MIDWEST STUDIES IN PHILOSOPHY, XVIII (1993)

Wittgensteinian Bayesianism

PAUL HORWICH

Belief is not an all-or-nothing matter. Rather, there are various *degrees* of conviction which may be represented by numbers between zero and one. Were we ideally rational, our full beliefs (of degree one) would comply with the laws of deductive logic; they would be consistent and closed under logical implication. And similarly, our *degrees* of belief should conform to the probability calculus.[1] This enrichment of epistemology—provided by the addition of degrees of belief and an appreciation of their probabilistic 'logic'—fosters progress with respect to many problems in the philosophy of science.

These statements form the core of a program, which I will call "therapeutic Bayesianism," whose primary goal is the solution of various puzzles and paradoxes that come from reflecting on scientific methods. Its creed is that many of these problems are the product of oversimplification, and that the above-mentioned elementary probabilistic model of degrees of belief often contains just the right balance of accuracy and simplicity to enable us to command a clear view of the issues and see where we were going wrong.[2] This somewhat Wittgensteinian goal and creed distinguishes therapeutic Bayesianism from more systematic enterprises in which probabilistic degrees of belief play a prominent role: for example, Bayesian decision theory, Bayesian statistics, Bayesian psychology, Bayesian semantics, and Bayesian history of science. It is especially important to appreciate the difference between the problem-solving orientation of therapeutic Bayesianism—that of exploiting a simple, idealized model in order to help illuminate notorious philosophical perplexities—and the quite distinct project of providing a perfectly true and complete (descriptive or normative) *theory* of scientific practice. The latter task might well involve the postulation of belief-gradations, and might also be done in the name of philosophy of science. However, its aims are quite different; and one must beware of judging one project by adequacy conditions appropriate to the other.[3]

Therapeutic Bayesianism is not self-evidently beneficial, but it does have some prima facie plausibility. Moreover, this plausibility is enhanced by substantial accomplishments, and, as we shall see, a great deal of the criticism it has received is misdirected—commonly for the reason just indicated. In this paper I would like to try to make a case for the program by discussing it from three, progressively abstract, points of view: substantial, foundational, and metaphilosophical. More specifically, there will follow sections on: (I) "The fruitfulness of therapeutic Bayesianism," in which I will sketch treatments of the 'raven' paradox and the question of diverse data and mention various other applications; (II) "Probabilistic foundations," in which the propriety of certain idealizations will be defended—particularly the representation of belief by numbers, the adoption of probabilistic canons of reason governing such beliefs, the definition of confirmation as increase in rational degree of belief, and the idea that induction may be codified in a confirmation function; and (III) "Misplaced scientism," in which I criticize a metaphilosophical perspective that does not properly distinguish science from the philosophy of science, and which overvalues the use of symbolic apparatus. Along the way, I shall respond to some criticisms of therapeutic Bayesianism that have recently been advanced.

I. THE FRUITFULNESS OF THERAPEUTIC BAYESIANISM

A good illustration of therapeutic Bayesianism at work is its way of treating the notorious 'raven paradox'. It is plausible to suppose that any hypothesis of the form 'All Fs are G' would be supported by the observation of an F that is also G. But if this is generally true, then the discovery of a non-black non-raven (e.g., a white shoe) confirms that all non-black things are non-ravens; and thereby confirms the logically equivalent hypothesis, 'All ravens are black'—a seemingly bizarre conclusion. This is 'the paradox of confirmation'. The Bayesian approach to this problem is to argue that observing a known raven to be black will *substantially* confirm "All ravens are black," whereas observing that a known non-black thing is not a raven will confirm it only *negligibly*—the difference being explained, roughly speaking, by the fact that, given our background beliefs about the chances of coming across ravens and black things, the first of these observations is more surprising, more of a test of the hypothesis, and therefore more evidentially powerful, than the second. Thus, the paradoxical flavor of our conclusion comes from the not unnatural confusion of negligible support with no support at all—a confusion sustained by inattention to degrees of belief and their bearing on confirmation.

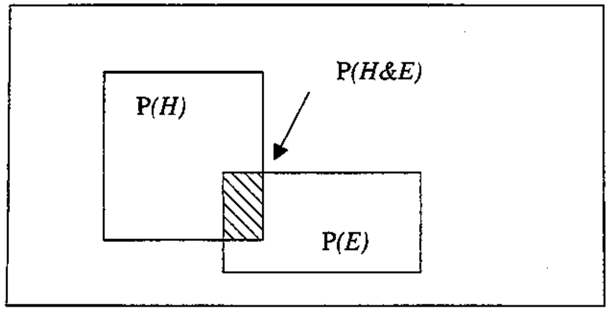
A formal version of this analysis proceeds from the following premises:

(a) That the amount of support for hypothesis H provided by evidence E is the factor by which the rational degree of belief in H is enhanced by the discovery of E—which is indicated by the ratio of subjective probabilities, $P(H/E)/P(H)$, for a rational person.

(b) That a rational person's degrees of belief will ideally conform to the probability calculus; and, in particular, will obey Bayes's Theorem:

$$\frac{P(H/E)}{P(H)} = \frac{P(E/H)}{P(E)}$$

(To appreciate the intuitive plausibility of this theorem, note that it derives from the fact that the conditional probability of H given E is equal to the probability of the conjunction of H and E, divided by the probability of E: i.e., $P(H/E) = P(H\&E)/P(E)$. See figure 1.



Therefore, since $P(H\&E) = P(E\&H)$, we obtain $P(H/E)P(E) = P(E/H)P(H)$, and hence Bayes's Theorem).

(c) That our degree of belief (prior to the investigation, and given the known scarcity of ravens) that a randomly selected non-black thing would turn out to be a non-raven is high.

(d) That our prior degree of belief (prior to the investigation, and given the known abundance of non-black things) that a randomly selected raven would turn out not to be black is substantial.

Now let us compare the support for the hypothesis, H, that all ravens are black, provided, first, by the discovery concerning a known raven that it is black (which is symbolized as $R^* B$), and, second, by the discovery that a known non-black thing is not a raven ($-B^* - R^*$). Applying premise (a) and then (b), we find:

$$\text{Support provided by } (R^* B) = \frac{P(H/R^* B)}{p(H)} = \frac{P(R^* B/H)}{P(R^* B)}$$

$$\text{Support provided by } (-B^* - R) = \frac{P(H/-B^* - R)}{p(H)} = \frac{P(-B^* - R/H)}{P(-B^* - R)}$$

But our hypothesis *entails* that any known raven would be black and any known non-black thing would not be a raven; therefore, $P(R^* B/H) = 1$ and $P(-B^* - R/H) = 1$; Therefore

$$\text{Support from raven found to be black} = \frac{P(H/R^* B)}{P(H)} = \frac{1}{P(R^* B)} = \text{1/prior degree of belief that a known raven would be black}$$

$$\text{Support from non-black thing found not to be a raven} = \frac{P(H/-B^* - R)}{P(H)} = \frac{1}{P(-B^* - R)} = \text{1/prior degree of belief that a known non-black thing would not be a raven}$$

Now one may assume (premise (c)) that a normal investigator of the hypothesis has prior background knowledge about the rough distribution of ravens and black things in his vicinity, and that this will lead him to expect that there is a very good chance that a randomly selected non-black thing will turn out not to be a raven. Thus $P(-B^* - R)$ is very nearly 1; and the amount of support for the hypothesis provided by observing that a non-black thing is not a raven is very little.

On the other hand, we would expect the background of investigation to dictate, in addition, (premise (d)) that the likelihood of a randomly selected raven being black is not especially high. After all, as far as we know at the outset of the research, there are many colors that the raven could perfectly well have. Thus $P(R^* B)$ is a good deal less than one. Therefore, the amount of support provided by observing that a known raven is black is substantial.

One might object to this reasoning that the final assumption is false, since the *objective* chances of finding that a raven is black are actually extremely high. However, this objection is based on a slip which is easy to identify. It confuses "probability" in the sense of *subjective degree of belief* and "probability" in the sense of *relative frequency*. All the probabilities mentioned in the argument are rational subjective probabilities, and it is under that construal that we may reasonably assume that $P(R^* B)$ is not near to 1. The feeling that this assumption is wrong derives from incorrectly reading $P(R^* B)$ as a relative frequency assertion. In that sense, since in fact almost all ravens are black, the probability that a randomly selected raven will be black is indeed very great. But this fact has no bearing on the argument.[4]

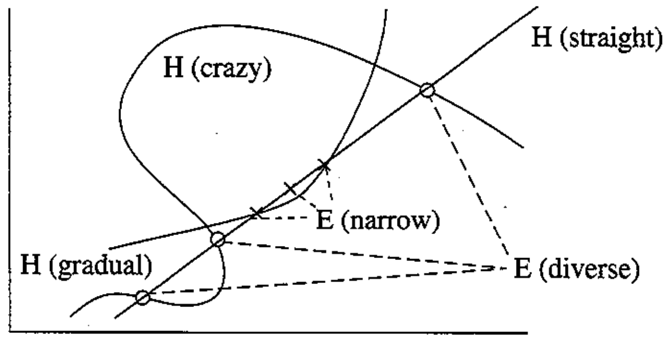
A similar objection is to deny that there could be any difference in evidential import between identifying a known raven as black and identifying a known black thing as a raven. Howson and Urbach,[5] for example, maintain that the only difference between these two data is the time order in which the elements of the observed fact are established. They think that in each case what is eventually known is the same, so there can be no variation in confirmation power between the two discoveries. Imagine, however, that an ornithologist instructs her assistant to go and find a black raven and bring it back to the lab for inspection. Surely, that inspection would count for nothing. And there is no paradox here, even though we might loosely speak of 'seeing a black raven' in all three cases. For a more precise characterization of the evidence shows that what is discovered in each case is not really the same. That a randomly selected raven turns out to be black, that a randomly selected black thing turns out to be a raven and that a randomly selected black raven turns out to be a black raven, are very different pieces of information, and it should not be surprising that they confirm our hypothesis to different degrees.

Therapeutic Bayesianism handles other issues in the philosophy of science similarly, putting a lot of weight on premises (a) and (b). By combining the idea of confirmation as enhancement of rational degree of belief, with the principle that rational degrees of belief should satisfy the probability calculus, we get a way of treating those problems that hinge upon considerations having to do with *degree of support*. Therefore the method has a wide scope. In particular, one can expect to shed light on why 'surprising' predictions have relatively great confirmation power, what is wrong with ad hoc hypotheses, whether prediction has more evidential value than mere accommodation of data, why a broad spectrum of facts can confirm a theory more than a narrow data set, why we base our judgments on as much data as possible, how statistical hypotheses can be testable despite their unfalsifiability, what is peculiar about 'grue-like' hypotheses, and various other problems.

These issues are unified by their involvement with the notion of 'varying evidential quality'; and this is why traditional epistemology, with its fixation on all-or-nothing belief, is not able to resolve them. It is only to be expected that the introduction of degrees of belief, together with an understanding of the rational constraints to which they are subject, would open the way to progress. Of course, there is not the space here to fully substantiate this thesis by describing all these applications of therapeutic Bayesianism. Let me, however, give one further illustration of the approach.

How is it that a broad spectrum of different kinds of fact, when entailed by a hypothesis, will confirm it to a greater degree than a uniform, repetitive set of data? It is natural to answer as follows. To the extent that our observations cover a broad range of phenomena, they are capable of falsifying a large number of alternative hypotheses, which then bequeath substantial credibility to those hypotheses that survive. Now, this solution does not quite work. For a narrow data set can preclude just as *many* hypotheses as a diverse data set. Nevertheless, we can repair the solution by noting that there is a significant difference in the *kinds* of hypotheses that are excluded by the two sets of facts. We should notice that the diverse data tend to exclude more of the *simple* hypotheses than do the narrow data. Given a representation of simplicity in terms of high prior probability,[6] this suggests that diverse data tend to rule out more high-probability alternatives than narrow data. But if so, then a hypothesis that survives relatively diverse observations becomes more probable than one that is left in the running by a narrow set of data. (See figure 2).

In particular, the data points E(narrow) exclude (given experimental error) just as many alternatives to the line H(straight) as does E(diverse). Nonetheless E(diverse) confirms H(straight) more strongly than E(narrow) does, because E(diverse) is better than E(narrow) at excluding simple alternatives to H(straight)—for example, gradual curves—which have an initially high



probability. Thus H(gradual) is ruled out by E(diverse) but not by E(narrow). On the other hand, the sort of hypothesis, like H(crazy), prohibited by E(narrow) yet not by E(diverse), is not very probable anyway; so excluding it does not greatly benefit those hypotheses that survive. Thus, with the help of a probabilistic representation of simplicity, we can begin to account for our methodological intuitions concerning diverse data.[7]

II. PROBABILISTIC FOUNDATIONS

In the last section I have tried to indicate something of the fruitfulness of therapeutic Bayesianism. Let me now consider various foundational questions that might be thought to cast doubt on the project:

(1) Do people actually have numerical degrees of belief?
(2) If so, can it be shown that *rational* degrees of belief conform to the probability calculus?
(3) Is it correct to identify degree of confirmation with rational enhancement of subjective probability?
(4) Are there objective facts of confirmation?
(5) Does reason require *merely* that one's beliefs conform to the probability calculus? Or is it the case (as Carnap thought) that a rational system of beliefs is subject to several further constraints?
(6) If further constraints are needed, then what are they?

On the first question, perhaps we should be agnostic. The successes of therapeutic Bayesianism will reinforce the evident fact that its basic principles are at least *roughly* correct. Thus we know that there are belief gradations of some sort, that there are rational constraints governing them (prohibiting, for example, a high degree of confidence in two contradictory propositions), and that confirmation is not wholly unrelated to increasing belief. Moreover, the Bayesian representation of these ideas has a great deal of plausibility. Consider a spectrum of situations in which we know that the propensities (as manifested by relative frequencies) of certain events are $x_1, x_2, \ldots, x_N$. For each such case there is a corresponding epistemic attitude—a degree of confidence—that the next trial will produce an event of the designated type. Presumably the appropriate attitude will vary with the relative frequency. Specifically, since the frequencies range over numbers between zero and one, so will the degrees of confidence.

However, despite the attractiveness of such considerations, one must of course acknowledge that the Bayesian framework might be wrong. The crude ideas that it represents should not be controversial. However, it is quite possible that the Bayesian articulation of those ideas is not absolutely right; and that, in particular, the assumption of precise-valued, numerical degrees of belief is incorrect.

Even so, such a model could be an excellent idealization, sufficing perfectly well for the primary purposes of therapeutic Bayesianism: namely, to dispel confusion, solve problems, and thereby improve our understanding of the scientific method. For the paradoxes are caused by forgetting the *crude* facts (that there are gradations of belief, etc.) or by failing to recognize their significance. And so the solutions will involve noticing that those rough ideas have been overlooked and coming to appreciate how they bear on the problems. This sort of treatment will not depend essentially on any particular theoretical refinements. The function of the Bayesian framework is merely to cast the crude, uncontroversial ideas into a form where their impact on our problems can have maximal clarity and force.

What then is the import of studies that cast doubt on the existence of numerical degrees of conviction and which develop more complex and allegedly more realistic conceptions of belief? Let me stress that this work falls well outside the focus of therapeutic Bayesianism, for there is no reason to believe that such improvements will help to solve the standard problems in the philosophy of science. Perhaps these developments are important in psychology, statistics, semantics, or decision theory; perhaps they will become important to philosophy when we have progressed enough in our understanding of science so that the *details* of an inductive logic become items of reasonable concern. But at this juncture, confusion is rampant, the traditional problems are still very much with us, and it seems rather unlikely that the slight gains in accuracy to be derived from a more realistic theory of belief would be worth the price—in terms of loss of simplicity—that we would have to pay for it.

Our treatment of the 'raven paradox' is a case in point. The problem was solved by exposing a certain misconception (that a non-black, non-raven would be irrelevant to our hypothesis), and by explaining why we are so tempted by that misconception: namely, that in forgetting about degrees of belief, we lose sight of the distinction between very slight confirmation and no confirmation at all. The simple Bayesian model of belief provides a sufficiently perspicuous representation of the situation to enable us to put this in a clear way. Further accuracy regarding the nature of belief would distract us from the main point, ruin the argument, and not help us to understand the basis of the paradox.

Let me give another example. The conflict between realism and instrumentalism with respect to the acceptance of scientific theories is fueled by a shared tendency to think in terms of all-or-nothing belief. The instrumentalist argues, in light of previous scientific revolutions, that it is foolishly optimistic to expect that our current theories are true and will not eventually be refuted. Whereas the realist complains that it is a distortion of science to distinguish rigidly between credible observation reports and incredible theoretical claims. However, once we see that the issue is not 'To believe or not to believe?' but rather, 'To what degree shall we believe?' then there is room for reconciliation. The crucial move is the elimination of the shared misconception. There is no reason to think that a fancy model of belief, even more accurate than the Bayesian idealization, would be any further help with the problem.

I do not mean to be suggesting that it is not worthwhile to investigate more sophisticated models of belief. On the contrary, I can readily imagine research programs—e.g., Bayesian psychology, or attempts to give a perfectly accurate description of scientific practice—in which this would be crucial. My point is that there is another enterprise—the one I am calling "therapeutic Bayesianism"—whose focus is on solving the traditional methodological puzzles and paradoxes, and for which the introduction of such complex models is likely to do more harm than good.

Suppose, then, that we do have numerical degrees of belief. Is there any way of justifying the Bayesian assumption that, to be rational, these degrees of belief must conform to the probability calculus? Although there are indeed various lines of reasoning which purport to establish this thesis, none is compelling. The best known of them is the 'dutch book' argument[8] and it goes roughly as follows. Defining a person's degree of belief in a proposition as a function of the odds at which he is prepared to bet on its truth, it can be proved that if his degrees of belief do not satisfy the probability calculus then he will be prepared to accept a collection of bets which is guaranteed to lead to a loss. Therefore, since it would surely be irrational for him knowingly to put himself in such a no-win situation, it would be irrational to have a system of degrees of belief that violates the probability calculus. QED. However, the definition of 'degree of belief' that is employed in this argument presupposes that people maximize their expected utility. And there is a lot of room for skepticism about that assumption (and about the preference axioms to which it is equivalent). So the 'dutch book' argument is far from airtight. Worse still, there is positive reason to think that its conclusion is false, for it requires logical omniscience. The probability of any logical truth is 1 and of any contradiction is 0. Yet it is surely quite rational to be less than perfectly confident in the truth of *some* logical truths—those that are especially hard to prove—and quite rational to give non-zero degrees of belief to contradictions that are hard to recognize as such.

The proper response to these difficulties is to repeat that the picture of rational degrees of belief obeying the probability calculus should be regarded as an *idealization* of the real normative facts. It is uncontroversial that one ought to be certain of elementary logical truths, and that one ought not be confident of the truth of obviously incompatible hypotheses. The probabilistic model of belief provides a sharp, perspicuous way of capturing these trivialities, and to the extent that it goes beyond them it need not be construed realistically.

A similar answer may be given to the third question concerning the definition of confirmation. In our discussion of the raven paradox we defined "the degree by which E confirms H" as "the ratio, $P(H/E)/P(H)$, for a rational person." Evidently, this explication has at least *some* prima facie plausibility, and it certainly helps us to give a neat, compelling solution to the problem. Nonetheless, it is often argued that this particular explication is 'wrong'—yielding counterintuitive consequences—and that there are better definitions of confirmation which should be used instead.[9]

However, these criticisms have little relevance to the project of therapeutic Bayesianism. No doubt our explication leads to some strange-sounding consequences. No doubt it is strictly speaking false that the ordinary meaning of "confirms" is given by our explication. No doubt there are definitions (perhaps involving non-probabilistic notions) that come closer to what we ordinarily mean. But the object of therapeutic Bayesianism is not to give a theory of science. We are not trying to find the most accurate analyses of our concepts, but rather to use explications that are at least roughly right, and which are conducive to simple, convincing dissolutions of philosophical problems. Since we assume that these problems are the product of confusion, it is desirable to look for ways of clarifying the issues, which have the proper blend of accuracy and simplicity. Of course it is possible to *over*simplify. But one can conclude that this has happened only after finding that the admittedly idealized models do not in fact help to solve our problems.[10]

On the fourth question—Are there objective facts of confirmation?—it seems evident that judgments of credibility and confirmation do purport to capture objective normative facts. They do not state what any individual's degrees of belief *actually* are, but rather they say something about what one's degrees of belief *ought* to be, or how they *ought* to change given the circumstances. Thus we should acknowledge non-subjective facts regarding confirmation.

A natural way of capturing this idea, due to Carnap, is to suppose that an attribution of probability to a hypothesis reflects the belief in an objective, logical fact about the degree to which one statement—a summary of the available evidence—probabilifies another statement—the hypothesis in question. Such logical facts might be codified in a confirmation function, $c(p/q) = x$, which would specify explicitly the degree, x, to which q confirms p, and would specify implicitly the degree to which one should believe p if the total evidence is q. Carnap says, for example:

> Probability-1 is the degree of confirmation of a hypothesis h with respect to an evidence statement e, e.g., an observation report. This is a logical, semantical concept. A sentence about this concept is based, not on observation of facts, but on logical analysis; if it is true, it is L-true (analytic). . . . Probability-2 [relative frequency] is obviously an objective concept. It is important to recognize that probability-1 is likewise objective.
>
> Let h be the sentence 'there will be rain tomorrow' and j the sentence 'there will be rain and wind tomorrow'. Suppose someone makes the statement in deductive logic: 'h follows logically from j'. . . . The statement 'the probability-1 of h on the evidence e is 1/5' has the same general character as the former statement. . . . Both statements express a purely logical relation between two statements. The difference between the two statements is merely this: while the first states a complete logical implication, the second states, so to speak, a partial logical implication; hence, while the first belongs to deductive logic, the second belongs to inductive logic.[11]

Thus, Carnap held that certain facts about confirmation are analytic and *objective*, and thought of inductive probability as a *partial* version of the logical relation of entailment.[12]

On the fifth question—Does reason impose constraints on belief *over and above* the requirement of conformity with the probability calculus?—there are grounds for sympathy with Carnap's view that it does. For it is hard to see how the probabilistic constraint alone can account for our intuitions about the relative plausibility of competing hypotheses that equally well fit the current data. In particular, it is hard to see how it can solve the 'grue' problem.[13]

Suppose that such further constraints are indeed required. Still, to take up the sixth question, it is no trivial matter to say what they are. Carnap tried out various constraints and employed them to derive confirmation functions for certain extremely simple formal languages. Unfortunately, these functions have the counterintuitive property that laws of nature are never able to acquire more than a negligible probability. And this shows a deficiency in Carnap's constructions: either the languages are too simple, or the constraints are wrong. However, one can certainly not conclude that anyone who endorses a Carnapian conception of logical probability *must* hold that general laws never attain a non-negligible probability. This is a non sequitur, arising from a failure to distinguish between the general conception of logical probability and the admittedly inadequate prototypes with which Carnap experimented.[14]

For the treatment of various problems it is helpful to suppose that inductive reasoning is represented by a specific (but unspecified) real-valued Carnapian confirmation function, c, allowing general laws to achieve a non-negligible credibility. In light of our responses to questions (1), (2), and (3), we see that it can be no objection to this procedure that our inductive practice is not in fact precisely described by a single c-function. For, once again, the intention is not to get at the exact truth, but merely to employ a useful idealization. Nor—as we have just said—is it fair to complain that *some* c-functions—

those Carnap toyed with—always give zero probability to general laws. For we can suppose that 'the right c-function' is one of those that do *not* have that counterintuitive feature.

III. MISPLACED SCIENTISM

Much criticism of therapeutic Bayesianism arises from a conflation of philosophy and science. More exactly, it derives from a failure to recognize the legitimacy (even the existence) of non-scientific philosophical projects—those prompted, not by a desire to expose the whole truth regarding some domain, but by an interest in the resolution of paradoxes. Let me elaborate.

What one might call '*theory*-oriented philosophy of science' aims for a systematic account of the scientific method. The criteria of success are just those that pertain to theory construction *within* particular sciences: namely, empirical adequacy, scope, depth, simplicity, internal consistency, and coherence with the rest of our knowledge. More specifically, a perfect theory of the scientific method would be expected to conform with specific intuitions about the way that good science is done, to cover all aspects of methodology in detail, to expose fundamental principles enabling the complex, superficial aspects of scientific practice to be unified and explained, and to respect results in psychology and sociology. Thus it seems appropriate to regard theory-oriented philosophy of science as itself a department of science—a branch of naturalized epistemology. This characterization neither ignores nor denies that scientific methods are normative. A description of science will contain a codification of the basic norms which are implicit in the evaluation of theories. Moreover, it is quite possible that an identification of basic normative principles will result in the exposure of cases in which science is being done badly. One might thereby effect an improvement in the conduct of some science. Indeed, this may well be a motive for engaging in theory-oriented philosophy of science.

In contrast, the approach portrayed in this paper, '*problem*-oriented philosophy of science', has very different goals, methods, and adequacy conditions. It aims at the resolution of deep puzzles and paradoxes that arise from reflection upon science. It includes in its domain, for example, the problem of induction, the paradox of confirmation, the question of total evidence, and the issue of prediction versus accommodation. The problems here are not simply to fill various undesirable gaps in our knowledge about science. Characteristically, they are conceptual tensions, contradictions, absurd conclusions—that is to say, symptoms of confusion. We have somehow gone astray, and the task is to understand how this has happened and to get a clear view of the issue so that our misguided ways of thinking will be exposed and no longer seem so attractive.

These two approaches to the philosophy of science do not compete with one another. They are distinct projects with distinct objectives—not *wholly* unrelated to one another, but by no means simply parts of the same enterprise. Thus, it is not the case that the sort of full understanding provided by a successful theory-oriented philosophy of science would automatically solve the puzzles that form the domain of problem-oriented philosophy of science. For the resolution of a paradox requires a great deal more than just locating the wrong move in a fallacious argument. It is crucial to a proper resolution that one comes to see why that fallacy was natural. And it is important that one obtains a new perspective on the issue—a point of view from which the old and troublesome habits of thought no longer seem plausible. These elements of the solution do not simply fall out of a complete theory of science. (Similarly, the explanation of conjuring tricks does not follow from physics.) Thus, theory-oriented philosophy of science is not simply a more thorough, systematic, and ambitious project than problem-oriented philosophy of science.

Neither is it necessary, in order to solve problems, that one be in possession of an adequate theory of science. For confusions can be identified, understood, and removed without a theory of any particular depth or generality. Granted, assumptions about methodology will often be involved in the diagnosis and treatment of a problem, and if these were *wildly* false then it is unlikely that the discussion would be helpful. However, there is no reason why such assumptions should be *true* as long as their replacement with the truth would not undermine the solution that is based on them. Indeed it is quite possible that the perfect theory of science would be a very bad tool for solving problems. For the truth may be so complicated that it cannot provide the sort of simple and relevant perspective that is needed.

If the practice of conceptual troubleshooting is confused, as it often is, with the scientific search for a theory of science, then therapeutic Bayesianism will be wrongly subjected to all of the methodological requirements that are properly applied only in science. Let me describe some of the bad effects of this confusion.

One consequence, discussed above, of not seeing the distinctive aim of therapeutic Bayesianism is a tendency to misjudge the function of various helpful idealizations. Thus one commonly finds objections to the use of precise-valued degrees of belief, to the assumption that these should conform to the probability calculus, to the adoption of a particularly simple explication of confirmation, and to the idea that our inductive practice may be represented by a single Carnapian confirmation function. Doubtless some of these assumptions are, strictly speaking, false. (Just as it is false that a gas is made of point masses.) And in a different kind of study—one aimed at truth—it would be very important to discuss more realistic models. However, for the purposes of therapeutic Bayesianism it is important to use the simplest roughly accurate models of degrees of belief and of confirmation that will help to clarify the issues, and it is sufficient to proceed on the basis of their intuitive plausibility and to justify these models in retrospect in terms of their utility.

Secondly, a scientific understanding of confirmation aims for the truth, the *whole* truth, and nothing but the truth. Consequently, those wedded to this conception of the philosophy of science will find fault with studies that do not discuss every significant aspect of the phenomenon of confirmation. Consider, for example, *prior probability assignment*, the procedures for deciding, before data have been gathered, the various 'intrinsic' plausibilities of hypotheses; *belief-kinematics*, the way that systems of belief change over time in the light of new discoveries; or *direct inference*, the impact on our degrees of belief of a knowledge of empirical probabilities. These are fascinating topics, and a good theory of science must deal with them. But there is no reason why a paradox-oriented Bayesian program should incorporate a complete, systematic account of all such elements of methodology.[15]

In the third place, scientism in philosophy engenders a 'hyperformalist' fixation on symbolic technique—an overvaluation of logico-mathematical machinery. Among the symptoms of this hyperformalist state are: (a) a blindness to the possibility of philosophical problems distinct from the scientific and mathematical issues that arise in statistics, decision theory, sociology of science, etc., further questions being dismissed as 'merely verbal';[16] (b) a dissatisfaction with informal discussions and conclusions; (c) an exaggerated concern with formal rigor for its own sake; and (d) an obsession with the elimination of any potential ambiguity or vagueness, leading to the feeling that the English language is too confusing and vague a medium for intellectual progress, and that it should, wherever possible, be replaced with mathematics or logic.

Thus, even if an approach employs formal techniques, as therapeutic Bayesianism clearly does, it may still be subjected to hyperformalist criticism. I think this is an unhealthy point of view—in philosophy generally, and particularly in the philosophy of science, where it is especially common. No doubt there are occasions when clarity is gained and confusion allayed with the help of formal apparatus. This, I believe, is one of the morals of Bayesianism's success. However, one can withdraw too quickly into the secure, regulated territory of a formal system. It is certainly a tempting relief from the frustrating vagaries of philosophy to be able to obtain definite, proven results and get clear answers to clear questions. But, unless we are very careful, these answers and results might have little to do with the problems that have traditionally motivated philosophy of science. Our methodological puzzles arise when we reflect informally about scientific practice; and they can be solved only with an appreciation of the misconceptions and confusions to which we are prone and an understanding of the ways in which they are fostered by the rich conceptual resources put at our disposal by natural language. It seems to me that only when that sort of understanding is eventually attained will we know what we are looking for in a fully fledged inductive logic; and then, perhaps, be in a better position to devise one. But this level of understanding will not be achieved by trying to express as many questions as possible within a formal system, proving some theorems, and dismissing the residue as intractable and uninteresting. At its worst, such scientistic hyperformalism betrays a lack of concern for truly philosophical problems. If "merely verbal" issues are any that do not make a scientific difference, and if only scientific problems are worth worrying about, then philosophy is truly an endangered enterprise.

I hope to have clarified what I believe is a valuable approach to the philosophy of science, and to have shown that many of the complaints about it derive from scientistic hyperformalism and are therefore misconceived. The goal is not a theory of science but the unravelling of puzzles surrounding our ideas about surprising data, prediction versus accommodation, ad hoc postulates, statistical hypotheses, our thirst for new data, the tenability of realism, and other aspects of methodology. And given some of the successes of therapeutic Bayesianism there is reason to have a fair amount of confidence in its basic principles.

Thus, the notion of rational degrees of belief conforming to the probability calculus has an important role in the philosophy of science. It would no doubt be easier to think in terms of all-or-nothing belief, but that oversimplification is part of what engendered our methodological puzzles in the first place. On the other hand, there are more complex and realistic conceptions of belief, but the cause of clarity is not served by using them. Therapeutic Bayesianism appears to offer the ideal compromise between accuracy and simplicity, enabling us to represent the issues starkly without neglecting the essential ingredients or clouding them with unnecessary details.[17]

NOTES

1. The axioms of elementary probability theory are as follows: (1) probabilities are less than or equal to one; (2) the probability of a necessary truth is equal to one; (3) if two statements are jointly impossible, then the probability that at least one of them is true is equal to the sum of their individual probabilities; and (4) the conditional probability of p given q equals the probability of the conjunction of p and q divided by the probability of q.

2. This project is attempted in my *Probability and Evidence* (Cambridge, 1982), henceforth abbreviated as *P&E*. The metaphilosophical outlook is inspired by Wittgenstein's *Philosophical Investigations*, paragraphs 88–133.

3. Bayesian programs of various kinds have been developed in the work of Rudolf Carnap, David Christensen, R. T. Cox, Bruno de Finetti, Ron Giere, I. J. Good, John Earman, Ellery Eells, Hartry Field, Allan Franklin, Ian Hacking, Mary Hesse, Jaakko Hintikka, Colin Howson, E. T. Jaynes, Richard Jeffrey, Harold Jeffreys, Mark Kaplan, J. M. Keynes, Henry Kyburg, Isaac Levi, Patrick Maher, Roger Rosencrantz, Wesley Salmon, L. J. Savage, Teddy Seidenfeld, Abner Shimony, Brian Skyrms, Patrick Suppes, Peter Urbach, Bas van Fraassen and others. Much of this work (especially the studies by Good, Hesse, Howson & Urbach, and Earman) contains contributions to therapeutic Bayesianism. However, I cannot attribute to these philosophers the project that I have in mind by that label, since their work is oriented towards the discovery of a 'theory of science', and thus reflects a metaphilosophical point of view that is quite distinct from that of the program which I am calling "therapeutic Bayesianism."

4. Stephen Spielman's objection is based on the mistake described here: the identification of the probabilities with objective proportions. (See his review of *P&E, Journal of Philosophy* [March 1984]: 168-73. Page references for Spielman are to this work).

To keep things relatively simple I have assumed that there are just *two* observations in question: namely, the discovery regarding a randomly selected raven that it is black and the discovery regarding a randomly selected non-black thing that it is not a raven. If we consider instead the discovery that a known black thing is a raven, or various other ways of seeing black ravens and non-black non-ravens, then the existence of confirmation depends on the presence of special additional background assumptions (e.g., that ravens are quite likely all to have the same color). Nonetheless a similar contrast between the degrees of confirmation provided by black ravens and non-black non-ravens may be established. In *P&E* I suggest that these other ways of seeing black ravens would provide *no* confirmation of the hypothesis. This is misleading. Sometimes our background theories include a belief in the projectibility of the generalization in question, and in that case all the ways of observing an instance of it will normally provide confirmation.

My treatment of the paradox is, in a couple of respects, different from Patrick Suppes's analysis ("A Bayesian Approach to the Paradoxes of Confirmation," in *Aspects of Inductive Logic*, edited by J. Hintikka and P. Suppes, [Amsterdam 1966]). In the first place, he does not distinguish between the discovery that a randomly selected object is a black raven and the discovery that a randomly selected raven is black; whereas it is a significant feature of my account that in certain circumstances only the latter datum would confirm the hypothesis. And secondly, he does not obtain his results from the basic principles of Bayesianism—the thesis that degrees of belief should conform to the probability calculus; rather, he starts with the assumption that surprising observations have greater confirmation power; and this, though correct, is much better derived than simply presupposed.

5. C. Howson and P. Urbach, *Scientific Reasoning: The Bayesian Approach* (La Salle, Ill., 1989).

6. An argument for associating simplicity with high prior probability is given in *P&E.*, 70–71.

7. Teddy Seidenfeld maintains that this account goes in the "wrong direction." But he gives no grounds for that claim other than to note the above-mentioned deficiencies in our understanding of simplicity—our inability to solve either the descriptive or the normative problems surrounding it. And it seems to me that his observation is irrelevant in the absence of any reason to believe either (a) that we can get a satisfactory explication of simplicity in terms of evidential diversity, or (b) that the Bayesian account would not withstand a better grasp of simplicity. (See his review of *P&E, Philosophical Review*, [July 1984].)

8. For a good assessment of this argument and various others see John Earman's *Bayes or Bust* (Cambridge, Mass., 1992). Bruno de Finetti perhaps deserves the credit for first having argued that degrees of belief *ought* to be 'coherent', i.e., conform to the probability calculus—though they *need* not be coherent if the believer is irrational ("Foresight: Its Logical Laws, Its Subjective Sources," translated in *Studies in Subjective Probability*, edited by H. E. Kyburg, Jr., and H. E. Smokler [New York, 1964]). I hesitate to credit Frank Ramsey's earlier paper (in *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, edited by D. H. Mellor [Atlantic Highlands, N.J., 1977]) with this result, since he defines "degrees of belief" in such a way that they *must* conform to the probability calculus. On Ramsey's account there is no room for the existence of someone who has degrees of belief that are not coherent.

9. For example, I. J. Good (in the *British Journal for the Philosophy of Science* 19 [1968]: 123–43) advocates:

$$\begin{array}{c}\text{Weight of evidence}\\ \text{concerning } H \text{ provided by } E\end{array} = \log \frac{\mathrm{P}(E/H)}{\mathrm{P}(E/-H)}$$

And Seidenfeld (op. cit.), noting that on our account E might confirm both H_1 and H_2 yet disconfirm the conjunction $(H_1 \ \& \ H_2)$, suggests that confirmation cannot be defined in terms of probability alone.

10. A further complaint is that our definition of confirmation seems to go badly wrong when we apply it to measure the evidential value of *already known* data. For in that case $\mathrm{P}(E) = 1$, therefore $\mathrm{P}(H/E) = \mathrm{P}(H)$. This problem for Bayesians was first posed by Clark Glymour (see his *Theory and Evidence* [Princeton, 1980]). It has been forcefully reiterated by James Woodward (in his review of *P&E, Erkenntnis* 23 (1985): 213–19) and treated thoroughly by John Earman (in *Bayes or Bust*). In order to deal with it we should remember that the idea of the definition is to compare the credibility of a hypothesis, H, given the knowledge that E is true, with its credibility in the absence of such knowledge. Thus we should take the prior probability to be that which H would have had if the truth of E had not been discovered. Then, in order to assess E's confirmation power, we should consider what the absolute subjective probability of E would have been in that counterfactual situation, and also what the conditional probability of E given H would have been. Then we can employ Bayes's Theorem to calculate the factor by which the prior probability of H would have been increased. Doubtless, there is substantial indeterminacy in the assessment of these counterfactual probabilities. But this is no objection, since we generally have no reason to expect the magnitude of E's confirmation-power to be an especially determinate matter.

11. Rudolf Carnap, *Logical Foundations of Probability* (Chicago, 1962), 19, 31.

12. According to Spielman, this construal of Carnap is a "distorted caricature" (170), for "any careful reading of LFP [Logical Foundations of Probability] would show that Carnap never talks about 'objective relations of probabilification' or 'objective' relations of partial entailment" (170). Here I am at a loss to explain how Spielman could have arrived at his interpretation, and I can only refer the reader back to Carnap's work.

13. For further discussion of this point see *P&E*, 32–36 and 74–81 and Earman's *Bayes or Bust*, chapter 6.

14. Spielman (171) falls into this error, complaining that one cannot endorse logical probability and yet still assume that laws can have a non-negligible credibility.

15. Thus Woodward writes: "The principal defect of *Probability and Evidence* is its unsystematic character. Horwich does not give us a fully worked out general theory of confirmation but rather a series of essays which offer solutions to various particular puzzles, where the interconnections among these solutions are by no means always clear" (214).

16. This is starkly revealed in Seidenfeld's dismissal of therapeutic Bayesianism on the grounds that it is no substitute for a combination of excellent, but highly technical, foundational studies in decision theory and statistical inference by Jeffrey, Fishburn, and Lindley—works that hardly touch upon the traditional philosophical puzzles that form the domain of therapeutic Bayesianism. In a similar vein, Spielman is bothered by the "fail(ure) to see that the only difference between an 'objectivist' account [of the 'grue' problem] and a personalist account would be verbal: an objectivist would say that we *ought* to assign a much higher probability to H_1 than to H_2, and a subjectivist says that this is what intelligent informed people in fact do" (170). Spielman thinks the issue between them is 'merely verbal'.

17. I have greatly benefited from James Woodward's thorough and perceptive criticism. I would also like to thank Ned Block, Susan Brison, Josh Cohen, Marcus Giaquinto, Mark Kaplan, and Judith Thomson for helping me to improve earlier drafts of this paper.

Wittgensteinian Bayesianism

PAUL HORWICH

Belief is not an all-or-nothing matter. Rather, there are various *degrees* of conviction which may be represented by numbers between zero and one. Were we ideally rational, our full beliefs (of degree one) would comply with the laws of deductive logic; they would be consistent and closed under logical implication. And similarly, our *degrees* of belief should conform to the probability calculus.[1] This enrichment of epistemology—provided by the addition of degrees of belief and an appreciation of their probabilistic 'logic'—fosters progress with respect to many problems in the philosophy of science.

These statements form the core of a program, which I will call "therapeutic Bayesianism," whose primary goal is the solution of various puzzles and paradoxes that come from reflecting on scientific methods. Its creed is that many of these problems are the product of oversimplification, and that the above-mentioned elementary probabilistic model of degrees of belief often contains just the right balance of accuracy and simplicity to enable us to command a clear view of the issues and see where we were going wrong.[2] This somewhat Wittgensteinian goal and creed distinguishes therapeutic Bayesianism from more systematic enterprises in which probabilistic degrees of belief play a prominent role: for example, Bayesian decision theory, Bayesian statistics, Bayesian psychology, Bayesian semantics, and Bayesian history of science. It is especially important to appreciate the difference between the problem-solving orientation of therapeutic Bayesianism—that of exploiting a simple, idealized model in order to help illuminate notorious philosophical perplexities—and the quite distinct project of providing a perfectly true and complete (descriptive or normative) *theory* of scientific practice. The latter task might well involve the postulation of belief-gradations, and might also be done in the name of philosophy of science. However, its aims are quite different; and one must beware of judging one project by adequacy conditions appropriate to the other.[3]

Therapeutic Bayesianism is no
some prima facie plausibility. More
stantial accomplishments, and, as
it has received is misdirected—co
this paper I would like to try to ma
from three, progressively abstract, p
metaphilosophical. More specificall
fruitfulness of therapeutic Bayesia
of the 'raven' paradox and the que
other applications; (II) "Probabilisti
certain idealizations will be defende
by numbers, the adoption of proba
beliefs, the definition of confirmatio
and the idea that induction may be
(III) "Misplaced scientism," in which
that does not properly distinguish sc
which overvalues the use of symboli
to some criticisms of therapeutic Bay

I. THE FRU
THERAPEUT

A good illustration of therapeutic B
the notorious 'raven paradox'. It is
of the form 'All Fs are G' would
F that is also G. But if this is ge
black non-raven (e.g., a white shoe)
ravens; and thereby confirms the logi
black'—a seemingly bizarre conclus
The Bayesian approach to this prob
raven to be black will *substantially*
observing that a known non-black
negligibly—the difference being exp
given our background beliefs about
black things, the first of these obser
of the hypothesis, and therefore mo
Thus, the paradoxical flavor of our
confusion of negligible support with
by inattention to degrees of belief a

A formal version of this analys

(a) That the amount of support
is the factor by which the
by the discovery of E—wh
probabilities, $P(H/E)/P($