Valuation and Acceptance of Scientific Hypotheses
Author(s): Richard C. Jeffrey
Source: *Philosophy of Science*, Vol. 23, No. 3 (Jul., 1956), pp. 237-246
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: https://www.jstor.org/stable/185418

DISCUSSION

Valuation and Acceptance of Scientific Hypotheses

RICHARD C. JEFFREY[1]

1. Introduction. Churchman (4), Braithwaite (1), and Rudner (7) have recently argued from premises acceptable to many empiricists to the conclusion that ethical judgments are essentially involved in decisions as to which hypotheses should be included in the body of scientifically accepted propositions.[2] Rudner summarizes the argument:

> Now I take it that no analysis of what constitutes the method of science would be satisfactory unless it comprised some assertion to the effect that the scientist as scientist accepts or rejects hypotheses.
>
> But if this is so then clearly the scientist as scientist does make value judgments. For, since no scientific hypothesis is ever completely verified, in accepting a hypothesis the scientist must make the decision that the evidence is *sufficiently* strong or that the probability is *sufficiently* high to warrant the acceptance of the hypothesis. Obviously our decision regarding the evidence and respecting how strong is "strong enough", is going to be a function of the *importance*, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis (7, p. 2).

The form of this reasoning is hypothetical: *if* it is the job of the scientist to accept and reject hypotheses, *then* he must make value judgments. Now I shall argue (in effect) that if the scientist makes value judgements, then he neither accepts nor rejects hypotheses. These two statements together form a *reductio ad absurdum* of the widely held view which our authors presuppose, that science consists of a body of hypotheses which, pending further evidence, have been *accepted* as highly enough confirmed for practical purposes ("practical" in Aristotle's sense).

In place of that picture of science I shall suggest that the activity proper to the scientist is the assignment of probabilities (with respect to currently available evidence) to the hypotheses which, on the usual view, he simply accepts or rejects. This is not presented as a fully satisfactory position, but rather as the standpoint to which we are led when we set the Churchman-Braithwaite-Rudner arguments free from the presupposition that it is the job of the scientist as such to accept and reject hypotheses.

In the following pages we shall frequently have to speak of probabilities in connection with rational choice, and this opens us to the danger of greatly complicating our task through involvement in the dispute between conflicting theories of probability.[3] To avoid this I shall make use of the notion that these

[1] The author wishes to express his thanks to Prof. C. G. Hempel, at whose suggestion this paper was written, and to Dr. Abner Shimony, for their helpful criticism.

[2] Rudner (7) has the most explicit and unqualified statement of the point of view in question. These views stem largely from recent developments in statistics, especially from the work of Abraham Wald; see (9) and references there to earlier writings.

[3] For accounts of this dispute, see (2), ch. 2, (3), §1, and (5).

theories are conflicting explications[4] of the concept, *reasonable degree of belief*. If this is so we shall usually be able to avoid the controversy by using the subjectivistic language ("rational degree of belief" or of "confidence") which is appropriate to the *explicandum*. This device is justified if the reader finds the relevant statements acceptable after he has freely translated them into the terminology of whichever *explicatum* he prefers.

2. Betting and Choosing. It is commonly held that although we have no certain knowledge we must often act as if probable hypotheses were known to be true. For example it might be said that when we decide to inoculate a child against polio we are accepting as certain the hypothesis that the vaccine is free from virulent polio virus. Proponents of this view speak of "accepting a hypothesis" as a sort of inductive jump from high probability to certainty.

On this account, betting is an exceptional situation. Let H be the hypothesis that the ice on Lake Carnegie is thick enough to skate on. Now if one is willing to give odds of 4:1 to anyone who will bet that not-H, and if these are the longest odds one is willing to give, one has pretty well expressed by a sort of action a degree of belief in H four times as great as one's belief in not-H. Here one is risking a good—money—which admits of degrees, so that in the bet one is able to adjust the degree of risk to the degree of belief in H. But in actually attempting to skate, degree of commitment cannot be nicely tailored to suit degree of belief: one cannot arrange to fall in only part way in case the ice breaks.

Part of the discrepancy between betting and other choosing can be disposed of immediately. Thus far we have stressed the case where the bettor himself proposes the stakes, and indeed there is nothing comparable to this in most other choosing. When you "bet" by trusting the ice with your own weight the stakes are, say, a dunking if you lose and an afternoon's skating if you win. These stakes are fixed by nature, not by the skater. But the same arrangement is common in actual betting, e.g. at a race track, where the bettor's problem is not to propose the stakes but rather to decide whether odds offered by someone else are fair or advantageous to him. In such cases the bettor cannot pick the exact degree of his commitment any more than the skater can.

In betting as in other choosing the rational agent acts so as to maximize his expectation of value. In betting, the values at stake seem especially easy to measure: the value or utility of winning is measured by the amount of money won, and the value of losing by the amount lost. But it is well known that this identification is vague and approximate. In the bet about the ice, for example, the identification of utility with money would lead one to accept as advantageous odds of 1:1 when the ratio of degrees of belief is 4:1. But clearly it makes a difference whether the 1:1 in question is \$1:\$1 or \$1000:\$1000. The former would be a good bet for a man of moderate means, but the latter would not.

The usual way out of this difficulty is to specify that the stakes be small compared with the bettor's fortune, but not so small as to bore him. The importance

[4] The view that a theory of probability is an explication of a vague concept in common use (the *explicandum*) by a precise concept (the *explicatum*) is due to Carnap; cf. (2), ch. 1.

of finding a way out is that the ratio of stakes which a man finds acceptable is a convenient measure of degree of belief. But we have seen that it is not always a reliable measure. Therefore it seems appropriate to interpret the relationship between odds and utilities in the same way we interpret the relationship between the height of a column of mercury and temperature; the one is a reliable sign of the other within a certain range, but is unreliable outside that range, where we accordingly seek other signs (e.g. alcohol thermometers below and gas thermometers above the range of reliability of mercury).

3. Rational Decision: The Bayes Criterion. The decision whether or not to accept a bet separates naturally into four stages; we illustrate in the case of the bet about the hypothesis (H) that the ice on Lake Carnegie is thick enough for skating. First we draw up a *table of stakes* indicating what will be won or lost in each of the four situations which can arise depending on whether the hypothesis is true or not, and whether the bet is accepted or not.

Choice	Actual state of the ice: H	Actual state of the ice: not-H
A: accept the bet.	Win $1.	Lose $1.
not-A: don't accept the bet.	Neither win nor lose.	

Table of Stakes

In this case we may suppose the utilities of the stakes to be proportional to the stakes themselves, so that the *table of utilities* looks like this:

	H	not-H
A	1	-1
not-A	0	0

Here we are concerned not with the numbers themselves, but rather with their ratios; the same information about the utilities is contained in any table which is the same as the one above except that all entries are multiplied by some positive number, e.g. $\begin{bmatrix} 5 & -5 \\ 0 & 0 \end{bmatrix}$

If somehow we know that the bettor's degrees of belief in H and not-H stand in the ratio 4:1 we can construct a *table of expectations*:

	H	not-H
A	4	-1
not-A	0	0

where the expectation in each of the four situations is the product of belief in and utility of that situation.

By adding the two numbers in the top row of this table we get the *total expectation* from choice A (accepting the bet): 3. The sum of the numbers in the bottom row is the total expectation from not-A. The Bayes criterion defines the rational choice to be the one with the greater expectation. Here, then, the rational decision would be to accept the bet.

The decision about actually trying the ice is exactly parallel. Here the table of stakes is

	H	not-H
A: try to skate.	Skate	Get wet
not-A: don't try.	Neither skate nor get wet.	

If skating and getting wet are equal and opposite goods, the utility table and the rest of the calculation is identical with that for the betting decision, and the recommendation is: try the ice.

The Bayes criterion has generally been accepted as a satisfactory explication of "rational choice" *relative to a set of numerical utilities and degrees of belief*. If the criterion is accepted then the rationality of a decision which conforms to it can be attacked only on grounds that the degrees of belief and utilities involved are themselves unreasonable. The most influential school of thought in statistics today holds that in many cases there are no reasonable grounds for assigning probabilities to sets of hypotheses. This does not mean that in such cases the reasonable degree of belief in each hypothesis is zero, or that it is the same for all hypotheses, but simply that no numerical assignment whatever can be justified. Accordingly, statisticians have developed alternatives to Bayes' criterion, one of which (the *minimax* criterion) we shall consider in section 5.

The question of how and whether it is possible to justify the assignment of numerical utilities to situations is even more difficult, and we do not propose to consider it here. But it should be noted that the use made of utilities by the Bayes criterion is not very exacting. As observed earlier, we are concerned not with the utilities themselves, but with their ratios. Further, it is easy to show that often not even that much is required. For example, in applying the Bayes criterion to a choice between two actions it is sufficient to know ratios of certain *differences* between the utilities.[5]

4. Choice between Hypotheses: Bayes' Method. Meno, in the dialogue bearing his name, makes a strong objection to the Socratic concept of inquiry:

> And how will you enquire, Socrates, into that which you do not know? What will you put forth as the subject of enquiry? And if you find what you want, how will you ever know that this is the thing which you did not know? (6, Steph. 80).

[5] Let $H_1, \ldots, H_n$ be mutually exclusive, collectively exhaustive hypotheses for which $\beta_1, \ldots, \beta_n$ are the corresponding degrees of belief. A decision is to be made between acts A_1 and A_2. Let Δ_i be the difference: (utility of choosing A_1 if H_i is true) $-$ (utility of choosing A_2 if H_i is true). Then in this case the Bayes criterion reduces to: choose A_1 if $\Delta_1\beta_1 + \ldots + \Delta_n\beta_n > 0$, and choose A_2 if the inequality goes the other way.

In reply, Socrates undertakes the famous demonstration of how geometrical ideas can be "recollected" by an ignorant boy. But then he goes on, and apparently weakens the force of the demonstration by admitting

> . . . Some things I have said of which I am not altogether confident. But that we shall be better and braver and less helpless if we think that we ought to enquire, than we should have been if we indulged in the idle fancy that there was no knowledge and no use in seeking to know what we do not know;—that is a theme upon which I am ready to fight, in word and deed, to the utmost of my power (6, Steph. 86).

This is not mere wishful thinking, but rather part of a rational argument, in the Bayes sense of "rational". Using our previous notation, the hypothesis under consideration is (H): "Knowledge is obtainable through inquiry," and the choice is between A, the decision to inquire, and not-A. Meno had made it plausible that H is very improbable; Socrates' first reply (the demonstration of recollection) was an attempt to undermine Meno's argument directly, by showing that in fact H is more probable than Meno would have us believe. Socrates' second reply, quoted above, concedes that Meno may be right, but goes on to say that even if H *is* improbable, the utility of knowledge is so great that even when it is multiplied by a small probability, *the total expectation from inquiry* (A) *exceeds that from not-*A.

	H	not-H
A	Possibility of obtaining knowledge.	Waste of effort.
not-A	No knowledge obtained and no effort wasted in seeking it.	

Table of Stakes

For amusement, we might assign numerical utilities to the table of stakes: $\begin{bmatrix}1000 & -1\\ 0 & 0\end{bmatrix}$. A little calculation shows that with these utilities it is rational to inquire even if the probability of H is as low as .001.

This pattern of argument is fairly common as a justification for faith—in God (Pascal's wager), in inquiry (Socrates), in the unity of the laws of nature (Einstein). But it should be noted that in all three cases what is meant by "faith" is not verbalized intellectual acceptance of the truth of a thesis, but rather commitment to a line of action which would be useless or even damaging if the thesis in question were false. Typically, in these cases, the thesis itself is extremely vague; but it is meaningful to the person who accepts it in the sense that it partly determines his activity.

To take a more precise thesis as an example, consider the problem of quality control in the manufacture of polio vaccine. A sample of the vaccine in a certain lot is tested and found to be free from active polio virus. Let us suppose that this imparts a definite probability to the hypothesis that the entire lot is good. Is this probability high enough for us rationally to accept the hypothesis?

Contrast this with a similar problem about roller skate ball bearings. Imagine that here, too, a sample has been taken and that all the bearings tested have proved satisfactory; and suppose that this evidence imparts to the hypothesis that all the bearings in the lot are good the same probability that we encountered before in the case of the vaccine. As Rudner points out, we might accept the ball bearings and yet reject the vaccine because although the probabilities are the same in the two cases, the utilities are different. If the probability were just enough to lead us to accept the bearings, we should reject the vaccine because of the graver consequences of being wrong about it.

But what determines these consequences? There is nothing in the hypothesis, "This vaccine is free from active polio virus", to tell us what the vaccine is *for*, or what would happen if the statement were accepted when false. One naturally assumes that the vaccine is intended for inoculating children, but for all we know from the hypothesis it might be intended for inoculating pet monkeys. One's confidence in the hypothesis might well be high enough to warrant inoculation of monkeys, but not of children.

The trouble is that implicitly we have been discussing a utility table with these headings

	H	not-*H*
Accept *H*		
Reject *H*		

but there is no way to decide what numbers should be written in the blank spaces unless we know what actions depend on the acceptance or rejection of *H*. Bruno DeFinetti sums up the case:

> I do not deem the usual expression "to accept hypothesis H_r" to be proper. The "decision" does not really consist of this "acceptance" but in *the choice of a definite action* A_r. The connection between the action A_r and the hypothesis H_r may be very strong, say "the action A_r is that which we would choose if we knew that H_r was the true hypothesis." Nevertheless, this connection cannot turn into an identification (5, p. 219).

This fact is obscured when we consider very specialized hypotheses of the sort encountered in industrial quality control, where it is clear from the contexts, although not expressly stated in the hypotheses, what actions are in view. But the vaccine example shows that even in these cases it may be necessary to make the distinction that DeFinetti urges. In the case of lawlike scientific hypotheses the distinction seems to be invariably necessary; there it is certainly meaningless to speak of *the* cost of mistaken acceptance or rejection, for by its nature a putative scientific law will be relevant in a great diversity of choice situations among which the cost of a mistake will vary greatly.

In arguing for his position Rudner concedes, "The examples I have chosen are from scientific inferences in industrial quality control. But the point is clearly general in application" (7, p. 2). Rudner seems to give his reason for this last statement later on:

> I believe, of course, that an adequate rational reconstruction of the procedures of science would show that every scientific inference is properly construable as a statistical inference (i.e., as an inference from a set of characteristics of a sample of a population to a set of characteristics of the whole population (7, p. 3).

But even if analysis should show that lawlike hypotheses are like the examples from industrial quality control in being inferences from characteristics of a sample to characteristics of an entire population, they are different in the respect which is of importance here, namely their generality of application. Braithwaite and Churchman are more cautious here; they confine their remarks to statistical inferences of the ordinary sort. But we have seen that even in statistics the feasibility of blurring the distinction between accepting a hypothesis and acting upon it depends on features of the statement of the problem which are not present in every inference.

5. Choice Between Hypotheses: Minimax Method. In applying the Bayes criterion to quality control we assumed that on the basis of the relative frequency of some property in a sample of a population, definite probabilities can be assigned to the various conflicting hypotheses about the relative frequency of that property in the whole population. In general such "inverse inference" presupposes a knowledge of the *prior probabilities* (or of an *a priori probability distribution*) for the hypotheses in question. On the other hand, "direct inference" from relative frequencies in an entire population to relative frequencies in samples involves no such difficulty. Wald writes,

> In many statistical problems the existence of an a priori distribution cannot be postulated, and, in those cases where the existence of an a priori distribution can be assumed, it is usually unknown to the experimenter and therefore the Bayes solution cannot be determined (9, p. 16).

For these cases Wald proposes a criterion which makes no use of inverse inference.

For simplicity we consider the case where somehow it is known that one or the other of two hypotheses, H_1 or H_2, must be true. The hypotheses assign different relative frequencies of a property P to a population. A sample consisting of only one member is drawn from the population and will be inspected for this property. Wald's problem is to choose, in advance of the inspection, an *inductive rule* which tells him which hypothesis to accept under every possible assumption as to the outcome of the inspection. In this case the choice is between four rules, since the relative frequency of P in the sample can only be 0 or 100%.

Rule 1. Accept H_1 in either case.

Rule 2. Accept H_1 in case the relative frequency of P in the sample is 0, H_2 if it is 100%.

Rule 3. Accept H_2 in case the relative frequency of P in the sample is 0, H_1 if it is 100%.

Rule 4. Accept H_2 in either case.

By direct inference one can find the conditional probabilities that each of the inductive rules will lead to the right or the wrong hypothesis on the assumption that H_1 is true, and separately on the assumption that H_2 is true. From these eight probabilities together with a knowledge of the losses (negative utilities) that would result from accepting one of the H's when in fact the other is true, one can calculate a table of risks, e.g.:

Inductive rule	Risk in using this rule in case H_1 is true	Risk in using this rule in case H_2 is true
1	7	0
2	½	5
3	4	2
4	10	18

Wald's *minimax criterion* is: minimize the maximum risk. Here this directs us to choose the rule—3—for which the larger of the two risks is least.

The minimax criterion is the counsel of extreme conservatism or pessimism. Wald proves this in two ways. (i) He shows that "a minimax solution is, under some weak restrictions, a Bayes solution relative to a least favorable a priori distribution" (9, p. 18). (ii) He shows that the situation in which an experimenter uses the minimax criterion to make a decision is formally identical with the situation in which the experimenter is playing a competitive "game" with a personalized Nature in the sense that the experimenter's losses are Nature's gains. Nature plays her hand by selecting a set of prior probabilities for the hypotheses between which the experimenter must choose; being intelligent and malevolent, Nature chooses a set of probabilities which are as unfavorable as possible when viewed in the light of the negative utilities which the experimenter attaches to the acceptance of false hypotheses.

Since different experimenters make different value judgments, it would seem that in applying the minimax criterion each experimenter implicitly assumes that this is the worst of all possible worlds *for him*. We might look at the matter in this way: the minimax criterion is at the pessimistic end of a continuum of criteria. At the other end of this continuum is the "minimin" criterion, which advises each experimenter to minimize his minimum risk. Here each experimenter acts as if this were the *best* of all possible worlds *for him*. The rules at both extremes of the continuum share the same defect: they presuppose a great sensitivity on the part of Nature to human likes and dislikes and are therefore at odds with a basic attitude which we all share, in our lucid moments.

Wald was aware of the sort of objection we have been making:

> The analogy between the decision problem and a two-person game seems to be complete, except for one point. Whereas the experimenter wishes to minimize the risk . . . , we can hardly say that Nature wishes to maximize [the risk]. Nevertheless, since Nature's choice is unknown to the experimenter, it is perhaps not unreasonable for the experimenter to behave as if Nature wanted to maximize the risk (9, p. 27).

This suggests that we have overstated our case. As a general inductive rule, the minimax criterion represents an unprofitable extreme of caution; nevertheless we feel that there are conditions under which a man would do well to act with a maximum of caution, even though it would be unwise to follow that policy for all decisions. What we lack is an account of the conditions under which it is appropriate to use the minimax criterion.[6]

Apart from this, our previous objections to the notion of "accepting" a hypothesis apply to the minimax as well as to the Bayes criterion. Wald's procedure leads us to accept an inductive rule which, once the experiment has been made, determines one of the competing hypotheses as the "best". But this means, best for making the specific choice in question, e.g. whether to inject a child with polio vaccine from a certain lot. Among the *same* hypotheses, a different one might be best with respect to a different choice, e.g. inoculating a pet monkey. Hence both the Bayes and minimax criteria permit choice between hypotheses only with respect to a set of utilities which in turn are relative to the intended applications of the hypotheses.

6. Conclusion. On the Churchman-Braithwaite-Rudner view it is the task of the scientist as such to accept and reject hypotheses in such a way as to maximize the expectation of good for, say, a community for which he is acting. On the other hand, our conclusion is that if the scientist is to maximize good he should refrain from accepting or rejecting hypotheses, since he cannot possibly do so in such a way as to optimize every decision which may be made on the basis of those hypotheses. We note that this difficulty cannot be avoided by making acceptance relative to the most stringent possible set of utilities (even if there were some way of determining what that is) because then the choice would be wrong for all less stringent sets. One cannot, by accepting or rejecting the hypothesis about the polio vaccine, do justice both to the problem of the physician who is trying to decide whether to inoculate a child, and the veterinarian who has a similar problem about a monkey. To accept or reject that hypotheses once for all is to introduce an unnecessary conflict between the interests of the physician and the veterinarian. The conflict can be resolved if the scientist either contents himself with providing them both with a single probability for the hypothesis (whereupon each makes his own decision based on the utilities peculiar to his problem), or if the scientist takes on the job of making a separate decision as to the acceptability of the hypothesis in each case. In any event, we conclude that it is not the business of the scientist as such, least of all of the scientist who works with law-like hypotheses, to accept or reject hypotheses.

We seem to have been driven to the conclusion that the scientist's proper role is to provide the rational agents in the society which he represents with probabilities for the hypotheses which on the other account he simply accepts or rejects. There are great difficulties with this view. (i) It presupposes a satisfactory theory of probability in the sense of *degree of confirmation* for hypotheses

[6] Savage ((8), ch. 13) discusses a number of other objections to the minimax criterion.

on given evidence. (ii) Even if such a theory were available there would be great practical difficulties in using it in the way we have indicated. (iii) This account bears no resemblance to our ordinary conception of science. Books on electrodynamics, for example, simply list Maxwell's equations as laws; they do not add a degree of confirmation. These are only some of the difficulties with the probabilistic view of science.

To these, Rudner adds a very basic objection.

> . . . the determination that the degree of confirmation is say, *p*, . . . which is on this view being held to be the indispensable task of the scientist *qua* scientist, is clearly nothing more than *the acceptance by the scientist of the hypothesis that the degree of confidence is p* . . . (7, p. 4).

But of course we must reply that it is no more the business of the scientist to "accept" hypotheses about degrees of confidence than it is to accept hypotheses of any other sort, and for the same reasons.[7] Rudner's objection must be included as one of the weightiest under heading (i) above as a difficulty of the probabilistic view of science. These difficulties may be fatal for that theory; but they cannot save the view that the scientist, *qua* scientist, accepts hypotheses.

Princeton University

REFERENCES

1. R. B. Braithwaite, *Scientific Explanation* (Cambridge University Press, Cambridge, 1953).
2. Rudolf Carnap, *Logical Foundations of Probability* (University of Chicago Press, Chicago, 1950).
3. Rudolf Carnap, *The Continuum of Inductive Methods* (University of Chicago Press, Chicago, 1952).
4. C. West Churchman, *Theory of Experimental Inference* (The Macmillan Co., New York, 1948).
5. Bruno DeFinetti, pp. 217–225 of *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. Jerzy Neyman (University of California Press, Berkeley, 1951).
6. Plato, "Meno", *The Dialogues of Plato*, tr. Benjamin Jowett (3rd ed., Random House, Inc., New York), I, 349–380.
7. Richard Rudner, "The Scientist *Qua* Scientist Makes Value Judgements", *Philosophy of Science*, vol. 20 (1953), pp. 1–6.
8. Leonard J. Savage, *The Foundations of Statistics* (John Wiley and Sons, Inc., New York, 1954).
9. Abraham Wald, *Statistical Decision Functions* (John Wiley and Sons, Inc., New York, 1950).

[7] In Carnap's confirmation theory there at first seems to be no difficulty since it is a logical rather than a factual question, what the degree of confirmation of a given hypothesis is, with respect to certain evidence. But the difficulty may appear at a deeper level in choosing a particular c-function; this Carnap describes as a practical decision. See (3), §18.