

*Rationality and Objectivity in Science*  
or  
*Tom Kuhn Meets Tom Bayes*

Twenty-five years ago, as of this writing, Thomas S. Kuhn published *The Structure of Scientific Revolutions*.<sup>1</sup> It has been an extraordinarily influential book. Coming at the height of the hegemony of *logical empiricism*—as espoused by such figures as R. B. Braithwaite, Rudolf Carnap, Herbert Feigl, Carl G. Hempel, and Hans Reichenbach—it posed a severe challenge to the logistic approach that they practiced.<sup>2</sup> It also served as an unparalleled source of inspiration to philosophers with a historical bent. For a quarter of a century there has been a deep division between the logical empiricists and those who adopt the historical approach, and Kuhn's book was undoubtedly a key document in the production and preservation of this gulf.

At a 1983 meeting of the American Philosophical Association (Eastern Division), Kuhn and Hempel—the most distinguished living advocates for their respective viewpoints—shared the platform in a symposium devoted to Hempel's philosophy.<sup>3</sup> I had the honor to participate in this symposium. On that occasion Kuhn chose to address certain issues pertaining to the rationality of science that he and Hempel had been discussing for several years. It struck me that a bridge could be built between the differing views of Kuhn and Hempel if Bayes's theorem were invoked to explicate the concept of scientific confirmation.<sup>4</sup> At the time it seemed to me that this maneuver could remove a large part of the dispute between standard logical empiricism and the historical approach to philosophy of science on this fundamental issue.

I still believe that we have the basis for a new consensus regarding the choice among scientific theories. Although such a consensus, if achieved, would not amount to total agreement on every problem, it would represent a major rapprochement on an extremely fundamental issue. The purpose of the present essay is to develop this approach more fully. As it turns out, the project is much more complex than I thought in 1983.

---

I should like to express my deepest gratitude to Adolf Grünbaum and Philip Kitcher for important criticism and valuable suggestions with respect to an earlier version of this paper.

## §1. Kuhn on Scientific Rationality

A central part of Kuhn's challenge to the logical empiricist philosophy of science concerns the nature of theory choice in science. The choice between two fundamental theories (or paradigms), he maintains, raises issues that "cannot be resolved by proof." To see how they are resolved we must talk about "techniques of persuasion," or about "argument and counterargument in a situation in which there can be no proof." Such choices involve the exercise of the kind of judgment that cannot be rendered logically explicit and precise. Such statements, along with many others that are similar in spirit, led a number of critics to attribute to Kuhn the view that science is fundamentally irrational and lacking in objectivity.

Kuhn was astonished by this response, which he regarded as a serious misinterpretation. In his "Postscript—1969," in the second edition of *The Structure of Scientific Revolutions*, and in "Objectivity, Value Judgment, and Theory Choice"<sup>5</sup> he replies to these charges. What he had intended to convey was the claim that the decision by the community of trained scientists *constitutes* the best criterion of objectivity and rationality we can have. In order better to understand the nature of such objective and rational methods we need to look in more detail at the considerations that are actually brought to bear by scientists when they endeavor to make comparative evaluations of competing theories.

For purposes of illustration, Kuhn offers a (nonexhaustive) list of characteristics of good scientific theories that are, he claims:

individually important and collectively sufficiently varied to indicate what is at stake. . . . These five characteristics—accuracy, consistency, scope, simplicity, and fruitfulness—are all standard criteria for evaluating the adequacy of a theory. . . . Together with others of much the same sort, they provide the shared basis for theory choice.<sup>6</sup>

Two sorts of problems arise when one attempts to use them.

Individually the criteria are imprecise: individuals may legitimately differ about their applicability to concrete cases. In addition, when deployed together, they repeatedly prove to conflict with one another; accuracy may, for example, dictate the choice of one theory, scope the choice of its competitor.<sup>7</sup>

For reasons of these sorts—and others as well—individual scientists may, at a given moment, differ regarding a particular choice of theories. In the course of time, however, the interactions among individual members of the community of scientists produce a consensus for the group. Individual choices inevitably depend upon idiosyncratic and subjective factors; only the outcome of the group activity can be considered objective and fully rational.

One of Kuhn's major claims seems to be that observation and experiment, in conjunction with hypothetico-deductive reasoning, do not adequately account for

the choice of scientific theories. This has led some philosophers to believe that theory choice is not rational. Kuhn, in contrast, has tried to locate the additional factors that are involved. These additional factors constitute a crucial aspect of scientific rationality.

## §2. Bayes's Theorem

The first step in coming to grips with the problem of evaluating and choosing scientific hypotheses or theories<sup>8</sup> is the recognition of the inadequacy of the traditional hypothetico-deductive (H-D) schema as a characterization of the logic of science. According to this schema, we confirm a scientific hypothesis by deducing from it, in conjunction with suitable initial conditions and auxiliary hypotheses, an observational prediction that turns out to be true. The H-D method has a number of well-known shortcomings. (1) It does not take account of alternative hypotheses that might be invoked to explain the same prediction. (2) It makes no reference to the initial plausibility of the hypothesis being evaluated. (3) It cannot accommodate cases, such as the testing of statistical hypotheses, in which the observed outcome is not deducible from the hypothesis (in conjunction with the pertinent initial conditions and auxiliary hypotheses), but only rendered more or less probable.

In view of these and other considerations, many logical empiricists agreed with Kuhn regarding the inadequacy of hypothetico-deductive confirmation. A number—including Carnap and Reichenbach—appealed to Bayes's theorem, which may be written in the following form:

$$P(T|E.B) = \frac{P(T|B)P(E|B.T)}{P(T|B)P(E|B.T) + P(\sim T|B)P(E|B.\sim T)} \quad (1)$$

Let "T" stand for the theory or hypothesis being tested, "B" for our background information, and "E" for some new evidence we have just acquired. Then the expression on the left-hand side of the equation represents the probability of our hypothesis on the basis of the background information and the new evidence. This is known as the *posterior probability*. The right-hand side of the equation contains four probability expressions. Two of these,  $P(T|B)$  and  $P(\sim T|B)$ , are called *prior probabilities*; they represent the probability, on the basis of background information alone, without taking account of the new evidence E, that our hypothesis is true or false respectively. Obviously the two prior probabilities must add up to one; if the value of one of them is known, the value of the other can be inferred immediately. The remaining two probabilities,  $P(E|T.B)$  and  $P(E|\sim T.B)$ , are known as *likelihoods*; they are, respectively, the probability that the new evidence would occur if our hypothesis were true and the probability that it would occur if our hypothesis were false. The two likelihoods, in contrast to the two prior probabilities, must be established independently; the value of one

does not automatically determine the value of the other. To calculate the posterior probability of our hypothesis, then, we need three separate probability values to plug into the right-hand side of Bayes's theorem—a prior probability and two likelihoods.

Before attempting to resolve any important issues concerning the nature of scientific reasoning, let us look at a simple and noncontroversial application of Bayes's theorem. Consider a factory that produces can openers at the rate of 6,000 per day. This factory has two machines, a new one that produces 5,000 can openers per day and an old one that produces 1,000 per day. Among the can openers produced by the new machine 1 percent are defective; among those produced by the old machine 3 percent are defective. We pick one can opener at random from today's production and find it defective. What is the probability that it was produced by the new machine?

We can get the answer to this question via Bayes's theorem. If we let "B" stand for the class of can openers produced in this factory today, "T" for the class of can openers produced by the new machine, and "E" for a can opener that is defective, then the probability we seek is the posterior probability  $P(T|B.E)$ —the probability that a defective can opener from today's production was produced by the new machine. The values of the prior probabilities and likelihoods have been given, namely:

$$\begin{array}{ll} P(T|B) = 5/6 & P(\sim T|B) = 1/6 \\ P(E|T.B) = 1/100 & P(E|\sim T.B) = 3/100. \end{array}$$

Plugging these values into equation (1) immediately yields  $P(T|B.E) = 5/8$ . Notice that the old machine has a greater probability of producing a defective can opener than does the new, but the probability that a defective can opener was produced by the new machine is greater than that it was produced by the old one. This results, obviously, from the fact that the new machine produces so many more can openers overall than does the old one.

One way to look at this example is to consider the hypothesis T that a given can opener was produced by the new machine. This is a causal hypothesis. Our background information B is simply that the can opener is part of today's production at this factory. On the basis of this prior information, we can evaluate the prior probability of T; it is 5/6. Now we add to our knowledge about this can opener the information E that it is defective. This knowledge is relevant to the hypothesis that it was produced by the new machine; the posterior probability is 5/8. Although one does not *need* to appeal to Bayes's theorem to establish this result,<sup>9</sup> the highly artificial example shows clearly just how Bayes's theorem can be used to ascertain the posterior probability of a simple causal hypothesis.

When we come to more realistic scientific cases, it is not so easy to see how to apply Bayes's theorem; the prior probabilities may seem particularly difficult. I believe that, in fact, they reflect the plausibility arguments scientists often bring

to bear in their deliberations about scientific hypotheses. I shall discuss this issue in §4; indeed, in subsequent sections we shall have to take a close look at all of the probabilities that enter into Bayes's theorem.

In this section I have been concerned to present Bayes's theorem and to make a few preliminary remarks about its application to the problem of evaluating scientific hypotheses. In the next section I shall try to spell out the connections between Bayes's theorem and Kuhn's views on the nature of theory choice. Before moving on to that discussion, however, I want to present two other useful forms in which Bayes's theorem can be given. In the first place, because of the theorem on total probability:

$$P(E|B) = P(T|B) P(E|T.B) + P(\sim T|B) P(E|\sim T.B), \tag{2}$$

equation (1) can obviously be rewritten as:

$$P(T|E.B) = \frac{P(T|B) \times P(E|B.T)}{P(E|B)}. \tag{3}$$

In the second place, equation (1) can be generalized to handle several alternative hypotheses, instead of just one hypothesis and its negation, as follows:

$$P(T_i|B.E) = \frac{P(T_i|B) \times P(E|T_i.B)}{\sum_{j=1}^k [P(T_j|B) \times P(E|T_j.B)]}, \tag{4}$$

where  $T_1$ -  $T_k$  are mutually exclusive and exhaustive alternative hypotheses and  $1 \leq i \leq k$ .

Strictly speaking, (4) is the form that is needed for realistic historical examples—such as the corpuscular ( $T_1$ ) and wave ( $T_2$ ) theories of light in the nineteenth century. In that case, although we could construe  $T_1$  and  $T_2$  as mutually exclusive, we could not legitimately consider them exhaustive, for we cannot be sure that one or the other is true. Therefore, we would have to introduce  $T_3$ —what Abner Shimony has called the *catchall hypothesis*—which says that  $T_1$  and  $T_2$  are both false.  $T_1$ –  $T_3$  thus constitute a mutually exclusive and exhaustive set of hypotheses. This is the sort of situation that obtains when scientists are attempting to choose a correct hypothesis from among two or more serious candidates.

### §3. Kuhn and Bayes

For purposes of discussion, Kuhn is willing to admit that “each scientist chooses between competing theories by deploying some Bayesian algorithm which permits him to compute a value for  $P(T|E)$ , i.e., for the probability of the theory  $T$  on the evidence  $E$  available both to him and the other members of his professional group at a particular period of time.”<sup>10</sup> He then formulates the crucial issue in terms of the question of whether there is one unique algorithm used by all rational scientists, yielding a unique value for  $P$ , or whether different scien-

tists, though fully rational, may use different algorithms yielding different values of  $P$ . I want to suggest a third possibility to account for the phenomena of theory choice—namely, that many different scientists might use the same algorithm, but nevertheless arrive at different values of  $P$ .

When one speaks of a Bayesian algorithm, the first thought that comes to mind is Bayes's theorem itself, as embodied in any of the equations (1), (3) or (4). We have, for instance:

$$P(T|E.B) = \frac{P(T|B) \times P(E|B.T)}{P(E|B)}, \quad (3)$$

which constitutes an algorithm in the most straightforward sense of the term. Let us call  $P(E|B)$  the *expectedness* of the evidence. Given values for the prior probability, likelihood, and expectedness, the value of the posterior probability can be computed by trivial arithmetical operations.<sup>11</sup>

If we propose to use equation (3) as an algorithm, the obvious question is how to get values for the expressions on the right-hand side. Several answers are possible in principle, depending on what interpretation of the probability concept is espoused. If one adopts a Carnapian approach to inductive logic and confirmation theory, all of the probabilities that appear in Bayes's theorem can be derived a priori from the structure of the descriptive language and the definition of degree of confirmation. Since it is extremely difficult to see how any genuine scientific case could be handled by means of the highly restricted apparatus available within that approach, not many philosophers are tempted to follow this line. Moreover, even if a rich descriptive language were available, it is not philosophically tempting to suppose that the probabilities associated with serious scientific theories are a priori semantic truths.

Two major alternatives remain. First, one might maintain that the probabilities on the right-hand side of (3)—especially the prior probability  $P(T|B)$ —are objective and empirical. I have attempted to defend the view that they refer, at bottom, to the frequencies with which various kinds of hypotheses or theories have been found successful.<sup>12</sup> Clearly, enormous difficulties are involved in working out that alternative; I shall return to the issue below. In the meantime, let us consider the other—far more popular—alternative.

The remaining alternative approach involves the use of personal probabilities. Personal probabilities are subjective in character; they represent subjective degrees of conviction on the part of the individual who has them, provided that they fulfill the condition of coherence.<sup>13</sup> Consider a somewhat idealized situation. Suppose that, in the presence of background knowledge  $B$  (which may include initial conditions, boundary conditions, and auxiliary hypotheses) theory  $T$  deductively entails evidence  $E$ . This is the situation to which the hypothetico-

deductive method appears to be applicable. In this case,  $P(E|T.B)$  must equal 1, and equation (3) reduces to:

$$P(T|E.B) = P(T|B)/P(E|B). \quad (5)$$

One might then ask a particular scientist for his or her plausibility rating of theory T on background knowledge B alone, quite irrespective of whether evidence E obtains or not. Likewise, the same individual may be queried regarding the degree to which evidence E is to be expected irrespective of the truth or falsity of T. According to the personalist, it should be possible—by direct questioning or by some less direct method—to elicit such *psychological* facts regarding a scientist involved in investigations concerning the theory in question. This information is sufficient to determine the degree of belief this individual should have in the theory T given the background knowledge B and the evidence E, namely, the posterior probability  $P(T|E.B)$ .

In the more general case, when T and B do not deductively entail E, the procedure is the same, except that the value of  $P(E|T.B)$  must also be ascertained. In many contexts, where statistical significance tests can be applied, a value of the likelihood  $P(E|T.B)$  can be calculated, and the personal probability will coincide with the value thus derived. In any case, whether statistical tests apply or not, there is no *new* problem in principle involved in procuring the needed degree of confidence. This reflects the standard Bayesian approach in which all of the probabilities are taken to be personal probabilities.

In any case, whether one adopts an objective or a personalistic interpretation of probability, equation (3)—or some other version of Bayes's theorem—can be taken as an algorithm for evaluating scientific hypotheses or theories. Individual scientists, using the same algorithm, may arrive at different evaluations of the same hypothesis because they plug in different values for the probabilities. If the probabilities are construed as objective, different individuals may well have different estimates of these objective values. If the probabilities are construed as personal, different individuals may well have different subjective assessments of them. Bayes's theorem provides a mechanical algorithm, but the judgments of individual scientists are involved in procuring the values that are to be fed into it. This is a general feature of algorithms; they are not responsible for the data they are given.

#### §4. Prior Probabilities

In §2 I remarked that the prior probabilities in Bayes's theorem can best be seen as embodying the kinds of plausibility judgments that scientists regularly make regarding the hypotheses with which they are concerned. Einstein, who was clearly aware of this consideration, contrasted two points of view from which a theory can be criticized or evaluated:

The first point of view is obvious: the theory must not contradict empirical facts. . . . [it] is concerned with the confirmation of the theoretical foundation by the available empirical facts. The second point of view is not concerned with the relation of the material of observation but with the premises of the theory itself, with what may briefly but vaguely be characterized as the “naturalness” or “logical simplicity” of the premises. . . . The second point of view may briefly be characterized as concerning itself with the “inner perfection” of a theory, whereas the first point of view refers to the “external confirmation.”<sup>14</sup>

Einstein’s second point of view is the sort of thing I have in mind in referring to plausibility arguments or judgments concerning prior probabilities.

Plausibility considerations are pervasive in the sciences; they play a significant—indeed, *indispensable*—role. This fact provides the initial reason for appealing to Bayes’s theorem as an aid to understanding the logic of evaluating scientific hypotheses. Plausibility arguments serve to enhance or diminish the probability of a given hypothesis prior to—i.e., without reference to—the outcome of a particular observation or experiment. They are designed to answer the question, “Is this the kind of hypothesis that is likely to succeed in the scientific situation in which the scientist finds himself or herself?” On the basis of their training and experience, scientists are qualified to make such judgments.

This point can best be explained, I believe, in terms of concrete examples. Since before the time of Newton, for instance, a well-known plausibility argument for the inverse square character of gravitational forces has been around. It is natural to think of the gravitational force emanating from a particle of matter as one that spreads spherically from it in a uniform manner. In the seventeenth and eighteenth centuries all competent physical scientists believed that physical space has a three-dimensional Euclidean structure. Since the surface of a Euclidean sphere increases as the square of the radius, it is reasonable to suppose that the force of gravity is diluted in just the same way, for the farther one goes from the particle, the greater the spherical surface over which the force must be spread.

A famous Canadian study of the effects of the consumption of large doses of saccharin provides another example.<sup>15</sup> A statistically significant association between heavy saccharin consumption and bladder cancer in a controlled experiment with rats lends considerable plausibility to the hypothesis that use of saccharin as an artificial sweetener in diet soft drinks increases the risk of bladder cancer in humans. This example, unlike the preceding one, is inherently statistical and does not have even the *prima facie* appearance of a hypothetico-deductive inference.

In order to come to a clearer understanding of the nature of prior probabilities, it will be necessary to look at them from the point of view of the personalist and



that of the objectivist (frequency or propensity theorist).<sup>16</sup> The frightening thing about pure unadulterated personalism is that nothing prevents prior probabilities (and other probabilities as well) from being determined by all sorts of idiosyncratic and objectively irrelevant considerations. A given hypothesis might get an extremely low prior probability because the scientist considering it has a hang-over, has had a recent fight with his or her lover, is in passionate disagreement with the politics of the scientist who first advanced the hypothesis, harbors deep prejudices against the ethnic group to which the originator of the hypothesis belongs, etc. What we want to demand is that the investigator make every effort to bring all of his or her *relevant* experience in evaluating hypotheses to bear on the question of whether the hypothesis under consideration is of a type likely to succeed, and to leave aside emotional irrelevancies.

It is rather easy to construct really perverse systems of belief that do not violate the coherence requirement. But we need to keep in mind the objectives of science. When we have a long series of events, such as tosses of fair or biased coins, or radioactive decays of unstable nuclei, we want our subjective degrees of conviction to match what either a frequency theorist or a propensity theorist would regard as the objective probability. Carnap was profoundly correct in his notion that inductive or logical or epistemic probabilities should be reasonable estimates of relative frequencies.

A sensible personalist, I would suggest, is someone who wants his or her personal probabilities to reflect objective fact. Betting on a sequence of tosses of a coin, a personalist wants not only to avoid Dutch books,<sup>17</sup> but also to stand a reasonable chance of winning (or of not losing too much too fast). As I read it, the whole point of F. P. Ramsey's famous article on degrees of belief is to consider what you get if your subjective degrees of belief match the relevant frequencies.<sup>18</sup> One of the facts recognized by the sensible personalist is that whether the coin lands heads or tails is not affected by on which side of the bed he or she got out that morning. If we grant that the personalist's aim is to do as well as possible in betting on heads and tails, it would be obviously counterproductive to allow the betting odds to be affected by such irrelevancies.

The same general sort of consideration should be brought to bear on the assignment of probabilities to hypotheses. Whether a particular scientist is dyspeptic on a given morning is irrelevant to the question of whether a physical hypothesis that is under consideration is correct or not. Much more troubling, of course, is the fact that any given scientist may be inadvertently influenced by ideological or metaphysical prejudices. It is obvious that an unconscious commitment to capitalism or racism might seriously affect theorizing in the behavioral sciences.

Similar situations may arise in the physical sciences as well; another historical example will illustrate the point. In 1800 Alessandro Volta invented the battery, thereby providing scientists with a way of producing steady electrical currents.

It was not until 1820 that Hans Christian Oersted discovered the effect of an electrical current on a magnetic needle. Why was there such a delay? One reason was the previously established fact that a static electric charge has no effect on a magnetic needle. Another reason that has been mentioned is the fact that, contrary to the expectation if there were such an effect, it aligns the needle perpendicular to the current carrying wire. As Holton and Brush remark, "But even if one has currents and compass needles available, one does not observe the effect unless the compass is placed in the right position so that the needle can respond to a force that seems to act in a direction *around* the current rather than *toward* it."<sup>19</sup> I found it amusing when, on one occasion, a colleague set up the demonstration with the magnetic needle oriented at right angles to the wire to show why the experiment fails if one begins with the needle in that position. When the current was turned on, the needle rotated through 180 degrees; he had neglected to take account of polarity. How many times, between 1800 and 1820, had the experiment been performed without reversing the polarity? Not many. The experiment had apparently not been tried by others because of Cartesian metaphysical commitments. It was undertaken by Oersted as a result of his proclivities toward *naturphilosophie*.

How should scientists go about evaluating the prior probabilities of hypotheses? In elaborating a view he calls *tempered personalism*,—a view that goes beyond standard Bayesian personalism by placing further constraints on personal probabilities—Shimony<sup>20</sup> points out that experience shows that the hypotheses seriously advanced by serious scientists stand some chance of being successful. Science has, in fact, made considerable progress over the past four or five centuries, which constitutes strong empirical evidence that the probability of success among members of this class is nonvanishing. Likewise, hard experience has also taught us to reject claims of scientific infallibility. Thus, we have good reasons for avoiding the assignment of extreme values to the priors of the hypotheses with which we are seriously concerned. Moreover, Shimony reminds us, experience has taught that science is difficult and frustrating; consequently, we ought to assign fairly low prior probabilities to the hypotheses that have been explicitly advanced, allowing a fairly high prior for the catchall hypothesis—the hypothesis that we have not yet thought of the correct hypothesis. The history of science abounds with situations of choice among theories in which the successful candidate has not even been conceived at the time.

In *The Foundations of Scientific Inference*, I proposed that the problem of prior probabilities be approached in terms of an objective interpretation of probability, in particular, the frequency interpretation. I suggested three sorts of criteria that can be brought to bear in assessing the prior probabilities of hypotheses: formal, material, and pragmatic.

*Pragmatic criteria* have to do with the circumstances in which a new hypothesis originates. We have already seen an example of a pragmatic criterion in Shimony's observation that hypotheses advocated by serious scientists have non-

vanishing chances of success. The opposite side of the same coin is provided by Martin Gardner, who offers an enlightening characterization of scientific cranks.<sup>21</sup> Since it is doubtful that a single useful scientific suggestion has ever been originated by anyone in that category, hypotheses advanced by people of that ilk have negligible chances of being correct. I recall when L. Ron Hubbard's *Dianetics* was first published. A psychologist friend, asked what he thought of it, said, "I can't condemn this book before reading it, but after I have read it, I will." When competent scientists offer hypotheses outside of their areas of specialization, we have a right to wonder whether appreciable plausibility accrues to such suggestions. Hubbard was, incidentally, an engineer with no training in psychology.

The *formal criteria* have to do not only with matters of internal consistency of a new hypothesis, but also with relations of entailment or incompatibility of the new hypothesis with accepted laws and theories. The fact that Immanuel Velikovski's *Worlds in Collision*<sup>22</sup> contradicts many of the accepted basic laws of physics—e.g., the law of conservation of angular momentum—renders his 'explanations' of such biblically reported incidents as the parting of the waters of the Red Sea and the brief interruption of the rotation of the earth (the sun standing still) utterly implausible.

It should be recalled that among his five considerations for the evaluation of scientific theories—mentioned above—Kuhn includes consistency of the sort we are discussing. I take this as a powerful hint that one of the main issues Kuhn has raised about scientific theory choice involves the use of prior probabilities and plausibility judgments.

The *material criteria* have to do with the actual structure and content of the hypothesis or theory under consideration. The most obvious example is simplicity—another of Kuhn's five items. Simplicity strikes me as singularly important, for it has often been treated by scientists and philosophers as an a priori criterion. It has been suggested, for example, that the hypothesis that quarks are fundamental constituents of matter loses plausibility as the number of different types of quarks increases, since it becomes less simple as a result.<sup>23</sup> It has also been advocated as a universal methodological maxim: *Search for the simplest possible hypothesis*. Only if the simpler hypotheses do not stand up under testing should one resort to more complex hypotheses.

Although simplicity has obviously been an important consideration in the physical sciences, its applicability in the social/behavioral sciences is problematic. In a recent article, "Slips of the Tongue," Michael T. Motley criticizes Freud's theory for being too simple—an oversimplification.

Further still, the categorical nature of Freud's claim that all slips have hidden meanings makes it rather unattractive. It is difficult to imagine, for example, that my six-year-old daughter's mealtime request to "help cut up my meef" was

the result of repressed anxieties or anything of that kind. It seems more likely that she simply merged “meat” and “beef” into “meef.” Similarly, about the only meaning one can easily read into someone’s saying “roon mock” instead of “moon rock” is that the m and r got switched. Even so, how does it happen that words can merge or sounds can be switched in the course of speech production? And in the case of my “pleased to beat you” error [to a competitor for a job], might Freud have been right?<sup>24</sup>

The most reasonable way to look at simplicity, I think, is to regard it as a highly relevant characteristic, but one whose applicability varies from one scientific context to another. Specialists in any given branch of science make judgments about the degree of simplicity or complexity that is appropriate to the context at hand, and they do so on the basis of extensive experience in that particular area of scientific investigation. Since there is no precise measure of simplicity as applied to scientific hypotheses and theories, scientists must use their judgment concerning the degree of simplicity a given hypothesis or theory possesses and concerning the degree of simplicity that is desirable in the given context. The kind of judgment to which I refer is not spooky; it is the kind of judgment that arises on the basis of training and experience. This experience is far too rich to be the sort of thing that can be spelled out explicitly. As Patrick Suppes<sup>25</sup> has pointed out, the assignment of prior probability by the Bayesian can be regarded as the best estimate of the chances of success of the hypothesis or theory on the basis of all relevant experience in that particular scientific domain. The personal probability represents, not an effort to contaminate science with subjective irrelevancies, but rather an attempt to facilitate the inclusion of all relevant evidence.

Simplicity is only one among many material criteria. Another closely related criterion—frequently employed in contemporary physics—is symmetry. Perhaps the most striking historical example is de Broglie’s hypothesis regarding matter waves. Since light exhibits both particle and wave behavior, which are linked in terms of linear momentum, he suggested, why should not material particles, which obviously possess linear momentum, also have such wave characteristics as wave length and frequency? Unbeknownst to de Broglie, experimental work by Davisson was, at that very time, providing positive evidence of wavelike behavior of electrons.

A third widely used material criterion is analogy, as illustrated by the saccharin study. The physiological analogy between rats and humans is sufficiently strong to lend considerable plausibility to the hypothesis that saccharin can cause bladder cancer in humans. I suspect that the use of arguments by analogy in science is almost always aimed at establishing prior probabilities. The formal criteria enable us to take account of the ways in which a given hypothesis fits *deductively* with what else we know. Analogy helps us to assess the degree to which a given hypothesis fits *inductively* with what else we know.

The moral I would draw concerning prior probabilities is that they can be understood as our best estimates of the frequencies with which certain kinds of hypotheses succeed. These estimates are rough and inexact; some philosophers might prefer to think of them in terms of intervals. If, however, one wants to construe them as personal probabilities, there is no harm in it, as long as we attribute to the subject who has them the aim of bringing to bear all his or her experience that is relevant to the success or failure of hypotheses similar to that being considered. The personalist and the frequentist need not be in any serious disagreement over the construal of prior probabilities.<sup>26</sup>

One point is apt to be immediately troublesome. If we are to use Bayes's theorem to compute values of posterior probabilities, it would appear that we must be prepared to furnish numerical values for the prior probabilities. Unfortunately, it seems preposterous to suppose that plausibility arguments of the kind we have considered could yield exact numerical values. The usual answer is that, because of a phenomenon known as "washing out of the priors" or "swamping of the priors," even very crude estimates of the prior probabilities will suffice for the kinds of scientific judgments we are concerned to make. Obviously, however, this sort of convergence depends upon agreement regarding the likelihoods.

### §5. The Expectedness

The term " $P(E|B)$ " occurring in the denominator of equation (3) is called the *expectedness* because it is the opposite of surprisingness. The smaller the value of  $P(E|B)$ , the more surprising  $E$  is; the larger the value of  $P(E|B)$ , the less surprising, and hence, the more expected  $E$  is. Since the expectedness occurs in the denominator, a smaller value tends to increase the value of the fraction. This conforms to a widely held intuition that the more surprising the predictions a theory can make, the greater is their evidential value when they come true.

A classic example of a surprising prediction that came true is the Poisson bright spot. If we ask someone who is completely naive about theories of light how probable it is that a bright spot appears in the center of the shadow of a brightly illuminated circular object (ball or disk), we would certainly anticipate the response that it is very improbable indeed. There is a good inductive basis for this answer. In our everyday lives we have all observed many shadows of opaque objects, and they do not contain bright spots at their centers. Once, when I demonstrated the Poisson bright spot to an introductory class, one student carefully scrutinized the ball bearing that cast the shadow because he strongly suspected that it had a hole through it.

Another striking example, to my mind, is the Cavendish torsion-balance experiment. If we ask someone who is totally ignorant of Newton's theory of universal gravitation how strongly they expect to find a force of attraction between a lead ball and a pith ball in a laboratory, I should think the answer, again, would

be that it is very unlikely. There is, in this example as well, a sound inductive basis for the response. We are all familiar with the gravitational attraction of ordinary-size objects to the earth, but we do not have everyday experience of an attraction between two such relatively small (electrically neutral and unmagnetized) objects as those Cavendish used to perform his experiment. Newton's theory predicts, of course, that there will be a gravitational attraction between any two material objects. The trick was to figure out how to measure it.

As the foregoing two examples show, there is a possible basis for assigning a low value to the expectedness; it was made plausible by assuming that the subject was completely naive concerning the relevant physical theory. The trouble with this approach is that a person who wants to use Bayes's theorem – in the form of equation (3), say – cannot be totally innocent of the theory  $T$  that is to be evaluated, since the other terms in the equation refer explicitly to  $T$ . Consequently, we have to recognize the relationship between  $P(E|B)$  and the prior probabilities and likelihoods that appear on the right-hand side in the theorem on total probability:

$$P(E|B) = P(T|B) P(E|T.B) + P(\sim T|B) P(E|\sim T.B). \quad (2)$$

Suppose that the prior probability of  $T$  is not negligible and that  $T$ , in conjunction with suitable initial conditions, entails  $E$ . Under these circumstances  $E$  cannot be totally surprising; the expectedness cannot be vanishingly small. Moreover, to evaluate the expectedness of  $E$  we must also consider its probability if  $T$  is false. By focusing on the expectedness, we cannot really avoid dealing with likelihoods.

There is a further difficulty. Suppose, for example, that the wave theory of light is true. It is surely *true enough* in the context of the Poisson bright spot experiment. If we want to evaluate  $P(E|B)$  we must include in  $B$  the initial conditions of the experiment – the circular object illuminated by a bright light in such a way that the shadow falls upon a screen. Given the truth of the wave theory, the *objective probability* of the bright spot is one, for whenever those initial conditions are realized, the bright spot appears. It makes no difference whether we know that the wave theory is true, or believe it, or reject it, or have ever thought of it. Under the conditions specified in  $B$  the bright spot invariably occurs. Interpreted either as a frequency or a propensity,  $P(E|B) = 1$ . If we are to avoid trivialization in many important cases, the expectedness must be treated as a personal probability. To anyone who, like me, wants to base scientific theory preference or choice on objective considerations, this result poses a serious problem.

The net result is a twofold problem. First, by focusing on the expectedness, we do *not* escape the need to deal explicitly with the likelihoods. In §6 I shall discuss the difficulties that arise when we focus on the likelihoods, especially the problem of the likelihood on the catchall hypothesis. Second, the expectedness defies interpretation as an objective probability. In §7 I shall propose a strategy for avoiding involvement with either the expectedness or the likelihood on the

catchall. That maneuver will, I hope, keep open the possibility of an objective basis for the evaluation of scientific hypotheses.

## §6. Likelihoods

Equations (1), (3), and (4) are different forms of Bayes's theorem, and each of them contains a likelihood,  $P(E|T.B)$ , in the numerator. Two trivial cases can be noted at the outset. First, if the conjunction of theory  $T$  and background knowledge  $B$  are logically incompatible with evidence  $E$ , the likelihood equals zero, and the posterior probability,  $P(T|E.B)$ , automatically becomes zero.<sup>27</sup> Second, as we have already noticed, if  $T.B$  entails  $E$ , that likelihood equals one, and consequently drops out, as in equation (5).

Another easy case occurs when the hypothesis  $T$  involves various kinds of randomness assumptions, for example, the independence of a series of trials on a chance setup.<sup>28</sup> Consider, for example, the case of a coin that has been tossed 100 times, with the result that heads showed in 63 cases and tails in 37. We assume that the tosses are independent, but we are concerned whether the system consisting of the coin and tossing mechanism is biased. Calculation shows that the probability, given an unbiased coin and tossing mechanism, of the actual frequency of heads differing from  $1/2$  by 20 percent or more on 100 tosses (i.e., falling outside of the range 40 to 60) is about .05. Thus, the likelihood of the outcome on the hypothesis that the coin and mechanism are fair is less than .05. On the hypothesis that the coin has a 60 to 40 bias for heads, by contrast, the probability that the number of heads in 100 trials differs from  $6/10$  by less than 20 percent (i.e., lies within the 48 to 72 range) is well above .95. These are the kinds of likelihoods that would be used to compare the *null hypothesis* that the coin is fair with the hypothesis that it has a certain bias.<sup>29</sup> This example typifies a wide variety of cases, including the above-mentioned controlled experiment on rats and saccharin, in which statistical significance tests are applied. These yield a comparison between the probability of the observed result if the hypothesis is correct and the probability of the same result on a null hypothesis.

In still another kind of situation the likelihood  $P(E|T.B)$  is straightforward. Consider, for example, the case in which a physician takes an X ray for diagnostic purposes. Let  $T$  be the hypothesis that the patient has a particular disease and let  $E$  be a certain appearance on the film. From long medical experience it may be known that  $E$  occurs in 90 percent of all cases in which that disease is present. In many cases, as this example suggests, there may be accumulated frequency data from which the value of  $P(E|T.B)$  can be derived.

Unfortunately, life with likelihoods is not always as simple as the foregoing cases suggest. Consider an important case, which I will present in a highly un-historical way. In comparing the Copernican and Ptolemaic cosmologies, it is easy to see that the phases of Venus are critical. According to the Copernican sys-

tem, Venus should exhibit a broad set of phases from a narrow crescent to an almost full disk. According to the Ptolemaic system, Venus should always present nearly the same crescent-shaped appearance. One of Galileo's celebrated telescopic observations was of the phases of Venus. The likelihood of such evidence on the Copernican system is unity; on the Ptolemaic it is zero. This is the decisive sort of case that we cherish.

The Copernican system did, however, face one serious obstacle. On the Ptolemaic system, because the earth does not move, the fixed stars should not appear to change their positions. On the Copernican system, because the earth makes an annual trip around the sun, the fixed stars should appear to change their positions in the course of the year. The very best astronomical observations, including those of Tycho Brahe, failed to reveal any observable stellar parallax.<sup>30</sup> However, it was realized that, if the fixed stars are at a very great distance from the earth, stellar parallax, though real, would be too small to be observed. Consequently, the likelihood  $P(E|T.B)$ , where T is the Copernican system and E the absence of observable stellar parallax, is not zero. At the time of the scientific revolution, prior to the advent of Newtonian mechanics, there seemed no reasonable way to evaluate this likelihood. The assumption that the fixed stars are almost unimaginably distant from the earth was a highly ad hoc, and consequently implausible, auxiliary hypothesis to adopt just to save the Copernican system. Among other things, Christians did not like the idea that heaven was so very far away.

The most reasonable resolution of this anomaly was offered by Tycho Brahe, whose cosmology placed the earth at rest, with the sun and moon moving in orbits around the earth, but with all of the other planets moving in orbits around the sun. In this way both the observed phases of Venus and the absence of observable stellar parallax could be accommodated. Until Newton's dynamics came upon the scene, it seems to me, Tycho's system was clearly the best available theory.

In §2 I suggested that the following form of Bayes's theorem is the most appropriate for use in actual scientific cases in which more than one hypothesis is available for serious consideration:

$$P(T_i|B.E) = \frac{P(T_i|B) \times P(E|T_i.B)}{\sum_{j=1}^k [P(T_j|B) \times P(E|T_j.B)]} \quad (4)$$

It certainly fits the foregoing example in which we compared the Ptolemaic, Copernican, and Tyconic systems. This equation involves a mutually exclusive and exhaustive set of hypotheses  $T_1, \dots, T_{k-1}, T_k$ , where  $T_1 - T_{k-1}$  are seriously entertained and  $T_k$  is the catchall. Thus, the scientist who wants to calculate the posterior probability of one particular hypothesis  $T_i$  on the basis of evidence E must ascertain likelihoods of three types: (1) the probability of evidence E given  $T_i$ , (2) the probability of that evidence on each of the other seriously considered



alternatives  $T_j$  ( $j \neq i, j \neq k$ ), and (3) the probability of that evidence on the catchall  $T_k$ .

In considering the foregoing example, I suggested that, although likelihoods in the first two categories are sometimes straightforward, there are cases in which they turn out to be quite problematic. We shall look at more examples in which they present difficulties as our discussion proceeds. But the point to be emphasized right now is the utter intractability of the likelihood on the catchall. The reason for this difficulty is easy to see. Whereas the seriously considered candidates are bona fide hypotheses, the catchall is a hypothesis only in a Pickwickian sense. It refers to all of the hypotheses we are *not* taking seriously, including all those that have not been thought of as yet; indeed, the catchall is logically equivalent to their disjunction. These will often include brilliant discoveries in the future history of science that will eventually solve our most perplexing problems.

Among the hypotheses hidden in the catchall are some that, in conjunction with present available background information, entail the present evidence  $E$ . On such as-yet-undiscovered hypotheses the likelihood is one. Obviously, however, the fact that its probability on one particular hypothesis is unity does not entail anything about its probability on some disjunction containing that hypothesis as one of its disjuncts. These considerations suggest to me that the likelihood on the catchall is totally intractable. To try to evaluate the likelihood on the catchall involves, it seems to me, an attempt to guess the future history of science. That is something we cannot do with any reliability.

In any situation in which a small number of theories are competing for ascendancy it is tempting, though quite illegitimate, simply to ignore the likelihood on the catchall. In the nineteenth century, for instance, scientists asked what the probability of a given phenomenon is on the wave theory of light and what it is on the corpuscular theory. They did not seriously consider its probability if neither of these theories is correct. Yet we see, from the various forms in which Bayes's theorem is written, that either the expectedness or the likelihood on the catchall is an indispensable ingredient. In the next section I shall offer a *legitimate* way of eliminating those probabilities from our consideration.

## §7. Choosing Between Theories

Kuhn has often maintained that in actual science the problem is never to evaluate one particular hypothesis or theory in isolation; it is always a matter of choosing from among two or more viable alternatives. He has emphasized that an old theory is never completely abandoned unless there is currently available a rival to take its place. Given that circumstance, it is a matter of choosing between the old and the new. On this point I think that Kuhn is quite right, especially as regards reasonably mature sciences. And this insight provides a useful clue on how to use Bayes's theorem to explicate the logic of scientific confirmation.

Suppose that we are trying to choose between  $T_1$  and  $T_2$ , where there may or may not be other serious alternatives in addition to the catchall. By letting  $i = 1$  and  $i = 2$ , we can proceed to write equation (4) for each of these candidates. Noting that the denominators of the two are identical, we can form their ratio as follows:

$$\frac{P(T_1|E.B)}{P(T_2|E.B)} = \frac{P(T_1|B) P(E|T_1.B)}{P(T_2|B) P(E|T_2.B)}. \quad (6)$$

No reference to the catchall hypothesis appears in this equation. Since the catchall is not a bona fide hypothesis, it is not a contender, and we need not try to calculate its posterior probability. The use of equation (6) frees us from the need to deal either with the expectedness of  $E$  or with its probability on the catchall.

Equation (6) yields a relation that can be regarded as a *Bayesian algorithm for theory preference*. Suppose that, prior to the emergence of evidence  $E$ , you prefer  $T_1$  to  $T_2$ ; that is,  $P(T_1|B) > P(T_2|B)$ . Then  $E$  becomes available. You should change your preference in the light of  $E$  if and only if  $P(T_2|E.B) > P(T_1|E.B)$ . From (6) it follows that

$$P(T_2|E.B) > P(T_1|E.B) \text{ iff } P(E|T_2.B)/P(E|T_1.B) > P(T_1|B)/P(T_2|B). \quad (7)$$

In other words, you should change your preference to  $T_2$  if the ratio of the likelihoods is greater than the reciprocal of the ratio of the respective prior probabilities. A corollary is that, if both  $T_1.B$  and  $T_2.B$  entail  $E$ , so that:

$$P(E|T_1.B) = P(E|T_2.B) = 1,$$

the occurrence of  $E$  can never change the preference rating between the two competing theories.

At the end of §4 I made reference to the well-known phenomenon of washing out of priors in connection with the use of Bayes's theorem. One might well ask what happens to this swamping when we switch from Bayes's theorem to the ratio embodied in equation (6).<sup>31</sup> The best answer, I believe, is this. If we are dealing with two hypotheses that are serious contenders in the sense that they do not differ too greatly in plausibility, the ratio of the priors will be of the order of unity. If, as the observational evidence accumulates, the likelihoods come to differ greatly, the ratio of the likelihoods will swamp the ratio of the priors. Recall the example of the tossed coin. Suppose we consider the prior probability of a fair device to be ten times as large as that of a biased device. If about the same proportion of heads occurs in 500 tosses as occurred in the aforementioned 100, the likelihood on the null hypothesis would be virtually zero and the likelihood on the hypothesis that the device has a bias approximating the observed frequency would be essentially indistinguishable from unity. The ratio of prior probabilities would obviously be completely dominated by the likelihood ratio.

## §8. Plausible Scenarios

Although, by appealing to equation (6), we have eliminated the need to deal with the expectedness or the likelihood on the catchall, we cannot claim to have dealt adequately with the likelihoods on the hypotheses we are seriously considering, for their values are not always straightforwardly ascertainable. We have already mentioned one example, namely, the probability of absence of observable stellar parallax on the Copernican hypothesis. We noted that, by adding an auxiliary hypothesis to the effect that the fixed stars are located an enormous distance from the Earth, we could augment the Copernican hypothesis in such a way that the likelihood on this augmented hypothesis is one. But, for many reasons, this auxiliary assumption could hardly be considered plausible in that historical context. By now, of course, we have measured the parallax of relatively nearby stars, and from those values have calculated these distances. They are extremely far from us in comparison to the familiar objects in our solar system.

Consider another well-known example. During the seventeenth and eighteenth centuries the wave and corpuscular theories of light received considerable scientific attention. Each was able to explain certain important optical phenomena, and each faced fundamental difficulties. The corpuscular hypothesis easily explained how light could travel vast distances through empty space, and it readily explained sharp shadows. The theory of light as a longitudinal wave explained various kinds of diffraction phenomena, but failed to deal adequately with polarization. When, early in the nineteenth century, light was conceived as a transverse wave, the wave theory explained polarization as well as diffraction quite straightforwardly. And Huygens had long since shown how the wave theory could handle rectilinear propagation and sharp shadows. For most of the nineteenth century the wave theory dominated optics.

The proponent of the particle theory could still raise a serious objection. What is the likelihood of a wave propagating in empty space? Lacking a medium, the answer is zero. So wave theorists augmented their theory with the auxiliary assumption that all of space is filled with a peculiar substance known as the *luminiferous ether*. This substance was postulated to have precisely the properties required to transmit light waves.

The process I have been describing can appropriately be regarded as the discovery and introduction of *plausible scenarios*. A theory is confronted with an *anomaly*—a phenomenon that appears to have a small, possibly zero, likelihood given that theory. Proponents of the theory search for some auxiliary hypothesis that, if conjoined to the theory, renders the likelihood high, possibly unity. This move shifts the burden of the argument to the plausibility of the new auxiliary hypothesis. I mentioned two instances involved in the wave theory of light. The first was the auxiliary assumption that the wave is transverse. This modification of the theory was sufficiently plausible to be incorporated as an integral part of

the theory. The second was the luminiferous ether. The plausibility of this auxiliary hypothesis was debated throughout the nineteenth, and into the twentieth, century. The ether had to be dense enough to transmit transverse waves (which require a denser medium than do longitudinal waves) and thin enough to allow astronomical bodies to move through it without noticeable diminution of speed. Attempts to detect the motion of the earth relative to the ether were unsuccessful. The Lorentz-Fitzgerald contraction hypothesis was an attempt to save the ether theory—that is, another attempt at a plausible scenario—but it was, of course, abandoned in favor of special relativity.

I am calling these auxiliaries *scenarios* because they are stories about how something could have happened, and *plausible* because they must have some degree of acceptability if they are to be of any help in handling problematic phenomena. The wave theory could handle the Poisson bright spot by deducing it from the theory. There seemed to be no plausible scenario available to the particle theory that could deal with this phenomenon. The same has been said with respect to Foucault's demonstration that the velocity of light is greater in air than it is in water.<sup>32</sup>

One nineteenth century optician of considerable importance who did not adopt the wave theory, but remained committed to the Newtonian emission theory, was David Brewster.<sup>33</sup> In a "Report on the Present State of Physical Optics," presented to the British Association for the Advancement of Science in 1831, he maintained that the undulatory theory is "still burthened with difficulties and cannot claim our implicit assent."<sup>34</sup> Brewster freely admitted the unparalleled explanatory and predictive success of the wave theory; nevertheless, he considered it false.

Among the difficulties Brewster found with the wave theory, two might be mentioned. First, he considered the wave theory implausible, for the reason that it required "an *ether* invisible, intangible, imponderable, inseparable from all bodies, and extending from our own eye to the remotest verge of the starry heavens."<sup>35</sup> History has certainly vindicated him on that issue. Second, he found the wave theory incapable of explaining a phenomenon that he had discovered himself, namely, *selective absorption*—dark lines in the spectrum of sunlight that has passed through certain gases. Brewster points out that a gas may be opaque to light of one particular index of refraction in flint glass, while transmitting freely light whose refractive indices in the same glass are only the tiniest bit higher or lower. Brewster maintained that there was no plausible scenario the wave theorists could devise that would explain why the ether permeating the gas transmits two waves of very nearly the same wave length, but does not transmit light of a very precise wave length lying in between:

There is no fact analogous to this in the phenomena of sound, and I can form no conception of a simple elastic medium so modified by the particles of the

body which contains it, as to make such an extraordinary selection of the undulations which it stops or transmits. . . . <sup>36</sup>

Brewster never found a plausible scenario by means of which the Newtonian theory he favored could cope with absorption lines, nor could proponents of the wave theory find one to bolster their viewpoint. Dark absorption lines remained anomalous for both the wave and particle theories; neither could see a way to furnish them with high likelihood.

With hindsight we can say that the catchall hypothesis was looking very strong at this point. We recognize that the dark absorption lines in the spectrum of sunlight are closely related to the discrete lines in the emission spectra of gases, and that they, in turn, are intimately bound up with the problem of the stability of atoms. These phenomena played a major role in the overthrow of classical physics at the turn of the twentieth century.

I have introduced the notion of a plausible scenario to deal with problematic likelihoods. Likelihoods can cause trouble for a scientific theory for either of two reasons. First, if you have a pet theory that confers an extremely small—for all practical purposes zero—likelihood on some observed phenomenon, that is a problem for that favored theory. You try to come up with a plausible scenario according to which the likelihood will be larger—ideally, unity. Second, if there seems to be no way to evaluate the likelihood of a piece of evidence with respect to some hypothesis of interest, that is another sort of problem. In this case, we search for a plausible scenario that will make the likelihood manageable, whether this involves assigning it a high, medium, or low value.

What does this mean in terms of the Bayesian approach I am advocating? Let us return to:

$$\frac{P(T_1|E.B)}{P(T_2|E.B)} = \frac{P(T_1|B) P(E|T_1.B)}{P(T_2|B) P(E|T_2.B)}, \tag{6}$$

which contains two likelihoods. Suppose, as in nineteenth-century optics, that both likelihoods are problematic. As we have seen, we search for plausible scenarios  $A_1$  and  $A_2$  to augment  $T_1$  and  $T_2$  respectively. If the search has been successful, we can assess the likelihoods of  $E$  with respect to the augmented theories  $A_1.T_1$  and  $A_2.T_2$ . Consequently, we can modify (6) so as to yield

$$\frac{P(A_1.T_1|E.B)}{P(A_2.T_2|E.B)} = \frac{P(A_1.T_1|B) P(E|A_1.T_1.B)}{P(A_2.T_2|B) P(E|A_2.T_2.B)}. \tag{8}$$

In order to use this equation to compare the posterior probabilities of the two augmented theories, we must assess the plausibilities of the scenarios, for the prior probabilities of both augmented theories— $A_1.T_1$  and  $A_2.T_2$ —appear in it. In §4 I tried to explain how prior probabilities can be handled—that is, how we can obtain at least rough estimates of their values. If, as suggested, the plausible

scenarios have made the likelihoods ascertainable, then we can use them in conjunction with our determinations of the prior probabilities to assess the ratio of the posterior probabilities. We have, thereby, handled the central issue raised by Kuhn, namely, what is the basis for preference between two theories.<sup>37</sup> Equation (8) is a Bayesian algorithm.

If either augmented theory, in conjunction with background knowledge B, entails E, then the corresponding likelihood is one and it drops out of (8). If both likelihoods drop out we have the special case in which:

$$\frac{P(A_1.T_1|E.B)}{P(A_2.T_2|E.B)} = \frac{P(A_1.T_1|B)}{P(A_2.T_2|B)}, \quad (9)$$

thereby placing the *whole* burden on the prior probabilities—the plausibility considerations. Equation (9) represents a simplified Bayesian algorithm that is applicable in this type of special case.

Another type of special case was mentioned above. If, as in our coin tossing example, the values of the prior probabilities do not differ drastically from one another, but the likelihoods become widely divergent as the observational evidence accumulates, there will be a washing out of the priors. In this case, the ratio of the posterior probabilities equals, for practical purposes, the ratio of the likelihoods.

The use of either (8) or (9) as an algorithm for theory choice does not imply that all scientists will agree on the numerical values or prefer the same theory. The evaluation of prior probabilities clearly demands the kind of scientific judgment whose importance Kuhn has rightly insisted upon. It should also be clearly remembered that these formulas provide no evaluations of individual theories; they furnish only comparative evaluations. Thus, instead of yielding a prediction regarding the chances of one particular theory being a component of “completed science,” they compare existing theories with regard to their present merits.

## §9. Kuhn’s Criteria

Early in this paper I quoted five criteria that Kuhn mentioned in connection with his views on the rationality and objectivity of science. The time has come to relate them explicitly to the Bayesian approach I have been attempting to elaborate. In order to appreciate the significance of these criteria it is important to distinguish three aspects of scientific theories that may be called *informational virtues*, *confirmational virtues*, and *economic virtues*. Up to this point we have concerned ourselves almost exclusively with confirmation, for our use of Bayes’s theorem is germane only to the confirmational virtues. But since Kuhn’s criteria patently refer to the other virtues as well, we must also say a little about them.

Consider, for example, the matter of *scope*. Newton’s three laws of motion and his law of universal gravitation obviously have greater scope than the conjunction

of Galileo's law of falling bodies and Kepler's three laws of planetary motion. This means, simply, that Newtonian mechanics contains more information than the laws of Kepler and Galileo taken together. Given a situation of this sort, we prefer the more informative theory because it is a basic goal of science to increase our knowledge as much as possible. We might, of course, hesitate to choose a highly informative theory if the evidence for it were extremely limited or shaky, because the desire to be right might overrule the desire to have more information content. But in the case at hand that consideration does not arise.

In spite of its intuitive attraction, however, the appeal to scope is not altogether unproblematic. There are two ways in which we might construe the Galileo-Kepler-Newton example of the preceding paragraph. First, we might ignore the small corrections mandated by Newton's theory in the laws of Galileo and Kepler. In that case we can clearly claim greater scope for Newton's laws than for the conjunction of Galileo's and Kepler's laws, since the latter is entailed by the former but not conversely. Where an entailment relation holds we can make good sense of comparative scope.

Kuhn, however, along with most of the historically oriented philosophers, has been at pains to deny that science progresses by finding more general theories that include earlier theories as special cases. Theory choice or preference involves *competing* theories that are *mutually incompatible* or *mutually incommensurable*. To the best of my knowledge Kuhn has not offered any precise characterization of scope; Karl Popper, in contrast, has made serious attempts to do so. In response to Popper's efforts, Adolf Grünbaum has effectively argued that none of the Popperian measures can be usefully applied to make comparisons of scope among mutually incompatible competing theories.<sup>38</sup> Consequently, the concept of scope requires fundamental clarification if we are to use it to understand preferences among competing theories. However, since scope refers to information rather than confirmation, it plays no role in the Bayesian program I have been endeavoring to explicate. We can thus put aside the problem of explicating that difficult concept.

Another of Kuhn's criteria is *accuracy*. It can, I think, be construed in two different ways. The first has to do with informational virtues; the second with economic. On the one hand, two theories might both make true predictions regarding the same phenomena, but one of them might give us precise predictions where the other gives only predictions that are less exact. If, for example, one theory enables us to predict that there will be a solar eclipse on a given day, and that its path of totality will cross North America, it may well be furnishing correct information about the eclipse. If another theory gives not only the day, but also the time, and not only the continent, but also the precise boundaries, the second provides much more information, at least with respect to this particular occurrence. It is not that either is incorrect; rather, the second yields more knowledge than the first. However, it should be clearly noted—as it was in the case of

scope—that these theories are not incompatible or incommensurable competitors (at least with respect to this eclipse), and hence do not illustrate the interesting type of theory preference with which Kuhn is primarily concerned.

On the other hand, one theory may yield predictions that are nearly, but not quite, correct, while another theory yields predictions that are entirely correct—or, at least, more nearly correct. Newtonian astrophysics does well in ascertaining the orbit of the earth, but general relativity introduces a correction of 3.8 seconds of arc per century in the precession of its perihelion.<sup>39</sup> Although the Newtonian theory is literally false, it is used in contexts of this sort because its inaccuracy is small, and the economic gain involved in using it instead of general relativity (the saving in computational effort) is enormous.

The remaining three criteria are *simplicity*, *consistency*, and *fruitfulness*; all of them have direct bearing upon the confirmational virtues. In the treatment of prior probabilities in §4, I briefly mentioned simplicity as a factor having a significant bearing upon the plausibility of theories. More examples could be added, but I think the point is clear.

In the same section I also made passing reference to consistency, but more can profitably be said on that topic. Consistency has two aspects, internal consistency of a theory and its compatibility with other accepted theories. While scientists may be fully justified in *entertaining* collections of statements that contain contradictions, the goal of science is surely to accept only logically consistent theories.<sup>40</sup> The discovery of an internal inconsistency has a distinctly adverse effect on the prior probability of that theory, to wit, it must go straight to zero.

When we consider the relationships of a given theory to other accepted theories we again find two aspects. There are *deductive* relations of entailment and incompatibility, and there *inductive* relations of fittingness and incongruity. The deductive relations are quite straightforward. Incompatibility with an accepted theory makes for implausibility; being a logical consequence of an accepted theory makes for a high prior probability. Although deductive subsumption of narrower theories under broader theories is probably something of an oversimplification of actual cases, nevertheless, the ability of an overarching theory to deductively unify diverse domains furnishes a strong plausibility argument.

When it comes to the inductive relations among theories, analogy is, I think, the chief consideration. I have already mentioned the use of analogy in inductively transferring results of experiments from rats to humans. In archaeology, the method of ethnographic analogy, which exploits similarities between extant primitive societies and prehistoric societies, is widely used. In physics, the analogy between the inverse square law of electrostatics and the inverse square law of gravitation provides an example of an important plausibility consideration.

Kuhn's criteria of consistency (broadly construed) and simplicity seem clearly to pertain to assessments of the prior probabilities of theories. They cry out for a Bayesian interpretation.



The final criterion in Kuhn's list is *fruitfulness*; it has many aspects. Some theories prove fruitful by unifying a great many apparently different phenomena in terms of a few simple principles. The Newtonian synthesis is, perhaps, the outstanding example; Maxwellian electrodynamics is also an excellent case. As I suggested above, this ability to accommodate a wide variety of facts tends to enhance the prior probability of a given theory. To attribute diverse success to happenstance, rather than basic correctness, is implausible.

Another sort of fertility involves the predictability of theretofore unknown phenomena. We might mention as familiar illustrations the prediction of the Poisson bright spot by the wave theory of light and the prediction of time dilation by special relativity. These are the kinds of instances in which, in an important sense, the expectedness is low. As we have noted, a small expectedness tends to increase the posterior probability of a hypotheses.

A further type of fertility relates directly to plausible scenarios; a theory is fruitful in this way if it successfully copes with difficulties with the aid of suitable auxiliary assumptions. Newtonian mechanics again provides an excellent example. The perturbations of Uranus were explained by postulating Neptune. The perturbations of Neptune were explained by postulating Pluto.<sup>41</sup> The motions of stars within galaxies and of galaxies within clusters are explained in terms of *dark matter*, concerning which there are many current theories. A theory that readily gives rise to plausible scenarios to deal with problematic likelihoods can boast this sort of fertility.

The discussion of Kuhn's criteria in this section is intended to show how adequately they can be understood within a Bayesian framework—insofar as they are germane to confirmation. If it is sound, we have constructed a fairly substantial bridge connecting Kuhn's views on theory choice with those of the logical empiricists—at least, those who find in Bayes's theorem a suitable schema for characterizing the confirmation of hypotheses and theories.

## §10. Rationality vs. Objectivity

In the title of this essay I have used both the concept of *rationality* and that of *objectivity*. It is time to say something about their relationship. Perhaps the best way to approach the distinction between them is to enumerate various grades of rationality. In a certain sense one can be rational without paying any heed at all to objectivity. It is essentially a matter of good housekeeping as far as one's beliefs and degrees of confidence are concerned. As Bayesians have often emphasized, it is important to avoid logical contradictions in one's beliefs and to avoid probabilistic incoherence in one's degrees of conviction. If contradiction or incoherence are discovered, they must somehow be eliminated; the presence of either constitutes a form of irrationality. But the removal of such elements of irrationality can be accomplished without any appeal to facts outside of the subject's corpus

of beliefs and degrees of confidence. To achieve this sort of rationality is to achieve a minimal standard that I have elsewhere called *static rationality*.<sup>42</sup>

One way in which additional facts may enter the picture is via Bayes's theorem. We have a theory T in which we have a particular degree of confidence. A new piece of evidence turns up—some objective fact E of which we were previously unaware—and we use Bayes's theorem to calculate a posterior probability of T. To accept this value of the posterior probability as one's degree of confidence in T is known as *Bayesian conditionalization*. Use of Bayes's theorem does not, however, guarantee objectivity. If the resulting posterior probability of T is one we are not willing to accept, we can make adjustments elsewhere to avoid incoherence. After all, the prior probabilities and likelihoods are simply personal probabilities, so they can be adjusted to achieve the desired result. If, however, the requirement of *Bayesian conditionalization* is added to those of static rationality we have a stronger type of rationality that I have called *kinematic*.<sup>43</sup>

The highest grade of rationality—what I have called *dynamic rationality*—requires much fuller reference to objective fact than is demanded by advocates of personalism. The most obvious way to inject a substantial degree of objectivity into our deliberations regarding choices of scientific theories is to provide an objective interpretation of the probabilities in Bayes's theorem. Throughout this discussion I have adopted that approach as thoroughly as possible. For instance, I have argued that prior probabilities can be given an objective interpretation in terms of frequencies of success. I have tried to show how likelihoods could be objective—by virtue of entailment relations, tests of statistical significance, or observed frequencies. When the likelihoods created major difficulties, I appealed to plausible scenarios. The result was that an intractable likelihood could be exchanged for a tractable prior probability—namely, the prior probability of a theory in conjunction with an auxiliary assumption.

We noted that the denominators of the right-hand sides of the various versions of Bayes's theorem—equations (1), (3), and (4)—contain either an expectedness or a likelihood on the catchall. It seems to me futile to try to construe either of these probabilities objectively. Consequently, in §7 I introduced equation (6), which involves a ratio of two instances of Bayes's theorem, and from which the expectedness and the likelihood on the catchall drop out. Confining our attention, as Kuhn recommends, to comparing the merits of competing theories, rather than offering absolute evaluations of individual theories, we were able to eliminate the probabilities that most seriously defy objective interpretation.

## §11. Conclusions

For many years I have been convinced that plausibility arguments in science have constituted a major stumbling block to an understanding of the logic of scientific inference. Kuhn was not alone, I believe, in recognizing that considera-

tions of plausibility constitute an essential aspect of scientific reasoning, without seeing where they fit into the logic of science. If one sees confirmation solely in terms of the crude hypothetico-deductive method, there is no place for them. There is, consequently, an obvious incentive for relegating plausibility considerations to heuristics. If one accepts the traditional distinction between the *context of discovery* and the *context of justification*, it is tempting to place them in the former context. But Kuhn recognized, I think, that plausibility arguments enter into the justifications of choices of theories, with the result that he became skeptical of the value of that distinction. If, as I believe, plausibility considerations are simply evaluations of prior probabilities of hypotheses or theories, then it becomes apparent via Bayes's theorem that they play an indispensable role in the context of justification. We do not need to give up that important distinction.

At several places in this paper I have spoken of Bayesian algorithms, mainly because Kuhn introduced that notion into the discussion. I have claimed that such algorithms exist—and attempted to exhibit them—but I accord *very little* significance to that claim. The algorithms are trivial; what is important is the scientific judgment involved in assessing the probabilities that are fed into the equations. The algorithms give frameworks in terms of which to understand the role of the sort of judgment upon which Kuhn rightly placed great emphasis. —

The history of science chronicles the successes and failures of attempts at scientific theorizing. If the Bayesian analysis I have been offering is at all sound, history of science—in addition to contemporary scientific experience, of course—provides a rich source of information relevant to the prior probabilities of the theories among which we are at present concerned to make objective and rational choices. This viewpoint captures, I believe, the point Kuhn made at the beginning of his first book:

But an age as dominated by science as our own does need a perspective from which to examine the scientific beliefs which it takes so much for granted, and history provides one important source of such perspective. If we can discover the origins of some modern scientific concepts and the way in which they supplanted the concepts of an earlier age, we are more likely to evaluate intelligently their chances for survival.<sup>44</sup>

I suggested at the outset that an appeal to Bayesian principles could provide some aid in bridging the gap between Hempel's logical-empiricist approach and Kuhn's historical approach. I hope I have offered a convincing case. However that may be, there remain many unresolved issues. For instance, I have not even broached the problem of incommensurability of paradigms or theories. This is a major issue. For another example, I have assumed uncritically throughout the discussion that the various parties to disputes about theories share a common body B of background knowledge. It is by no means obvious that this is a tenable assumption. No doubt other points for controversy remain. I do not for a moment

maintain that complete consensus would be in the offing even if both camps were to buy the Bayesian line I have been peddling. But I do hope that some areas of misunderstanding have been clarified.

### Notes

1. Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962; 2d ed., 1970). "Postscript—1969," added to the second edition, contains discussions of some of the major topics that are treated in the present essay.

2. Such philosophers are often characterized by their opponents as *logical positivists*, but this is an egregious historical inaccuracy. Although some of them had been members of or closely associated with the Vienna Circle in their earlier years, none of them retained the early positivistic commitment to phenomenalism and/or instrumentalism in their more mature writings. Reichenbach and Feigl, for example, were outspoken realists, and Carnap regarded physicalism as a tenable philosophical framework. Reichenbach never associated himself with positivism; indeed, he regarded his 1938 book, *Experience and Prediction* (Chicago: University of Chicago Press), as a refutation of logical positivism. I could go on and on . . .

3. "Symposium: The Philosophy of Carl G. Hempel," *Journal of Philosophy* LXXX, 10 (Oct., 1983): 555–72.

4. I had offered a similar suggestion in "Bayes's Theorem and the History of Science," in *Minnesota Studies in the Philosophy of Science*, vol. 5 *Historical and Philosophical Perspectives of Science*, ed. Roger H. Stuewer (Minneapolis: University of Minnesota Press, 1970), 68–86.

5. Thomas S. Kuhn, *The Essential Tension* (Chicago: University of Chicago Press, 1977), 320–39. The response is given in greater detail in this article than it is in the Postscript.

6. "Objectivity, Value Judgment, and Theory Choice," 321–22.

7. *Ibid.*, p. 322.

8. Throughout this paper I shall use the terms "hypothesis" and "theory" more or less interchangeably. Kuhn tends to prefer "theory," while I tend to prefer "hypothesis," but nothing of importance hinges on this usage here.

9. As Adolf Grünbaum pointed out to me, if we assume that in a given day the actual frequency of defective can openers produced by the two machines matches precisely the respective probabilities, we can calculate the result as follows. The new machine produces 50 defective can openers and the old machine produces 30, so that 50 out of a total of 80 are produced by the new machine. However, it would be *incorrect* to assume that the frequencies match the probabilities each day; in fact, the probability of an exact match is quite small.

10. "Objectivity, Value Judgment, and Theory Choice," 328.

11. I remarked above that three probabilities are required to calculate the posterior probability—a prior probability and two likelihoods. Obviously, in view of (2), the theorem on total probability, if we have a prior probability, one of the likelihoods, and the expectedness, we can compute the other likelihood; likewise, if we have one prior probability and both likelihoods, we can compute the expectedness.

12. *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1967), chap. 7.

13. A set of degrees of conviction is coherent provided that its members do not violate any of the conditions embodied in the mathematical calculus of probability.

14. From "Autobiographical Notes," in Paul A. Schilpp, ed., *Albert Einstein: Philosopher-Scientist* (Evanston, Ill: Library of Living Philosophers, 1949), 21–22.

15. This example is discussed in Ronald N. Giere, *Understanding Scientific Reasoning*, 2d ed. (New York: Holt, Rinehart, and Winston, 1984), 274–76.

16. I reject the so-called propensity interpretation of probability because, as Paul Humphreys

pointed out, the probability calculus accommodates inverse probabilities of the type that occur in Bayes's theorem, but the corresponding inverse propensities do not exist. In the example of the can opener factory, each machine has a certain propensity to produce defective can openers, but it does not make sense to speak of the propensity of a given defective can opener to have been produced by the new machine.

17. A so-called Dutch book is a combination of bets such that, no matter what the outcome of the event upon which the wagers are made, the subject is bound to suffer a net loss.

18. "Truth and Probability," in Frank Plumpton Ramsey, *The Foundations of Mathematics* ed. R. B. Braithwaite (New York: Humanities Press, 1950).

19. Gerald Holton and Stephen G. Brush, *Introduction to Concepts and Theories in Physical Science*, 2d ed. (Reading, Mass.: Addison-Wesley, 1973), 416, italics in original.

20. Abner Shimony, "Scientific Inference," in Robert G. Colodny, ed., *The Nature and Function of Scientific Theories* (Pittsburgh: University of Pittsburgh Press, 1970), 79–172.

21. Martin Gardner, *Fads and Fallacies in the Name of Science* (New York: Dover Publications, 1957), 7–15.

22. Doubleday & Company, 1950.

23. Haim Harari, "The Structure of Quarks and Leptons," *Scientific American* 248, (April 1983): 56–68.

24. *Scientific American*, 253 (Sept. 1985): 116. Adolf Grünbaum, *The Foundations of Psychoanalysis* (Berkeley/Los Angeles/London: University of California Press, 1984), 202–4, criticizes Motley's account of Freud's theory; he considers Motley's version a distortion, and points out that Freud's motivational explanations were explicitly confined to a very circumscribed set of slips. He defends Freud against Motley's criticism on the grounds that Freud's actual account has greater complexity than Motley gives it credit for.

25. "A Bayesian Approach to the Paradoxes of Confirmation," in Jaakko Hintikka and Patrick Suppes, eds., *Aspects of Inductive Logic* (Amsterdam: North Holland, 1966), 202–3.

26. I have discussed the relations between personal probabilities and objective probabilities in "Dynamic Rationality: Propensity, Probability, and Credence," in James H. Fetzer, ed., *Probability and Causality* (Dordrecht: Reidel 1988), 3–40.

27. As Duhem has made abundantly clear, in such cases we may be led to reexamine our background knowledge B, which normally involves auxiliary hypotheses, to see whether it remains acceptable in the light of the negative outcome E. Consequently, refutation of T is not usually as automatic as it appears in the simplified account just given. Nevertheless, the probability relation just stated is correct.

28. Exchangeability is the personalist's surrogate for randomness; it means that the subject would draw the same conclusion regardless of the order in which the members of an observed sample occurred.

29. Note that, in order to get the posterior probability—the probability that the observed results were produced by a biased device—the prior probabilities have to be taken into account.

30. Indeed, stellar parallax was not detected until the nineteenth century.

31. This question was, in fact, raised by Adolf Grünbaum in a private communication.

32. See, for example, Gerald Holton and Stephen G. Brush, *Introduction to Concepts and Theories in Physical Science*, 2d ed. (Reading, Mass.: Addison-Wesley, 1973), 392–93.

33. An excellent account of Brewster's position can be found in John Worrall, "Scientific Revolutions and Scientific Rationality: The Case of the Elderly Holdout," this volume, pp. 319–36.

34. Quoted by Worrall, p. 321.

35. Quoted by Worrall, p. 322.

36. Quoted by Worrall, p. 323.

37. If more than two theories are serious candidates, the pairwise comparison can be repeated as many times as necessary.

38. Adolf Grünbaum, "Can a Theory Answer More Questions Than One of Its Rivals," *British Journal for the Philosophy of Science* 27 (1976): 1-23.

39. Steven Weinberg, *Gravitation and Cosmology* (New York: John Wiley & Sons, 1972), 198. Note that this correction is smaller by an order of magnitude than the correction of 43 seconds of arc per century for Mercury.

40. See Joel Smith, *The Status of Inconsistent Statements in Scientific Inquiry* (doctoral dissertation, University of Pittsburgh, 1987).

41. Unfortunately, recent evidence regarding the mass of Pluto strongly suggests that Pluto is not sufficiently massive to explain the perturbations of Neptune. A different plausible scenario is needed, but I do not know of any serious candidates that have been offered.

42. See "Dynamic Rationality" (note 26 above), 5-12, for a more detailed discussion of various grades of rationality. The term "static" was chosen to indicate the lack of any principled method for changing personal probabilities in the face of inconsistency or incoherence.

43. *Ibid.*, esp. pp. 11-12

44. Thomas S. Kuhn, *The Copernican Revolution* (Cambridge: Harvard University Press, 1957), 3-4.