; they are *mine.* Unless I am very opinionated, they are not many, and they leave gaps: they may for instance not include, either directly or by implication, any judgment nontrivially comparing in such terms as the above, rain and newspaper reports of murder, or Darwin's theories and Einstein's.

It will be clear how an assignment of numbers to propositions could in principle reflect these judgments, because we are all familiar with their counterparts in the terminology of subjective probability. A person has, in the technical sense, a *coherent* state of opinion only if there exists at least one probability function P such that $P(A) > P(B)$ if it seems more likely to him that A than that B, $P(A) \geqslant P(B)$ if it seems no less likely to him that A than that B, $P(A \mid C) > P(B \mid C)$ if on the supposition that C it seems more likely to him that A than that B, and so forth. Let us say that such a function P *satisfies* his judgments. The lack of precision and other gaps in his judgments entail now that, if any one probability function P satisfies his judgments, then so do a number of others. The class of all that do, we may call the *representor* for his state of opinion. Unless that representation contains only a single function, we also

say that his degrees of belief, or subjective probabilities, are to some extent vague or indeterminate.[19]

We can now introduce a quite exact concept of implication among epistemic judgments for coherent states of opinion: if all probability functions satisfying each of a class X of judgments also satisfy judgment J, then (and only then) does X *coherently entail* J. It is exactly in such a case, when a person overtly makes all the judgments in X, that we say that he is also committed to J, on pain of incoherence.

Obviously a coherent state of opinion can be re-expressed in judgments formulated in the language of vague probability theory. "My subjective probability for A is no less than x, no greater than y" characterizes my state of opinion correctly if and only if, for every member P of my representor, $x \leq P(A) \leq y$. Similarly for subjective conditional probability, subjective odds, and subjective expectation. We see, therefore, that subjective-probability talk is merely the formulation, in sophisticated and flexible language, of judgments that have exactly the same status as, and indeed are entailed by, the epistemic judgments with which we began our discussion—for coherent states of opinion.

Therefore we must now look closely at exactly what an epistemic judgment is. Suppose I express my opinion as follows: "It seems more likely to me—supposing that it stays this cold—that it will snow than that it will rain." What exactly have I just done? One answer, the answer I wish to dispute, is that I have just made an autobiographical statement, describing my own psychological state.[20] Certainly, if you hear me say the above, you will be able to infer something about my psychological state, and perhaps this fact even provided the motive for my utterance. But that is very different from saying that what I did was to make an autobiographical statement of fact. (I belabor the point only because

[19] This emphasis on vagueness, and this sort of way to represent it, is especially to be found in Isaac Levi's and Richard Jeffrey's writings. For more technical details see also my "Rational Belief and Probability Kinematics," *Philosophy of Science*, XLVII, 2 (June 1980): 165-187.

[20] It is never easy to gauge one's agreement with other writers, but I think that in this I side with de Finetti—see p. 189 of his *Probability, Induction and Statistics* (New York: Wiley, 1972)—against Ramsey—see "Truth and Probability" in his *Foundations of Mathematics and Other Essays* (New York: Humanities Press, 1950). I would also like to refer to Stuart Hampshire's discussions of the connections between intention and knowledge or belief, in his *Freedom of the Individual* (Princeton, N.J.: University Press, 1975). Let me emphasize, however, with reference to the examples used here, that I regard acceptance of scientific theories as involving both more and less than belief; see my *The Scientific Image* (New York: Oxford, 1980), pp. 12/3, 80-83, 198-200.

John Austin is not generally discussed in writings on subjective probability.) Consider this story: yesterday morning I said to you "I promise you a horse by nightfall." This morning you point out that I have not got you a horse, and you accuse me of the heinous immorality of breaking my promise. Not at all, I say, I am guilty only of the lesser sin of lying; what I said yesterday morning was only a false autobiographical statement, for I was not in fact promising you a horse.

The sentence "I promised you a horse yesterday" is clearly a statement of fact, the fact that became true yesterday when I made the promise (perhaps by saying "I promise you a horse"). I wish to make the same sort of distinction with respect to the terminology of personal probability. In the preceding sections I already introduced a symbolic distinction, with the capital and lower-case distinction in $P(p_t(A) = r) = s$. If I were to say that, I would be expressing my opinion concerning a factual proposition about what my opinion was (is, will be) at time t. As analogue, consider "I promise you that I will not make you any promises concerning future dividends until I have carefully looked into the chances of success."

I do not mean that to express an opinion is to make a promise. The latter is a sort of ceremony in which I take upon myself, bring into being, an obligation to someone else. Two other alternatives suggest themselves: to express my opinion is to express my feelings, or it is to express an intention or commitment. There is something to be said for the first. A promise properly made will *follow* the agent's realization that he is willing, and able, to enter the corresponding contract or obligation. But expressing one's feelings generally involves, and may be the only means for, exploration of those feelings—I know that I feel strongly about this subject, but I don't know what I feel until I begin to talk or act or paint or write, and I discover almost as much about what I feel as the onlooker does. In this respect expressing one's opinions is often less like promising and more like emotive expression. But in this respect, expression of intentions is often the same. A difference is that, both in the case of opinion and of intention, and not in the case of feeling, the act of expression does not typically turn from genuine expression into something else, if one deliberately repeats the act.

Suppose, for example, that I have looked at my calibration score, found that I have generally overestimated the chances of rain, and now have exactly the same feelings on the question of rain as I did yesterday. Then my judgment about rain will now be different from what it was yesterday, for this judgment does not have the

function of merely expressing my feelings—properly made, formulating my judgment follows deliberation.

It seems then that, of the alternatives examined, epistemic judgments are most like expressions of intention. I may expres an intention either by simply stating the outcome of what I have decided upon ("You will be my successor") or by choosing a form of words traditionally suited to such expression ("You shall be my successor"). In either case, it is conveyed that I have made a decision, have formed an intention, am committed to a certain stance or program or course of action. There is no direct obligation to anyone else to fulfill this intention, but I have, as it were, entered a contract with myself. If I express this intention to an audience, then, just as in the case of a promise, I invite them to rely on my integrity and to feel assured that they now have knowledge of a major consideration in all my subsequent deliberation and courses of action. In this respect, expressing a considered judgment is similar.

Returning now to the principle (Reflection), consider the following analogies. I say, "I promise you a horse," and you ask, "And what are the chances that you'll get me one"? I say, "I am starting a diet today," and you ask, "And how likely is it that you won't overeat tomorrow? In both cases, the *first* reply I must give is "You heard me"! To express anything but a full commitment to stand behind my promises and intentions, is to undermine my own status as a person of integrity and, hence, my entire activity of avowal. This applies equally in the case of conditional questions. "If you promise to marry me, will you actually do it"? "If you decide to join our crusade, will you really participate"? In the first instance these questions are not invitations to an academic discussion of the objective chances, but challenges or probes of one's avowed intentions and commitments. It is confusing that the same words can be used for either purpose—not confusing in actual dialogue where contextual factors disambiguate, but confusing in written discussion.

Avowal, qua avowal, has its own constraints, which affect the logic of expressions of avowal. In none of the above cases do we have a simple way of characterizing what it is to be "false" to one's commitment. Having made a promise, I also have some obligation to prevent circumstances that would make it impossible to keep the promise. Having decided on a program of regular exercise, I have obliged myself to some extent to prevent travel arrangements, hangovers, lack of proper clothes and shoes, and so forth, that would interfere. It may not be easy for the onlooker, or even for me, to allocate blame or to decide whether I was false to myself or merely a

victim of circumstances. In the same way, if I express my opinion, I invite the world to rely on my integrity and to infer from this what advice to myself and anyone else in like circumstances, concerning the carrying of umbrellas, purchase of insurance policies, entering wagers, I would presently consider the best. Only in clinically hygienic cases would it be uncontroversially clear whether or not I really stood behind my expressed opinion. But that is so in the case of any expression of commitment or intention.

I conclude that my integrity, qua judging agent, requires that, if I am presently asked to express my opinion about whether A will come true, on the supposition that I will think it likely tomorrow morning, I must stand by my own cognitive *engagement* as much as I must stand by my own expressions of commitment of any sort. I can rationally and objectively discuss the possibility of a discrepancy between objective chance and my previsions. But I can no more say that I regard A as unlikely on the supposition that tomorrow morning I shall express my high expectation of A, than I can today make the same statement on the supposition that tomorrow morning I shall promise to bring it about that A. To do so would mean that I am now less than fully committed (a) to giving due regard to the felicity conditions for this act, or (b) to standing by the commitments I shall overtly enter.

VI. TRADITIONAL EPISTEMOLOGY REVISITED

This paper began with a statement of what I regard as a traditional epistemological view in philosophy of science: that we may rationally decide or come to believe propositions, hypotheses, theories which are not entailed (and which we ourselves do not regard as being made certain by) the evidence at hand. In addition—still spelling out this view—evidence itself is only the body of propositions that we have taken as evidence, and what we take to be evidence on a particular occasion may in fact be false. The refutation, along familiar Bayesian lines, was quick and sure and deadly: anyone who even regards himself as not totally unlikely to do what this view calls rational, is diachronically incoherent: vulnerable in that he implicitly regards as fair, disastrous combinations of wagers.

But then we also saw that the refutation is blocked by adherence to a principle, which goes well beyond the probability calculus, but which is equally required for the diachronic coherence of agents that Bayesian writers regard as rational. So the refutation is no refutation: we need not stop at conditionalization on the evidence on pain of incoherence, as long as we adhere to this principle, which even the strict conditionalizer himself (and also the less

committal observer described by Jeffrey) needs equally badly. Of course, the more improbable the proposition we decide to believe, or equivalently, the more we raise our credence in an uncertain proposition, the more risk we take. But that is merely a matter of degree, and there is no violation of coherence or any other criterion of rationality. Any accusation of epistemic extravagance is in any case to be met, by Jamesian and Bayesian alike, with the cool judgment "My credence that A is true, on the supposition that tomorrow I shall accord it credence to degree r, equals r." We can put the matter in either of two ways, depending on how we value the epithet of "Bayesian". Either that non-Bayesian epistemic behavior is defensible by exactly the same defense needed for Bayesian behavior; or, if you like, that apparently non-Bayesian behavior described by James and other traditional epistemologists, turns out to be, after all, entirely acceptable as far as Bayesian standards go. It may be a bit scary to think that such leaps of faith as James described in "The Will to Believe" or St. Augustine in "On Belief in Things Unseen"—he included his own belief in the existence of the Ocean—are not ruled out by the Bayesian's standards of coherence. But it is also a welcome thought, if we regard considerations of coherence as eminently rational, yet hope to find room for independence and enterprise in forming our world picture.

But then there is still the matter of the defense of the defence. I have argued that it is in fact indefensible if we regard the epistemic judgment—whether formulated in probabilistic or more qualitative terms—as a statement of autobiographical fact. The principle (Reflection) can be defended, namely as a form of commitment to stand behind one's own commitments, if we give a different, voluntarist interpretation of epistemic judgment. I call it "voluntarist," because it makes judgment in general, and subjective probability in particular, a matter of cognitive commitment, intention, *engagement*. Belief is a matter of the will.

BAS C. VAN FRAASSEN

Princeton University