

Bayesian Methods and the Initial Probabilities of Theories

I. Introduction

In the past, objections have been raised to the notion of a probabilistic confirmation theory on the grounds that it required assignment of initial probabilities to expressions of hypotheses and evidence and that these probability values are not directly verifiable as are probabilities in the classic theories of equipossible alternatives or in the statistical frequency theory. Such objections have often come from philosophers and statisticians of positivist bent¹ and are closely parallel to objections within science itself to introduction of theoretical concepts that are not directly verifiable. The deductivist analysis of science has, however, released us from the tyranny of the "directly verifiable," in showing how unobservables can be given a function in theories, and it prompts the suggestion with regard to a confirmation theory that initial probabilities might be construed as "theoretical terms" to be judged by how well they enable the theory to systematize and explicate inductive arguments that are

AUTHOR'S NOTE: This paper is an expanded version of my "Theories and the Transitivity of Confirmation," *Philosophy of Science*, 37 (1970), 50-63, which was in turn an expanded version of a paper delivered at the Colloquium on Confirmation held in the Minnesota Center for Philosophy of Science in July 1968. I am indebted to Professor Allan Birnbaum for helpful comments on a first draft of Section II. Since preparing the paper for publication in this volume, I have incorporated much of sections II-V into my book *The Structure of Scientific Inference* (London: Macmillan, 1974). These passages are reproduced here with the publisher's permission.

¹ For example, the frequentist statistician R. von Mises makes the quasi-operational demand, "A quantitative probability concept must be defined in terms of potentially unlimited sequences of observations and experiments." *Probability, Statistics, and Truth*, 2nd Eng. ed. (London: Allen & Unwin, 1957), p. viii, and the positivist philosopher H. Reichenbach claims for his account of probability that "there is no doubt that this finitization [of the frequency limit for finite sequences] will satisfy all requirements of the verifiability theory of meaning in its most rigorous form." *The Theory of Probability*, Eng. ed. (Berkeley and Los Angeles: University of California Press, 1949), p. 348.

actually used in scientific contexts. Just as the justification for assigning a numerical mass to the unobservable electron is found in the verifiable consequences of this assignment, so the justification for assigning a probability value to a hypothesis prior to evidence is to be found in the agreement or disagreement of the consequences of this assignment with the posterior probabilities of hypotheses which are implicit in our attitudes of acceptance or rejection of these hypotheses on the basis of evidence.

Not only is a more liberal climate in philosophy of science more congenial to the introduction of initial probabilities, but also developments within statistics itself have tended in the same direction. The subjective view of probability as "degree of belief" is interpreted in more objective fashion in the personalist analysis of "degree of rational belief," and the introduction of Bayesian methods has clarified use of inverse arguments from evidence and initial probabilities to the posterior probabilities of hypotheses.

Criticism of probabilistic methods in confirmation theory now comes on more detailed grounds, namely that our intuitive requirements for the confirmation of hypotheses by evidence do not satisfy the probability axioms.² Much confusion has arisen in this debate from failure to recognize that there are many different explicata closely connected in ordinary discourse with the concept of confirmation, but not all behaving in the same way. Among others, the notions of acceptability of hypotheses, desirable risk in action, utilities, power, content, expectation of truth or success, and support by evidence have all entered the discussion. All are initially as unclear as the notion of confirmation itself. But in this area it is impossible to solve all problems until we commit ourselves to a particular solution of some of them. I shall therefore concentrate here on the notion of confirmation or support of hypotheses by evidence, and begin by adopting the personalist interpretation of confirmation as

² Especially K. R. Popper, *Logic of Scientific Discovery* (London: Hutchinson, 1959), ch. 10 and appendix *vii, and *Conjectures and Refutations* (London: Routledge, 1963), chs. 1, 11, and pp. 388-91; and L. J. Cohen, *The Implications of Induction* (London: Methuen, 1970). For critical surveys of Popper's and Carnap's positions, see I. Lakatos, "Changes in the Problem of Inductive Logic," in I. Lakatos, ed., *The Problem of Inductive Logic* (Amsterdam: North Holland, 1968), p. 315; and Alex C. Michalos, *The Popper-Carnap Controversy* (The Hague: Nijhoff, 1971). For arguments on behalf of Bayesian methods, see W. C. Salmon, *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1966), ch. 7.

degree of rational belief in hypotheses or predictions relative to evidence.

The first half of the paper (sections II–IV) rehearses the personalist's arguments from rational decision procedures for taking degree of rational belief to be measured by a probability function, and the objections to them. It is by now generally accepted that these arguments provide no grounds for particular assignments of initial probabilities, nor for the use of Bayes's theorem to obtain posterior distributions of belief from initial distributions together with observed evidence. I next investigate some suggested transformations of belief which are not Bayesian and also find these wanting. Finally in this critical analysis of Bayesian methods, the claims of personalists for "convergence of belief" with increasing evidence are investigated and found largely inapplicable to the conditions of confirmation in science.

The first half of the paper shows the inadequacy of relatively a priori approaches to Bayesian confirmation theory. The methods criticized are those in which the probability axioms are claimed to be derivable from rationality alone, and no constraints are put on the initial probabilities (except that some are required to be non-zero), since such constraints would not be derivable a priori. The second half of the paper (sections V and VI) attempts to develop a more modest Bayesian confirmation theory for science, in which intuitions concerning purely local inductive methods are used to suggest stronger constraints on the initial distribution of belief and on transformations from initial to posterior distributions. Some intuitive criteria of adequacy for a confirmation theory are found to be explicated immediately by a Bayesian theory; others, involving universal generalizations and theories, are more problematic, but are found to raise interesting general issues regarding the function of laws and theories.

II. Games against Nature

1. The Probability Axioms

The starting point of a personalist theory of confirmation is the assimilation of the scientist's problem in confronting hypotheses with evidence to *betting behavior*, and more generally to *decision-making under uncertainty*, or in a more perspicuous phrase, "games against nature." In order to bring to bear upon a man's subjective beliefs some objective norms to which all rational men should assent, the personalist

proposes (following the original suggestion of F. P. Ramsey)³ to interpret "degree of belief" in a prediction behaviorally in terms of the maximum betting odds a man would be prepared to accept on the occurrence of the predicted event, provided no considerations other than his belief influence his attitude to the bet (for example, being honest and having no money, and therefore being unwilling to take any risk of loss however strong his belief). Then it is held that a rational man will not allow his betting system to be such that an opponent who knows no more than he does can make a book against him which will result in his loss whatever the outcome of the bet (a *Dutch book*). A betting function is called *coherent* if it assigns real numbers in the closed interval (0,1) to the possible outcome of the bet in such a way that if betting odds are laid on the occurrence of the outcomes in the ratio of the numbers, a Dutch book cannot be made by the opponent. It can then be shown that a coherent betting function is a probability measure over the possible outcomes and conversely that any probability measure corresponds to some coherent betting function. Thus the personalist's case for a probabilistic confirmation theory rests on two proposed equivalences:

- (A) Degree of rational belief $\equiv$ coherent betting behavior.
- (B) Coherent betting function $\equiv$ probability measure.

We shall need first a convenient formulation of the probability axioms. We are concerned with expressions of hypotheses, evidence, and predictions satisfying the axioms of propositional logic. These propositions will be denoted by lowercase letters such as *h*, *g*, *f*, *e*, sometimes with suffixes, to indicate that some of the terms in a probability expression will generally be interpreted as hypotheses or laws (as *h,g,f*) and some as evidence or observable predictions (as *e*). All the axioms and theorems to be discussed will, however, be valid for any substitution of *h*-type letters for *e*-type letters, and conversely, unless something is specifically indicated to the contrary. Probability will be understood as a measure over a pair of propositions; thus, $p(h/e)$ is to be read as "the probability of *h* conditionally upon *e*," or "the probability of *h* given *e*." The *initial probability* of *h* will be written $p(h)$, but this is to be understood consistently with what has just been said as the probability of *h* given either the tautology or previous background evidence which does not enter into the current calculations.

³*Foundations of Mathematics*, ed. R. B. Braithwaite (London: Routledge, 1931), ch. 7.

The probability axioms are as follows:

- (P.1) $0 \leq p(h/e) \leq 1$;
 (P.2) *Equivalence*: If h and e are logically equivalent to h' and e' respectively, then $p(h/e) = p(h'/e')$;
 (P.3) *Entailment*: If e logically implies h , then $p(h/e) = 1$;
 (P.4) *Addition*: If e logically implies $\sim(h \& h')$, then $p(h \vee h'/e) = p(h/e) + p(h'/e)$;
 (P.5) *Multiplication*: If e, e' are consistent and not self-contradictory, $p(h \& e'/e) = p(e'/e) p(h/e' \& e)$.

Logical implication or entailment will be written $\rightarrow$.

If the requirement of coherence is strengthened by requiring that the odds are not such that a bettor must certainly either come out even or lose for every possible outcome (*strict coherence*), then it can be shown that the necessary and sufficient conditions are obtained by replacing axiom (P.3) by (P.3'):

(P.3') If and only if e logically implies h , then $p(h/e) = 1$.

With this replacement the axioms define a regular probability measure.

Some immediate corollaries of the axioms will be useful.

(C.1) From the addition axiom it follows immediately that if on evidence e we have a set $S = \{h_1, h_2, \dots, h_n\}$ of mutually exclusive and exhaustive hypotheses (of which $\{h, \sim h\}$ is the simplest special case), then their probabilities on e must sum to 1 for all e . For the disjunction $h_1 \vee h_2 \dots \vee h_n$ must be true given e , and hence $p(h_1 \vee h_2 \dots \vee h_n/e) = 1$, and by the addition axiom

$$p(h_1/e) + p(h_2/e) + \dots + p(h_n/e) = 1.$$

(C.2) From entailment and addition, if e falsifies h , that is, entails that $\sim h$ is true, we have

$$p(h/e) = 1 - p(\sim h/e) = 0.$$

(C.3) If h_1 and h_2 are independent, that is, if the truth of one has no relevance to the truth or falsity of the other, then $p(h_1/h_2 \& e) = p(h_1/e)$, and $p(h_2/h_1 \& e) = p(h_2/e)$. The multiplication axiom then gives

$$p(h_1 \& h_2/e) = p(h_1/e) p(h_2/e)$$

which can be taken as the definition of the *probabilistic independence* of h_1 and h_2 .

(C.4) The equivalence and multiplication axioms also yield the following result: if h_1 entails h_2 , then the probability of h_1 cannot be greater than that of h_2 , given any e . That is,

$$\text{if } h_1 \rightarrow h_2, \text{ then } p(h_1/e) \leq p(h_2/e).$$

(C.5) A rearrangement of the multiplication axiom itself is often wrongly called Bayes's theorem — wrongly, because, historically, Thomas Bayes proved a different expression from the multiplication axiom which should properly go by his name. Consider the multiplication axiom expressed as a relation between the initial probability of h and its posterior probability on evidence e , $p(h/e)$. Provided $p(e)$ is non-zero, we have

$$\begin{aligned} (1) \quad p(h/e) &= p(h \& e) / p(e) \\ &= p(h) p(e/h) / p(e). \end{aligned}$$

This is what is often called Bayes's theorem, and we shall follow this historical misuse here.

In some statistical applications, $p(e/h)$ is called the *likelihood* of h on e . So we have informally

Posterior probability of h =

$$\frac{(\text{Initial probability of } h) \text{ times } (\text{likelihood of } h \text{ on } e)}{\text{Initial probability of } e}.$$

In deductive situations, where $h \rightarrow e$, using axiom (3), this expression reduces to

$$\text{Posterior probability of } h = \frac{\text{Initial probability of } h}{\text{Initial probability of } e}.$$

(C.6) It is useful to notice as a corollary of Bayes's theorem that, provided neither $p(e/h)$ nor $p(e)$ is zero, if $p(h)$ is zero, so is $p(h/e)$ for every e , and by replacing h by $\sim h$ it also follows that if $p(h) = 1$, $p(\sim h) = 0$, and hence $p(h/e) = 1$ for all e .

(C.7) Consider now the expression $p(e)$. Taking S as the set of exhaustive and exclusive hypotheses as before, since $h_1 \vee h_2 \dots \vee h_n$ is logically true for this set, we have the logical equivalence

$$e \equiv (h_1 \vee h_2 \dots \vee h_n) \& e.$$

Therefore, by the equivalence axiom

$$p(e) = p((h_1 \vee h_2 \dots \vee h_n) \& e).$$

And by the addition and multiplication axioms

$$(2) \quad p(e) = \sum_{s=1}^n p(h_s) p(e/h_s)$$

or, the prior probability of e is the sum over all h 's of the product of initial probability and likelihood of each h .

(C.8) Bayes's theorem strictly speaking is obtained from (1) and (2):

$$p(h_r/e) = \frac{p(h_r) p(e/h_r)}{\sum_{s=1}^n p(h_s) p(e/h_s)}$$

for all $h_r (r = 1, 2, \dots, n)$.

2. Betting Rates

With these specifications of coherence and probability, equivalence (B) for betting functions has been uncontroversially justified by the theorems of Kemeny and Shimony.⁴ But the probability axioms place only very weak constraints upon the measures that can be assigned to propositions, since except in the circumstances mentioned in (P.3) and (P.3'), particular probability measures are not specified. In fact, if h_s is a set of mutually exclusive and exhaustive hypotheses, given e , then any set of measures $p(h_s/e)$ in the interval $(0,1)$ such that

$$\sum_s p(h_s/e) = 1$$

is consistent with the axioms.

Justification of equivalence (A) is a much more difficult matter. The suggestion that degrees of belief should be explicated as betting behavior originally arose from a behaviorist or empiricist desire to unpack the apparently subjective notion of belief into terms of observable behavior. The highest odds a man is prepared to offer on an unknown outcome is taken to represent his subjective degree of belief in that outcome. Adding the objective requirement of coherence, we obtain the notion of rational betting behavior as an index of degrees of rational belief. "Rationality" is of course still a very weak notion here — a man is rational in this sense if he does not deliberately court loss; it is not necessary that

⁴J. G. Kemeny, "Fair Bets and Inductive Probabilities," *Journal of Symbolic Logic*, 20 (1955), 263; A. Shimony, "Coherence and the Axioms of Confirmation," *Journal of Symbolic Logic*, 20 (1955), 1.

he use any other criteria of comparison between his beliefs. He may assign a probability $5/6$ to six coming up at the next throw or to the constitutional election of an anarchist government in Britain this year; so long as the rest of his beliefs cohere with these assignments he is rational.

The concept of coherent betting is clearly insufficient to determine useful probability assignments to hypotheses, and we shall later have to consider further constraints upon such assignments. Meanwhile, is it even necessary for a confirmation measure? Indeed is it at all applicable? We neglect for a moment the complication that a man might not be willing to back his actual beliefs by corresponding betting odds if he is unwilling to risk much loss under any circumstances. Someone who put an infinite value on his life would hardly ever cross any busy road, however small he believed the chances of accident to be. So we assume the stakes low enough to allow our bettor to contemplate the risk of loss and allow his bets to reflect his actual beliefs. Are his actual beliefs unreasonable if the corresponding betting rates are not coherent?

There are two ways in which an affirmative answer might be justified, one of which seems conclusive but is only dubiously relevant to "games against nature"; the other of which is more relevant but far from conclusive.

Suppose X and Y are tossing a die. X is of an optimistic turn of mind and offers even odds for each fall of the die. He might, of course, win on one or more tosses. But if the rules of the game are such that Y can propose any combination of bets on the same toss at the odds accepted by X , Y has only to place a sum m against each of the six possible falls of the die to be sure of winning an amount $(5/6 \times m/2) - (1/6 \times m/2) = m/3$ whatever the outcome. In the context of fair games the odds were intuitively unreasonable as well as incoherent.

But what has this situation to do with "games against nature"? Here we suppose the scientist asking himself what is his degree of belief in a hypothesis on the basis of evidence by considering how much he would bet on its truth. (We will suppose for the moment that the hypothesis is a singular prediction in order to avoid the difficulty that no bet on a universal hypothesis in an infinite domain can be conclusively settled in a finite time if the hypothesis is true.) Must his betting rates at modest stakes faithfully reflect his degrees of belief, and even if it is assumed that they must and that his degrees of belief are reasonable in an intuitive sense, must the rates be coherent? The argument from ideal

games certainly does not prove either of these assumptions. The hypothesis in question is not the outcome of a closely specified situation which can be made to repeat itself in all relevant respects an indefinite number of times — it may in fact be a once-for-all prediction in conditions never again even approximately repeated. Nature is not an ingenious opponent who seeks out just those bets which will result in certain loss. And there is no clear relation between reasonable expectation of truth and the willingness to gamble on financial reward. The point does not need laboring that there is no close analogy between games and games against nature.

In a second form of the argument we concede most of the disanalogies referred to, but redescribe the gambling situation as one in which an addicted gambler plays many different games for short and long periods, picking his odds at random and playing against a variety of opponents, both subtle and disinterested. Then, by a generalization of the first argument, although nothing is certain, his reasonable expectation on coherent betting odds is to come out approximately even in the long run. This form of the argument is obviously not conclusive for coherence. It refers to expectations not certainties, and it may be intuitively reasonable on some occasions to assume expectations will not happen, and consequently to bet incoherently. We may be in a highly unexpected corner of the world, even in the long, but necessarily finite, run.

Further props can be knocked away from equivalence (A) by removing the restriction to singular predictions that was implicit in the notion of an "outcome." Many philosophers argue that scientific laws and theories are in their essence universal in scope over infinite domains. We shall later find reason to question this, or at least to question its relevance to the logic of confirmation, but if it is accepted, then degrees of belief in laws cannot be represented by betting rates in a game in which there is a fair chance of settling the bet. Nor could the rates in any case be strictly coherent. For, if we assume that a law can be falsified but never shown to be true, any bet upon its truth may be either lost or not settled, but it can never be won. The only odds on its truth a reasonable man would freely choose under these circumstances would surely be 0:1, for since in any case he cannot win anything, any positive sum he places on its truth may be lost, and at best he will simply not have to pay. A zero degree of belief would thereby be assigned to all laws alike, that is, the same degree that is assigned to statements known

to be false. We might suggest forcing the scientist to adopt non-zero coherent odds however small in order to elicit his relative degrees of belief in different laws, but the concept of a forced bet only weakens still further the analogy between reasonable belief and betting behavior in ideal games.

3. Decision-Making under Uncertainty

The concept of coherent betting as a test for rational choice between hypotheses has been very considerably strengthened and generalized in statistical decision theory.⁵ Initially this theory does not depend on inverse probability or the assignment of initial probabilities to propositions. It starts from the relatively uncontroversial assumption that a statistical hypothesis h of the form "If x is A , the probability that x is B is p " determines for h a *likelihood* p which is equal to the probability p of an A being a B if h is true. Suppose now we have a problem of choice between two hypotheses h_1 and h_2 with likelihoods p_1 and p_2 respectively, and we consider the possible outcomes of a test consisting of taking a sample of one individual which is A and observing whether it is B . We can set up a table of likelihoods for possible outcomes on the assumption first that h_1 is true and second that h_2 is true, as in Table 1. Before making the test we decide on a particular *decision strategy*, which is equivalent to a general rule for choosing one of the hypotheses, given a particular outcome of the test. In this case there are four strategies, as follows:

T_{11} :	Choose h_1 on outcome B , and h_1 on outcome $\sim B$.
T_{12} :	" h_1 " B , " h_2 " $\sim B$.
T_{21} :	" h_2 " B , " h_1 " $\sim B$.
T_{22} :	" h_2 " B , " h_2 " $\sim B$.

We can now set up a table for each T of the probability of making the wrong choice if h_1 is true and if h_2 is true (Table 2). By examining this table it can be seen that some of these strategies are "rational" in a sense similar to that adopted in coherent betting, and some are "irra-

⁵ For introductory accounts, see N. Chernoff and L. E. Moses, *Elementary Decision Theory* (New York: John Wiley, 1959); R. B. Braithwaite, *Scientific Explanation* (Cambridge: At the University Press, 1953), ch. 7. See also R. D. Luce and H. Raiffa, *Games and Decisions* (New York: Wiley, 1957), and R. M. Thrall, C. H. Coombs, and R. L. Davis, *Decision Processes* (New York: John Wiley, 1954), especially J. Milnor, "Games against Nature," ch. 4.

		Outcomes	
		B	$\sim B$
h_1 true		p_1	$1 - p_1$
h_2 true		p_2	$1 - p_2$

Table 1

		T_{11}	T_{12}	T_{21}	T_{22}
h_1 true		0	$1 - p_1$	p_1	1
h_2 true		1	p_2	$1 - p_2$	0

Table 2

tional." Suppose $p_1 = 2/3$, $p_2 = 1/4$. Then, whatever hypothesis is true, T_{21} gives a greater probability of wrong choice than T_{12} , and T_{21} is said to be *dominated* by T_{12} . Any strategy which is dominated by some other strategy is said to be *inadmissible*. Strategies which are not dominated by any strategy are said to be *admissible*. In this example T_{11} , T_{12} , T_{22} are all admissible.

The very simple example just described can be generalized by the introduction of a continuum of admissible randomized strategies, defined by some probabilistic method of choosing between the strategies of the original admissible set. The example can also be generalized to cases of more than two hypotheses and more than one individual in the test sample, with corresponding multiplication of the number of possible outcomes and strategies. Further expansion of the concept of a decision theory involves introduction of *utilities* and *initial probabilities*. In most decision situations it is only in the artificial case of low stakes that choice of hypothesis should depend only upon minimizing the probability of false choice. In general we want to minimize the acceptable risk of false choice, whether this is measured in terms of financial, emotional, or some other undesirable loss. Thus our betting behavior would not generally reflect the full degree of our belief where the consequences of losing are in some way disastrous. Someone might not choose to enter into a marriage even though the probability of its success was somewhat greater than its failure. These considerations can be taken into account by defining a *utility function* over the hypotheses between which we are choosing in such a way that the utility of a hypothesis reflects the value to us of the hypothesis being true. Utility values are assumed additive for mutually exclusive hypotheses. In our example let the utilities, which are taken constant for each hypothesis, be represented by u_1 , u_2 for h_1 , h_2 respectively. Then if we are interested in somehow minimizing expected loss or utility, we shall be concerned with the products of u_1 or u_2 with the probabilities of false choice if h_1 or h_2 respectively are true, rather than with the probabilities of false choice themselves.

Initial probabilities may be introduced by noticing that if in calculating the expected loss of utility resulting from false choices of h_1 or h_2 , we had had some means of knowing that their initial probabilities were unequal, say equal to v , $(1 - v)$, respectively, then the expected losses of utility would have been multiplied in this ratio. This consideration shows that although initial probabilities do not appear in the formulation of the decision problem, since likelihoods of hypotheses on evidence are the only probabilities admitted in this formulation, nevertheless there is a mathematical equivalence between the decision problem and calculations based on the assumption that the initial probability v is equal to $1/2$.

Thus, if initial probabilities are admitted, the probabilities of false choice of Table 2 would be multiplied through in the first row by $w_1 = u_1 v$, and in the second row by $w_2 = u_2 (1 - v)$. In fact only these products w_1 , w_2 can affect subsequent choice of strategy and hypothesis. With the new representation of expected loss of utility exemplified by Table 3 we can now consider what criteria might be applied to selection among admissible strategies.

	T_{11}	T_{12}	T_{21}	T_{22}
h_1 true	0	$w_1(1 - p_1)$	$w_1 p_1$	w_1
h_2 true	w_2	$w_2 p_2$	$w_2(1 - p_2)$	0

Table 3

There is one type of admissible strategy with uniquely desirable properties. This is the so-called *minimax strategy*, which is obtained by finding the maximum expected loss in each strategy column, and picking that strategy or one of those strategies for which this maximum is least. The minimax strategy ensures that, independent of which hypothesis is true, the expected loss is smaller than the worst possible loss on any other strategy. In this sense the strategy is prudent, not to say timid. In general it leads to playing safe; if the decision is either to take action with risk or to take no action with no risk, it will choose inaction, and between hypotheses of equal prior probability and utility it will choose that which goes as little as possible beyond the evidence. Therefore in scientific contexts it is not often an interesting strategy. But it is not forced upon us by probabilistic decision methods, and it may be remarked parenthetically that the general criticism leveled, in particular

by Popper, against probabilistic methods as always selecting the weakest hypothesis is therefore groundless.

The prudence of the minimax strategy is bought at the cost of the possibility of greater expected gain if the true hypothesis is chosen, and also of course by guarding against the possibility of greater expected loss if the false hypothesis is chosen. Introduction of a greater element of risk in exchange for higher potential winnings suggests averaging the expected losses on each strategy, and taking that strategy which minimizes the average instead of the maximum expected loss. In the special case of $w_1 = w_2$ this strategy is called the *maximum likelihood strategy* ("maximum" referring, with greater optimism than minimax adherents are likely to exhibit, to the possibility of gain rather than loss). The strategy is so-called after Fisher's method of hypothesis-choice, which asserts that the best hypothesis is that whose likelihood on the observed sample is a maximum. That is to say, in the above example, if the sample is B the best hypothesis is h_1 , while if the sample is $\sim B$ the best hypothesis is h_2 , giving T_{12} as the maximum likelihood strategy.

Fisher's method was developed in response to objections to inverse probability methods depending on initial probabilities of hypotheses. But, as we have seen, initial probabilities and also utilities can be introduced into a more general viewpoint, which yields Bayesian decision theory. When maximum likelihood is extended to become a method of maximizing expected gain of utility, the best strategy in this sense is called a *Bayes strategy*, and is obtained by minimizing the average value of the losses in each column of Table 3 and selecting the strategy that corresponds to this minimum loss or maximum gain. The name "Bayes" is appropriate because it can be shown that the Bayes strategy with parameters $w_1, w_2, \dots, w_n$ for n hypotheses always selects that hypothesis which would be chosen on given evidence by calculating the posterior probabilities of all the hypotheses by Bayes's theorem and choosing that hypothesis whose posterior probability is maximum, provided the initial probabilities assumed in this calculation are in the ratios $w_1: w_2: \dots : w_n$. Moreover, it can be shown that every admissible strategy is a Bayes strategy for some set of w 's (all positive), and conversely that Bayes's strategies for positive w 's are always admissible.

This brief account of the concepts of elementary decision theory is sufficient to indicate its generality as a theory of hypothesis-choice. It is tempting to conclude that if any theory of confirmation is going to

be adequate for scientific inference, then Bayesian decision theory will be. The theory is certainly preferable to the analogy of confirmation with betting rates, since it removes some of the artificial restrictions under which that analogy describes "games against nature." It indicates how to take account of utilities if these are known. It is not, however, obvious how the notion of "utility" should be interpreted in relation to pure science, since at least monetary gains and losses do not seem to be directly involved. It might be held that all theories are equally valuable if true, so that utilities can be neglected. Or the notion of "epistemic utility" may be adopted, depending on the range or content of a hypothesis, thus incorporating the view that a scientific theory is more valuable the more powerful it is.⁶

With regard to the more basic objections leveled against the rationality of coherent betting rates, however, we are in no better case with admissible strategies. In the short run (and compared with the universal or near-universal pretensions of science all runs are short) inadmissible strategies may give more right answers than admissible ones. Suppose in the example above we adopt strategy T_{21} . Since we don't know whether we are in a typical corner of the universe, we don't know that in a comparatively small number of choices we are going to be faced with samples which will lead to greater loss than on, say, T_{12} . We may be presented with a series of B 's and no not- B 's, and yet h_2 may be true. Let us simplify by supposing that each sample is forgotten after it is observed, so that the choice of hypotheses starts afresh at each sampling. This enables us to consider the example as a sequence of independent single-sample decisions. Under these circumstances T_{21} will be right every time and T_{12} will be wrong. There are, however, two things that can be said in support of the irrationality of T_{21} in this case. First, if h_2 is true it becomes decreasingly probable with larger samples that we shall continue to sample not- B 's and therefore that T_{21} will always choose correctly. And more telling than this long-run justification is the fact that if T_{21} is always successful for a given finite sample, there is always some admissible strategy which would have been equally successful for that sample, namely one which makes h_2 initially very probable. And in general, for any finite sample, there is always some set of initial prob-

⁶ Suggested by C. G. Hempel, "Inductive Inconsistencies," *Aspects of Scientific Explanation* (New York: Free Press, 1965), p. 76, and developed by I. Levi, *Gambling with Truth* (New York: Knopf, 1967), ch. 5ff.

abilities which will choose the true hypothesis each time by Bayesian methods, if only we knew what that set is. But if a Bayes strategy with, say, $1 - v$ very near to zero is considered, it might be that no justification can be found for assuming h_2 so antecedently improbable, and a decision-maker who "irrationally" adopted T_{21} might prefer to describe his behavior in terms of a lucky break with an inadmissible strategy.

4. Bayesian Transformations

There is a further important respect in which the arguments for coherence and admissibility establish weaker conclusions than might at first sight appear. To see what this is, we need a more explicit formulation of the decision procedure. If only admissible strategies are adopted and utilities are assumed equal, it follows that any hypothesis-choice is practically equivalent to maximizing the posterior probability conditional upon evidence, where this posterior probability measure is defined over the set of all possible hypotheses. Admissibility requires only that this conditional measure satisfy the probability axioms.

Consider the position of the decision-maker at a given point of time, say t_0 , when he has a certain amount of evidence, e_0 , and a probability distribution over hypotheses. Suppose he adopts Bayes's strategy T . T tells him, for each possible combination of evidence statements he may subsequently observe to be true, which hypothesis he should choose, and it does this in accordance with the rule of taking the maximally probable hypothesis for each combination of evidence, which is equivalent to using Bayes's theorem. In the case of a bettor this means that he is prepared at time t_0 with coherent odds for any bet his opponent may suggest conditional upon whatever evidence may be observed. Under these circumstances the bettor cannot be presented with a Dutch book, and the decision-maker does not face the expectation of loss whatever the state of nature. It is very important to stress that all this refers only to possibilities, none of which are yet realized at t_0 . It refers to what the decision-maker should be prepared to choose in hypothetical circumstances before more evidence is collected. It describes the static situation at t_0 in terms of a certain probability distribution satisfying the probability axioms.

The situation is, however, quite different when more evidence is collected at time t_1 . Denote this new evidence by e_1 . There is now a new

decision situation in which a probability distribution $p_1(h/e_o \& e_1)$ is defined over all h consistent with e_o & e_1 . Admissibility demands only that p_1 be a probability measure; it does not demand that p_1 be related to p_o by the equations

$$p_1(h/e_o \& e_1) = p_o(h/e_o \& e_1) = \frac{p_o(h/e_o) p_o(e_1/h \& e_o)}{p_o(e_1/e_o)}$$

for all mutually consistent h, e_o, e_1 . In other words, p_1 need not be related to the initial distribution of probabilities p_o in accordance with axiom (P.5). Only if it is so related, is the transformation Bayesian.⁷

Since the coherence requirement has nothing to say about what particular distribution of initial probabilities should be adopted, it cannot prevent this distribution from being changed with new evidence. Indeed a man may change his betting rates without new evidence by just changing his mood, or thinking more deeply, and yet remain coherent. Coherence does not merely leave the question of initial probabilities completely open; it also allows them to be changed for any reason or no reason, in ways that do not satisfy the Bayesian transformation. Consequently we have to envisage two different kinds of situation that may face the decision-maker. He may have to choose a set of hypotheses in various domains all at once at t_o , when his actual evidence is e_o . This corresponds, for example, to the scientist's situation in predicting the outcome of a complex of future experiments. In such a case arguments for admissibility indicate that he should adopt a Bayesian strategy, and this involves calculating conditional probabilities and maximizing. But on the other hand the decision-maker may be in a situation in which he is allowed to collect evidence sequentially. Then he has not only to select a particular Bayesian strategy T_o at t_o , but also to select strategies at $t_1, t_2, \dots$, as evidence $e_1, e_2, \dots$, comes in. These subsequent strategies may be the same as or different from T_o . The difference between the situations is that between prediction and learning. The learner, of course, also remains a predictor, because it is assumed that he is not merely soaking information in as time goes on, but learning from it to

⁷ The non-derivability of Bayesian transformations from coherence was explicitly pointed out by J. M. Vickers, "Some Remarks on Coherence and Subjective Probability," *Philosophy of Science*, 32 (1965), 32, and I. Hacking, "Slightly More Realistic Personal Probability," *Philosophy of Science*, 34 (1967), 311. For a detailed discussion see D. H. Mellor, *The Matter of Chance* (Cambridge: At the University Press, 1971), ch. 2.

adapt to the future. And the learner has the predictor's problem of choice of initial probabilities, renewed on each occasion upon which he receives new evidence.

To conclude this brief survey, the best that can be said for a decision-theoretic formulation of confirmation theory is that it provides a sufficient framework in terms of which any choice-problem and subsequent decision could be represented, though possibly at the cost of highly implausible assumptions about utilities, initial probabilities, and posterior transformations. But clearly in itself this representation is vacuous as a method of induction; for unless there are some restrictions upon the initial distribution and some grounds for using a Bayes or some other posterior transformation rule on that distribution, the effect of evidence on the posterior distribution is quite arbitrary. Even if the overall probability distribution is required to be coherent or strictly coherent, that is, so long as it satisfies the probability axioms, a long run of black crows, for example, might be represented as good evidence for the statistical generalization "Most crows are white," if only the probability of this generalization was initially high enough or if Bayes's theorem is held to be irrelevant to the posterior transformation of probabilities. Rational decision procedures are not sufficient to capture any interesting *inductive* sense of rationality, for which further constraints must be found.

III. Non-Bayesian Transformations

The possibilities of assignment of initial probabilities and posterior transformations have been left wide open by the arguments of the previous section. The weakest possible response to this situation, which might be called pure personalism, is the view that the succession of probability distributions over hypotheses held by a believer is a purely *physical* or causal matter of belief-response to stimuli deriving from the evidence. If no normative constraints are put upon belief distributions other than coherence, this is equivalent to adopting an inductively arbitrary, and arbitrarily changing, initial probability distribution as a result of each new piece of evidence, and cannot in itself yield inductively or predictively useful beliefs except by accident. No serious inductive logician could be a pure personalist in just this sense, and in fact those who have emphasized the ultimately physical nature of the effect of evidence on

belief have also placed other normative constraints upon the sequence of probability distributions.

An inductive theory of this type is suggested by Richard Jeffrey,⁸ who makes a radical departure from Bayesian theory by rejecting the notion of incorrigible evidence. In orthodox Bayesian theory it is assumed that the posterior probabilities are conditionalized upon evidence which must be taken to have probability unity once it has been observed, that is to say, after observation of evidence e_r , the initial probability of e_r is necessarily replaced by the posterior probability $p(e_r/e_r) = 1$. But, Jeffrey suggests, observation is not always such a conclusive affair. We may not be quite sure whether we have observed e_r or not; we may be slightly more sure of its truth than before, but the light may have been bad, and we may wish to leave open the possibility that at some future time we may be led to withdraw e_r from the evidence under pressure of other evidence or at least to give it probability less than 1. A Bayesian might attempt to accommodate this idea by allowing withdrawals as well as accumulations of evidence in his sequence of posterior distributions, but he cannot deal in this way with "partly observed" evidence to which it is appropriate to give a probability value between 0 and 1. Or he might attempt to get beneath the uncertain evidence to something more conclusive which can be given probability 1 and can communicate a higher probability still short of 1 to the uncertain evidence. But any such attempt would be reminiscent of the search for more and more basic incorrigible observation statements, a program which has met conclusive refutation in recent philosophy of science.

Jeffrey's solution is a radical personalism in which not only is the initial distribution a physical rather than a rational bequest, but also every subsequent distribution shares this physical character. As we go around with our eyes more or less open, our degrees of belief in various propositions are physically caused to change. We usually cannot specify any evidence that can be asserted conclusively; thus the sequence of distributions is best not specified by conditional probabilities, but merely by suffixes indicating the order of the sequence. The distribution $p_r(h_s)$, for example, is the equivalent of the posterior distribution at t_r , and this distribution includes such probabilities as $p_r(e_j)$ with values less than 1, which are physically related to the uncertain and inexpressible results of

⁸ R. Jeffrey, *The Logic of Decision* (New York: McGraw-Hill, 1965), chs. 11, 12. I have somewhat modified Jeffrey's notation in what follows.

observation. Observation-type propositions are distinguished from hypothesis-type propositions by being those propositions in which the changed probabilities physically originate. Changes are disseminated over the distribution for hypothesis-type propositions in such a way that coherence is preserved, according to the set of equations

$$p_r(h_s) = p_r(h_s/e_j) p_r(e_j) + p_r(h_s/\sim e_j) p_r(\sim e_j),$$

where e_j is the only evidence-type proposition whose probability has changed, or according to a corresponding set of equations in the general case where several probabilities of evidence have changed. Jeffrey does require that in the sequence of distributions the conditional probabilities such as $p_r(h_s/e_j)$ remain invariant, so that for these probabilities the r -suffix can be dropped, and the above equation becomes

$$(3) \quad p_r(h_s) = p(h_s/e_j) p_r(e_j) + p(h_s/\sim e_j) p_r(\sim e_j).$$

Jeffrey seems to introduce this condition as just a device for distinguishing between those propositions which are directly affected by observation and those which are not.⁹ If $p_r(h_s/e_j)$ does not satisfy the stability condition for some h_s , then this h_s must be included among those propositions in which changes of probability originate. It might be objected that if h_s is hypothesis-like, it is contrary to the spirit of the physical account (which is an attempt to retain a vestige of empiricism) to suppose that our belief in a hypothesis can be directly affected by observation in a way not directly related to the probabilities of evidence for that hypothesis. But even if it is accepted that there must be some rule for dissemination of probability-changes from evidence to hypotheses, why choose the invariance of conditional probability and not, for example, the likelihood $p_r(e_j/h_s)$ or some other function of the $p_r - 1$ -distribution?

Jeffrey's attempt to deal with the problem of uncertain evidence is courageous and ingenious.¹⁰ Unfortunately it presupposes only coherence

⁹ In private correspondence, which he kindly permits me to quote, Jeffrey tells me that he intends the condition just as one which defines an interesting situation, namely, that in which the observer judges that he made an observation directly relevant to e and to nothing else.

¹⁰ I have proposed an alternative, Bayesian, method of accommodating uncertain evidence without assuming a stable observation language in "A Self-Correcting Observation Language," in B. van Rootselaar and J. F. Stahl, eds., *Logic, Methodology, and Philosophy of Science*, vol. 3 (Amsterdam: North Holland, 1968), 297, and in "Duhem, Quine, and a New Empiricism," *Knowledge and Necessity*, Royal Institute of Philosophy Lectures, vol. 3 (London: Macmillan, 1970), 191. After making my

constraints upon the sequence of distributions and could hardly form a sufficient basis of a confirmation theory. First, it should be noticed that in Jeffrey's theory, if (3) holds and $p_r(e_j) \neq p_{r-1}(e_j) \neq 1$, then the likelihoods of statistical hypotheses with respect to e must be assumed to change. For we have

$$p_r(e_j/h_s) = p_r(e_j)p(h_s/e_j) / p_r(h_s)$$

and by (3) this is not equal to

$$p_{r-1}(e_j/h_s) = p_{r-1}(e_j)p(h_s/e_j) / p_{r-1}(h_s).$$

Some personalists and most objectivists have regarded the likelihoods of statistical hypotheses as the most "objective" of the various probability functions entering Bayes's theorem.¹¹ Of course, Bayesians themselves replace $p_{r-1}(e_j/h_s)$ by $p_r(e_j/h_s \& e_j) = 1$ when e is said to have been observed. But for them there is at all times t_r a constant conditional likelihood $p(e_j/h_s)$ which is dependent only on the statistical hypothesis h_s and a random sample e_j , whether or not e_j has been observed at t_r . This constant likelihood cannot be expressed by Jeffrey, because all his probabilities, including likelihoods, are in effect conditional upon unexpressed (and inexpressible) evidence. The probability of e_j on h_s is well defined only if the truth of h_s is the only factor upon which e_j is conditional. Where the value of $p_r(e_j/h_s)$ in Jeffrey's theory differs from this "objective" likelihood, e_j must be construed as a sample conditional not only upon h_s but also upon some other unexpressed evidence, for example that only part of the domain of h_s is being sampled. It must be conceded to Jeffrey that such nonrandom sampling is nearer to practical scientific experimentation than the ideal situations of statistics, and indeed I shall accept this argument in the next section.

The most serious drawback to Jeffrey's proposal from the point of view of confirmation theory, however, is that it allows no control over confirmation relations. It is impossible to calculate, except conditionally at t_{r-1} , what the effect of e_j upon the probability of h_s will be when e_j has actually been observed at t_r . The conditional probability $p_{r-1}(h_s/e_j)$ bears no regular relation to $p_r(h_s)$, because the effect of actually

suggestion I discovered that essentially the same method is to be found in C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle, Ill.: Open Court, 1946), p. 235f.

¹¹ This point has been made against Jeffrey by I. Levi, "Probability Kinematics," *British Journal for the Philosophy of Science*, 18 (1967), 197.

"observing" e_i has an unpredictable effect upon the whole distribution. But to have a confirmation theory at all we must be able to answer explicitly some general questions concerning the effect of evidence upon the probability of hypotheses. For example, we may wish to calculate, on our present belief distribution, what will be the effect on our belief in given hypotheses of the future observation of a certain piece of evidence. This might be done by means of Bayes's theorem or by some other posterior transformation so long as we have a rule for it which can be applied at t_{i-1} to calculate the effects of evidence at t_i . But Jeffrey's proposal provides no rule at all, neither Bayesian nor any other, and if there is no rule, no answer can now be given to questions of the type: Which experiment will most usefully test a given hypothesis, or Which hypothesis will be most acceptable on the basis of specified future evidence? Without some such rule, there can be no confirmation theory.

It does not yet follow, however, that the rule must be Bayesian. An attempt to provide an a priori argument for Bayesian transformations has been made by Abner Shimony,¹² by examining a purely analytic derivation of the probability axioms, where the derivation does not depend on the notion of coherent betting. The proof concerns only values of probability assumed to hold at a given time, but it might be adapted to the discussion of Bayesian transformations if its conditions were applicable to posterior probability at time t_1 in terms of initial probabilities at t_0 , as well as to the conditional probabilities at t_0 which yield the multiplication axiom. The nontrivial assumptions involved in the proof assert a functional dependence of such expressions as $p(h \& e)$ upon $p(h/e)$ and $p(e)$, and of $p(h \vee h')$ upon $p(h)$ and $p(h')$, where h , h' are mutually exclusive. But Shimony concludes that such analytic assumptions, weak as they are, strain our intuitions about belief too far to be cogent a priori, even where axioms relative to a given evidence-set are concerned. A fortiori, it seems unlikely that such analytic arguments are capable of yielding Bayesian transformations from a belief distribution on one evidence-set to a belief distribution on another.

Shimony suggests that the justification of Bayesian methods must therefore rest on investigation of their consequences and the consequences of alternatives. I shall pursue the first type of investigation in sections V and VI below. As for the second, there is so far little to

¹² "Scientific Inference," in R. G. Colodny, ed., *Pittsburgh Studies in the Philosophy of Science*, vol. 4 (Pittsburgh: University of Pittsburgh Press, 1970), p. 79.

report, since few discussions of confirmation theory have explicitly abandoned Bayesian transformations and replaced them by an alternative rule. The most notable exception to this general neglect is Reichenbach's so-called "straight rule" of inference, and the related suggestions in Hempel and Oppenheim's early "Definition of 'Degree of Confirmation.'" ¹³ Reichenbach's rule is equivalent in the cases to which it is applicable to the maximum likelihood method, that is, it is the direction to expect the proportion of occurrences of a given property in the population to be the same as that in the observed sample, and this does not yield Bayesian transformations. The rule is not, however, generally applicable without further assumptions to probability distributions over all hypotheses, since it specifies only which the most likely hypothesis is on given evidence (in the technical sense of likelihood mentioned in C.5, Section II). It does not specify the distribution of probabilities on that evidence over competing hypotheses. To obtain this, as with maximum likelihood itself, some initial probability distribution must be assumed, even if this is a uniform distribution. The rule also suffers from the drawbacks of the maximum likelihood method for small samples, for example, it gives probability 1 to the next positive instance of a universal generalization on the basis of only one or a small number of observed positive instances.

Examples of restricted Bayesian methods are given by Suppes ¹⁴ in an analysis of various types of learning in behavioral psychology, where the constraints upon transformations necessary for learning are inferred in models simulating the actual behavior of organisms. In a simple case of choice between two possibilities in a sequence of trials with reinforcement for correct choice, the rule may be to guess on each occasion that possibility which turned out correct on the previous trial. This rule corresponds to "forgetting" all evidence before the immediately preceding trial, but if the probabilities of occurrence remain the same from trial to trial, and one possibility occurs more frequently than the other, eventually more guesses will be correct than not, and learning has been partially successful.

In general, the rule to forget all or most previous evidence, and use

¹³ Reichenbach, *The Theory of Probability*, ch. 11; C. G. Hempel and P. Oppenheim, "Definition of 'Degree of Confirmation,'" *Philosophy of Science*, 12 (1945), 98.

¹⁴ P. Suppes, "Concept Formation and Bayesian Decisions," in J. Hintikka and P. Suppes, eds., *Aspects of Inductive Logic* (Amsterdam: North Holland, 1966), p. 21, and "Probabilistic Inference and the Concept of Total Evidence," *ibid.*, p. 49.

the posterior distribution at t_0 as if it were the initial distribution in calculating the effect of subsequent evidence, is almost inescapable in any situation that is at all complex. It is certainly exemplified in the state of belief of the scientific community at any given time, which is far more dependent on theories that have recently proved successful than upon the detailed evidence which originally led to their acceptance and which may be forever lost in the back numbers of journals and the vaults of libraries. The rule to forget, however, is not Bayesian, for calculation of posterior probabilities by Bayes's theorem requires total evidence to be taken into account, and failing to take it into account may make the time-order of collection of evidence relevant to posterior distributions. If the total evidence of science since Galileo had been collected backwards, our beliefs calculated by the rule of forgetting would not now be the same as they are. (However, it must be remembered that the decision what to collect at any given time is itself dependent on the order in which evidence and theories have emerged, so this conclusion is not as paradoxical as it may look at first sight.) That Bayes's theorem cannot be a general requirement for historical development of scientific theories follows most obviously in cases of observation of new properties and invention of new concepts. If these have not previously been named in the language, then no hypotheses containing them can have explicitly been given any initial probability. Conversely, terms may disappear from the scientific language (for example, "phlogiston"), thus removing some hypotheses by giving them probability identically zero and causing non-Bayesian readjustment in the probabilities of others.

Much work remains to be done on the question of the relative efficiency of Bayes and other transformation rules in given sorts of circumstances. It may be assumed that which rules are most efficient will be a function both of the complexity of the learning mechanism (the size of its memory, for example), and also of the general characteristics of the world in which it is learning. With sufficiently complex learning devices one can envisage a "monitor" which checks the predictive success attained by a lower level probabilistic inductive device and changes the initial probabilities until better success rates are attained. Something like this process is suggested by Carnap for choosing a probability distribution out of his continuum of different distributions defined by a param-

eter λ in the real open interval $(0, \infty)$.¹⁵ He shows that, if the universe is on the whole homogeneous, the distribution defined by $\lambda = 0$ will be expected to give greatest success in the long run, while if the universe is grossly heterogeneous, λ infinite will give greatest expected success. $\lambda = 0$ is in fact equivalent to the statistical method of maximum likelihood, and, in simple cases, to Reichenbach's straight rule, which is the direction to predict that the proportion of properties in the world will continue to be the proportion that has already been observed in samples. The success of this method clearly relies heavily on the assumption that the world is relatively homogeneous with respect to the samples, especially when these are small.

But of course the essence of the inductive problem is that we do not know a priori whether the universe is on the large scale homogeneous or heterogeneous, whether we are in a typical part of it, or how far the past success of any inductive method will be maintained in the future. If choice of pragmatically efficient initial distributions and transformation rules depends crucially on how well adapted the learning organism is to actual contingent conditions in the world, then it is unlikely that any detailed confirmation theory can be developed a priori, that is to say, independently of empirical investigations into how organisms (including ourselves) actually learn. It cannot be expected, therefore, that Bayesian methods will be useful except locally. But before going on to investigate how far they will carry us in this situation, I shall consider one further influential attempt to show that a priori conditions are sufficient for establishing a normative confirmation theory.

IV. Convergence of Opinion

Some personalist statisticians have attempted to provide a long-run justification of all types of scientific inference by adopting a single constraint on initial distributions, namely, that all hypotheses to be confirmed have a finite initial probability. Appeal is made to a convergence theorem for statistical hypotheses, according to which, under certain conditions that must be examined, the posterior probabilities of such hypotheses tend to 1 or 0 with increasing evidence for true or false hypotheses respectively, irrespective of the numerical values of the initial

¹⁵ R. Carnap, *The Continuum of Inductive Methods* (Chicago: University of Chicago Press, 1951), secs. 18, 22.

probabilities of these hypotheses, provided only that the initial probabilities of true hypotheses are not less than some small fixed value. This theorem, if relevant to scientific inference in general, would be a very powerful aid to confirmation theory, for it would discharge us from discussing the details of the initial probability distribution if only this one condition upon it were satisfied.

Attempts to apply the convergence theorem in the context of confirmation theory are, however, subject to several kinds of misunderstanding and overoptimism. Thus, Savage claims to show rather generally how a person "typically becomes almost certain of the truth" when the amount of his experience increases indefinitely, and Edwards, Lindman, and Savage remark in a well-known expository article, "Although your initial opinion about future behavior of a coin may differ radically from your neighbor's, your opinion and his will ordinarily be so transformed by application of Bayes' theorem to the results of a long sequence of experimental flips as to become nearly indistinguishable."¹⁶

Assimilation of such results to scientific inference is certainly tempting, if only because it gives access to a large and familiar body of mathematical theory and technique. But personalist formulations of this kind, when applied to science, neglect the very stringent conditions on hypotheses and evidence under which the relevant convergence theorem is true and therefore omit to take account of the relevance of these conditions to typical cases of scientific inference. The second quotation also gives the misleading impression that Bayesian transformations are required for convergence of opinion. If this were the case it might be held to provide an argument for adoption of Bayesian transformations along the following lines. Inductive intuition suggests that in a satisfactory confirmation theory a true hypothesis should be one whose posterior probability tends to 1 after collection of a large amount of evidence. The convergence theorem says that this is the case under a number of necessary conditions. If these conditions included Bayesian transformations, it would be reasonable to adopt such transformations, since they would be necessary for explicating the inductive intuition. But, as we shall see, Bayesian transformations are *not* among the necessary conditions of the theorem. This indeed makes the theorem rather strong, since it is inde-

¹⁶ L. J. Savage, *The Foundations of Statistics* (New York: John Wiley, 1954), p. 46; W. Edwards, H. Lindman, and L. J. Savage, "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, 70 (1963), 197

pendent of choice of nonzero initial probabilities at every stage of collection of evidence, and not just at the first stage. However, the remaining necessary conditions of the theorem do restrict its application in scientific contexts quite stringently. We shall now examine these conditions in detail.¹⁷

Consider a finite set S of N mutually exclusive and exhaustive hypotheses $\{h_s\}$. Let E_r be the conjunction of all evidence propositions $e_1, e_2, \dots, e_r$ collected before time t_r . The convergence theorem proceeds as follows. By means of a law of large numbers it is first shown that, for h_1 true, the likelihood ratio $[p(E_r/h_1)]/[p(E_r/h_s)]$, for all $s \neq 1$, becomes greater than any fixed number η , however large, with second-order probability which approaches 1 as r increases. Then assuming for the moment the Bayesian transformation (which will be relaxed later), we have

$$(4) \quad p(h_s/E_r) = \frac{p(h_s) p(E_r/h_s)}{\sum_s p(h_s) p(E_r/h_s)} \quad (1 \leq s \leq N).$$

It follows at once that if $p(h_s) = 0$, no finite observation can make $p(h_s/E_r)$ nonzero, that is, if h_s is judged a priori to be impossible, no evidence can make it probable.

Suppose, however, that $p(h_s) > 0$ for all s . Consider the second-order probability P that $p(h_1/E_r) > \alpha$ for any fixed α as near to 1 as we please if h_1 is true. Using the result above for the likelihood ratios, it can be shown that, after a sufficiently large number of observations yielding E_r , P approaches 1 as r increases. It follows also that the posterior probabilities of all false hypotheses of the set S , given E_r , become as near to zero as we please with probability which approaches 1 as r increases.

The significance of the convergence theorem, more informally, is that, with sufficient evidence, the posterior probability of a true hypothesis is overwhelmingly likely to approach 1 and that of a false hypothesis to approach zero, whatever the set of nonzero initial probabilities when Bayesian transformations are assumed. The convergence of the second-order probability ensures not only that there is an E_r such that $p(h_1, E_r) > \alpha$ for fixed α , but also that it becomes progressively less probable that E_r is freak evidence which only temporarily favors a false hypothesis. Thus, the posterior probability of a true hypothesis is not only overwhelmingly likely to approach 1, but also to remain near 1 as evidence increases.

¹⁷ See Savage, *The Foundations of Statistics*, pp. 46–50.

Conversely, if the posterior probability of a given hypothesis approaches 1 as evidence increases, there is only a small and decreasing second-order probability that it is not the true hypothesis. Moreover, if there are non-zero lower limits on the $p(h_s)$, lower limits of these second-order probabilities can be calculated. If, however, no lower limits on the $p(h_s)$ can be stated, nothing can be said about the speed of convergence, and in any finite time any given hypothesis of the set will have the highest probability for some distribution of initial probabilities. Thus for the theorem to be of any use in a finite time (that is, for it to be of any use at all), not only must the initial probabilities be nonzero, but they must also have some fixed finite lower limit, say ϵ , as near to zero as we please.

Under this condition, the requirement of Bayesian transformation can itself be relaxed. Denote probabilities at time t_r by p_r and assume fixed likelihoods, that is,

$$p_0(e_i/h_s) = p_1(e_i/h_s) = \dots = p(e_i/h_s).$$

The limiting property of the likelihood ratios then remains as before, but we cannot now assume that $p_r(h_s) = p_{r-1}(h_s)$. The ratios (4), however, now yield

$$p_r(h_1/E_r) > a'p_r(h_1)$$

for any fixed a' near 1 and h_1 true, with second-order probability approaching 1 with increasing r . But since $p_r(h_1) \geq \epsilon$, all r , this means $p_r(h_1/E_r) > a$ with a as near to 1 as we please, and we have the convergence theorem as before, but without the Bayesian assumption of constant initial probabilities. In other words, as long as the lower limit of $p(h_1)$ is fixed, however small it is, the posterior probability provided by sufficient evidence for the true hypothesis will approach maximal probability.

Now let us investigate the conditions of the theorem in more detail in relation to a set of scientific hypotheses. The conditions are as follows:

(i) $p(h_s) \geq \epsilon > 0$, all s . The possibility of satisfying this condition for scientific hypotheses will have to be investigated later on; meanwhile we assume it satisfied.

(ii) No two hypotheses, $h_s, h_{s'}$, yield the same value of $p(e_i/h_s)$ for all possible e_i . Or, if this condition is not satisfied by different hypotheses $h_s, h_{s'}$, then only the disjunction of h_s and $h_{s'}$ can be confirmed by any e_i .

(iii) Given a particular hypothesis h_s , the probability of making the test which yields e_i is independent of the particular order in which it and other evidence are observed. This will be called the *randomness assumption*. It is a substantial assumption in scientific contexts, because it is often known to be false and usually not known to be true; for it contemplates a situation in which we know nothing about the conditions which differentiate one observation from another. If, for example, the structure of the universe includes the true hypothesis "A proportion p of aircraft are jets," and we know that we are sampling aircraft, and this is *all we know*, then the probability that this aircraft is a jet is p on any occasion of observation. But the situations in which we wish to use the convergence theorem are unlikely to be so simple, because we are likely to want to confirm h_s on evidence which is known not to be random in this sense. This might be because we are able to sample only over a limited part of the spatial and temporal domain of h_s , or only over certain sorts of instances of h_s (the "observable" as opposed to the "theoretical" instances, for example). More obviously, there is something paradoxical about a requirement of random sampling in the context of scientific experiment, because it is universally accepted that science requires deliberately structured experiments, not random observation, and that this requirement is precisely what differentiates science from statistics. The difference of method must be regarded as a *consequence* of the nonapplicability of many statistical assumptions in the domain of science, not as its cause. Its cause lies rather in the inaccessibility of the whole domain to random sampling and to the desire to reach conclusions in as short a run as possible. That this is so is shown by the fact that where the random sampling assumptions seem reasonable or even inescapable, as in social surveys, star counts, genetics, and the like, statistical methods form an integral part of scientific procedure. Where these assumptions are not reasonable, convergence is inapplicable and structured experiment is required.

(iv) The e_i 's are independent given h_s . That is, for all s ,

$$p(e_1 \& e_2/h_s) = p(e_1/h_s) p(e_2/h_s).$$

I shall call this the *independence assumption*. Like the randomness assumption, it is not always valid in scientific contexts. It implies that the probability of observing a particular piece of evidence is unaffected by having already observed particular other pieces of evidence, an assump-

tion that is satisfied, for example, by "random sampling with replacement" of balls out of a bag containing a finite number of balls. But, on the contrary, scientific experiments are often deliberately designed in the light of experimental evidence already obtained. Rather than conducting a large number of independent tests of the same hypothesis, scientific hypotheses are often themselves modified in the light of evidence, and new tests devised for the new hypothesis, with the results of previous evidence assumed. Especially where the hypothesis is universal rather than statistical in form, and may therefore need modification as a result of *falsifying* evidence, a limited number of experiments structured in the light of previous evidence, and directed to progressively modified hypotheses, is a much more efficient procedure than blindly conducting a long series of independent tests of the same hypothesis, such as might be appropriate for statistical hypotheses concerning large populations.

Comparison of "convergence of opinion" with the conditions of scientific inference therefore yields the following results: where the convergence theorem applies there is indeed convergence of opinion with regard to hypotheses which have nonzero probability, and this convergence is independent of the actual distribution of initial probabilities and of Bayesian transformations. But the conditions of the theorem, especially the randomness and independence assumptions, are not valid for typical examples of scientific inference in limited domains and from controlled experiments.

We have now investigated three types of attempt to minimize constraints upon confirmation functions, namely, decision theory, non-Bayesian methods, and convergence of opinion. All of them have turned out in some respects inadequate. It is time to make a bolder approach and to reformulate the confirmation problem as one of finding some theory which is *sufficient* (though not necessary) to explicate intuitive criteria of confirmation in science. In this reformulation we shall not be afraid to adopt both the probability axioms and the Bayesian transformation and to draw further constraints from intuitive criteria of adequacy for inductive inference. The justification of such a confirmation theory will be in its consequences rather than its premises.

V. Criteria of Adequacy

1. *Confirmation as Positive Relevance*

What should be expected from a confirmation theory which is regarded as one among the possible explications of inductive inference rather than as an a priori and necessary framework for such inference? There are three types of conditions it should satisfy.

1. It should explicate and systematize inductive methods actually used in science, without grossly distorting any of them. This will generally imply clarifications of confused inductive intuitions and of the relations between them, and insofar as such clarification is provided, confirmation theory will escape the charge of being an ad hoc rehearsal of intuitions which gives out no more than was put in.

2. A confirmation theory should suggest fruitful new problems of inductive inference and new insights into the structure of science, which may even issue in revision of the intuitively accepted methods themselves. For example, problems about criteria for confirming generalizations have led to fruitful discussions, and it will be suggested below that the probabilistic approach to these problems leads to a revised view of the function of generalizations and theories in scientific inference which is of general philosophical importance.

3. A confirmation theory may indicate that biological learning and inference are at least isomorphic with the model of inference provided by the theory, since the existence of the model will at least show that there is a self-consistent way of satisfying certain inductive and epistemological requirements. This suggestion should at present be taken only in a heuristic sense, however. It should be interpreted neither as giving license to confirmation theory to dictate the results of psychological and physiological research nor as issuing in a "discovery machine" for future scientific progress. In its present stage of development, confirmation theory can have only the more modest aim of providing one logically possible model of scientific inference. Even this will leave much to contingencies, such as the basic language presupposed and the presuppositions of learning, which, as has already been argued, cannot be expected to be wholly normative. There are in fact likely to be at least as many contingencies in a confirmation theory as there are in respect of actual human learning; hence no confirmation theory can be expected,

at least at the present stage of research, to describe a more efficient discovery machine than the human brain itself.

The strongest possible arguments for adoption of a Bayesian confirmation theory will not be a priori approaches of the kinds already considered, but the demonstration that such a confirmation theory does satisfy these three conditions. In the past there have been many objections to Bayesian theory on the grounds that it cannot in principle explicate this or that piece of inductive practice: confirmation of universal generalizations, preference for theories of great power or simplicity, and so on. But it follows from the decision theoretic formulation of Bayesian theory already considered that all such objections are beside the point. There will always be some distribution of initial probabilities which will provide the required explication in particular cases; the significant question is whether such a distribution appears at all perspicuous and illuminating for general aspects of theory structure or whether it is entirely implausible and ad hoc. It is only in the latter case that cogent objections can be brought against Bayesian theory, and not enough effort has yet been put into development of the theory to permit summary rejection of it on these grounds.

Objections to Bayesian theory have also been made from the point of view of empirical study of learning, as we saw in connection with restricted Bayesian methods proposed by Suppes. But initially a Bayesian theory may have much more modest aims than are contemplated by these objections. If a Bayesian model of even some aspects of inference can be shown to be self-consistent, this may be sufficient to elucidate important epistemological problems, and whether this proves to be so or not can only be discovered by pushing the theory as far as it will go. Moreover, at present there are no workable alternatives. Study of learning machines and rats has as yet yielded no explicit body of theory such as could be used for a confirmation logic, and in any case the learning problems involved are elementary compared to those of scientific inference. Confirmation theories of nonprobabilistic character have indeed been suggested, but it remains the case that only probabilistic theory has a sufficiently developed logical structure to be easily applicable to even elementary inductive intuitions.

Bayesian theory can lay claim to some immediate successes which have not come to light in other approaches. In this section I shall list some of these, and go on in the next section to discuss the requirements of

confirmation of scientific laws and theories which are more problematic and also more illuminating.

What relation has to hold between data and hypothesis if the data are to count as evidence for, confirmation of, or support for the hypothesis? We have to try to pick out what we regard as the essential features of that relation which may be expressed as, "Assuming the evidence expressed in true propositions, under what conditions do we regard the evidence as increasing the reliability of a hypothesis, or as making the hypothesis more acceptable as a guide to truth?" We have first to decide how to interpret "confirmation" in terms of probability functions. There are two obvious possibilities here. We may regard a hypothesis h as confirmed by evidence e if and only if the probability of h on e attains at least some fixed value k between $1/2$ and 1 . Thus we obtain the k -criterion

(5) e confirms h iff $p(h/e) \geq k > 1/2$.

Roughly speaking, we then believe that h has at least a better than even chance of being true. Such a belief might well dispose us to act upon or accept h , particularly if k is near 1 , or if a choice between h and $\sim h$ is forced upon us. But it may be questioned whether it is a suitable measure of the confirmation of h relative to e , for (5) may be satisfied even if e has decreased the confirmation of h below its initial value, in which case e has relatively disconfirmed h . For the relative measure of confirmation it seems better to adopt the condition, which I call with Carnap the positive relevance criterion,¹⁸ which requires the posterior probability of h on e to be greater than its initial probability:

(6) e confirms h iff $p(h/e) > p(h)$.

I shall also say that e is negatively relevant to h , or disconfirms h , iff $p(h/e) < p(h)$, and that e is irrelevant to h iff $p(h/e) = p(h)$.

There are several useful corollaries of these criteria.

(R.1) e confirms/is irrelevant to/disconfirms h according as h confirms/is irrelevant to/disconfirms e .

For $p(h/e) = p(e/h) p(h)/p(e)$, hence $p(h/e) \geq p(h)$ according as $p(e/h) \geq p(e)$.

¹⁸ R. Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950), ch. 6.

(R.2) If e confirms h , then e disconfirms $\sim h$.

For $p(\sim h/e) = 1 - p(h/e) < 1 - p(h) = p(\sim h)$.

(R.3) If e confirms h , then $\sim e$ disconfirms h .

For then h confirms e , from (R.1); h disconfirms $\sim e$, from (R.2); and hence $\sim e$ disconfirms h from (R.1).

All these properties of positive relevance are intuitively desirable for a relation of confirmation.

Some elementary intuitive conditions of adequacy for inductive inference can also be seen at once to be satisfied by positive relevance, independently of any assignment of initial probabilities. Not all of these conditions are uncontroversial, but in line with the self-correcting explicatory approach adopted here let us first state them and then explore their consequences.

(i) *Equivalence*. A condition that seems fairly unproblematic is that logically equivalent expressions should have identical effects in confirming logically equivalent expressions. That is to say, we have the *equivalence condition*

If $g \equiv g'$, and $h \equiv h'$, then if g confirms h , g' confirms h' .

This is satisfied by the positive relevance criterion of confirmation immediately in virtue of probability axiom (P.2).

The main criticisms of this condition have been in connection with the so-called "raven paradoxes." There are, as we shall see, methods of dealing with these paradoxes other than abandoning the equivalence condition, which would, of course, entail also abandoning a probabilistic theory of confirmation, since the condition follows directly from the probabilistic axioms.

(ii) *Entailment*. Another condition that seems inescapable for the relation of confirmation is that any entailment of a proposition h must be confirmed by h . Thus we have the *entailment condition*

If $h \rightarrow g$, then h confirms g .

This is satisfied immediately for empirical propositions by probability axiom (P.3):

If $h \rightarrow g$, then $p(g/h) = 1$,

hence for any g whose initial probability is less than maximal, $p(g/h) >$

$p(g)$, and h confirms g . It is also satisfied by the k -criterion, for $p(g/h) = 1 > k$.

(iii) *Converse entailment*. This is a more controversial condition. In the deductive model of theories it is generally assumed that if, given some initial conditions, a hypothesis entails an experimental law or singular prediction, and the entailment is accepted as true on the basis of observation, the hypothesis is thereby confirmed. This is indeed the only way in the deductive model in which the reliability of hypotheses can be increased. It is also assumed that a generalization is confirmed by a substitution instance which it entails. Thus we have the *converse entailment condition*

If $h \rightarrow g$, then g confirms h .

We have immediately from Bayes's theorem, if $h \rightarrow g$, where neither $p(h)$ nor $p(g)$ are 0 or 1, then

$$p(h/g) = p(h)/p(g) > p(h).$$

Hence under the stated conditions the converse entailment condition is satisfied in Bayesian theory.

One objection to this condition is that, together with the equivalence condition (i) above, it entails the so-called "raven paradox" of confirmation first stated by Nicod and discussed by Hempel. The paradox will be examined in a later section, and a resolution will be suggested that involves abandoning neither the equivalence nor converse entailment conditions.

It has been further objected to the converse entailment condition that sometimes a single positive substitution instance of a general law is enough to confirm it and that scientists do not usually go on accumulating positive instances of the same kind as those already observed, since such instances would not be regarded as adding significantly to the evidence. What does add to the confirmation, it is held, is observation of positive instances in a variety of circumstances, the elimination of rival hypotheses, or the placing of the law in a logical relation with other laws and theories, which are themselves well confirmed. I shall consider elimination and variety immediately, and later we shall see how far a Bayesian theory can account for the greater confirmation expected from incorporation of a law within a general theory. With these supplementations I shall maintain the condition that any observed entailment provides some

confirmation for a hypothesis whose initial probability is nonzero, however little that confirmation may be under some circumstances.

A useful consequence of converse entailment may be noted here. That is, that it ensures transitivity of confirmation in a deductive system in the following sense. If evidence f confirms a generalization or low-level theory g in virtue of being entailed by g , then f should also confirm a higher-level theory h of which g is an entailment. Since then h entails f , this confirmation follows at once from converse entailment by the transitivity of entailment.

(iv) *Eliminative induction and variety*. Bayesian theory provides an immediate explication and generalization of eliminative induction.

Suppose $S \equiv \{h_1, h_2, \dots, h_N\}$ is a finite, exhaustive, and exclusive hypothesis set. Then since $p(h_1/e) + p(h_2/e) + \dots + p(h_N/e) = 1$, refutation, or more generally disconfirmation, of any proper subset of S by evidence e increases the sum of the probabilities of the remaining subset. Without further information one cannot conclude that the probability of any particular remaining h_r is increased, since the evidence may in fact reduce it. The only conclusive result that can be obtained is in the case of elimination of *all* competing hypotheses, in which case $p(h_r/e)$ becomes 1, and h_r maximally certain. This result is, of course, ideally the aim of crucial experiments, in which the alternatives are supposed to be reduced to two, one of which is refuted by the experiment, although it is rare to find circumstances in which no other alternative hypotheses are possible.

The requirement of variety of instances can also be seen as a consequence of eliminative induction. Consider the generalization "All vegetarians are teetotallers," which would generally be said to be better confirmed by a sample of vegetarians taken from all sections of the population than by a sample taken, for example, exclusively from Methodists. General explication of this intuitive inductive rule turns out to be more difficult than might be expected, no doubt because what seems reasonable in examples of this kind depends heavily upon extra information, such as that many Methodists are teetotallers, and the more complex the background evidence assumed the more unmanageable becomes the explication.¹⁹ We can, however, regard the requirement as a corollary of eliminative induction by supposing that the more varied the popula-

¹⁹ Cf. P. Achinstein, "Variety and Analogy in Confirmation Theory," *Philosophy of Science*, 30 (1963), 207, and Carnap's reply, *ibid.*, p. 223.

tion from which samples are taken, the more rival generalizations about teetotallers are refuted if they are false, and hence the greater is the confirmation obtained for those generalizations that remain unrefuted.

(v) *Initially unexpected evidence.* It is generally assumed that if evidence e_1 is entailed by a hypothesis h , and e_1 would have been initially unexpected without consideration of h , then actual observation of e_1 does more for the confirmation of h than would initially expected evidence e_2 which is also entailed by h . The rule of searching for variety of instances can also be seen as an example of this consideration, for it is initially more unexpected that a general law, such as "All crows are black," should be true for a great number of different kinds of instances of crows (of all ages, for example, and observed in very various geographic regions), than that it should be true only for old crows in the south of England. Again, Einstein's relativity theory entailed that light rays should be detectably bent in the neighborhood of large gravitating masses. This result was highly unexpected in the context of theories and evidence available before relativity theory, and when the bending was actually observed in the neighborhood of the sun in the eclipse of 1919, it provided more support for relativity theory than would have been provided by evidence that was also accounted for, and therefore already expected, on older theories.²⁰

The requirement of unexpected evidence can be rather easily explicated by probabilistic confirmation as follows. Interpret "initially unexpected" to mean "initially improbable," that is, improbable on the evidence available prior to e_1 or e_2 . We can then compare the effect on the posterior probability of h of observing e_1 or e_2 , where e_1 is initially much less probable than e_2 , and $h \rightarrow e_1 \& e_2$. We have

$$\begin{aligned} p(h/e_1) &= p(h)/p(e_1) \\ p(h/e_2) &= p(h)/p(e_2). \end{aligned}$$

But $p(e_1)$ is much less than $p(e_2)$, hence $p(h/e_1)$ is much greater than $p(h/e_2)$. Thus, the confirming effect of e_1 on h is much greater than that

²⁰ A requirement similar to this seems to be what Popper has in mind in speaking of the "severity" of tests. In *Logic of Scientific Discovery* he asserts (in italics) that "the probability of a statement (given some test statements) simply does not express an appraisal of the severity of the tests a theory has passed, or of the manner in which it has passed these tests" (p. 394, cf. p. 267). But in *Conjectures and Refutations*, p. 390, he suggests a measure of severity proportional to the initial improbability of the test, as is also suggested here.

of e_2 , as required, and in a situation of limited experimental resources it would be wise to do the experiment to test c_1 rather than that to test e_2 .

2. *Transitivity of Confirmation*

Some of the easy successes of Bayesian theory have now been outlined. However, it is well known that difficulties soon arise, even in relation to the comparatively elementary requirements of enumerative induction. These difficulties concern, first, the possibility of assigning finite initial probabilities to all the universal generalizations that may be confirmed by evidence, for if there are more than a strictly finite number of these, assignment of finite probabilities, however small, will violate the probability axioms. Hence arises the problem of choosing which finite set out of all possible hypotheses to permit to be confirmed by suitable evidence. Secondly, problems arise about what is to count as confirming evidence for a generalization in order to evade Hempel's paradoxes of confirmation. As we shall see, difficulties about lawlike generalizations are multiplied when we come to consider confirmation of scientific theories and their resulting predictions.

I shall not consider in detail the various attempts that have been made to solve these problems, but I shall rather use the problems, in conjunction with Bayesian confirmation theory, to suggest an interpretation of the structure of scientific theory which conflicts to some extent with the generally received deductivist model of science. In this way I hope to indicate the potentially revisionary function of Bayesian theory, that is to say, the way a systematic attempt to explicate inductive intuitions in science may interact with and modify these intuitions themselves. The intuitions concerned here are those relating to the *universality* of scientific laws and theories. I have argued elsewhere²¹ that it is possible to interpret all of a scientist's behavior with regard to lawlike generalizations without assuming that he has any nonzero degree of belief in their universal truth, and indeed that it is more plausible to suppose that he does not have any such belief. One motivation for this argument is the difficulty in a Carnap-type confirmation theory of assigning finite initial probabilities to enough universal generalizations. I now want to provide another motivation for the argument by considering some more general

²¹ M. Hesse, "Confirmation of Laws," in S. Morgenbesser, P. Suppes, and M. White, eds., *Philosophy, Science, and Method* (New York: St. Martin's, 1969), p. 74.

features of Bayesian confirmation theory, which seem to indicate that the function of general theories in science is not what it is assumed to be in the deductivist model and that theoretical inference, like Mill's eductive inference or instance confirmation in Carnap's theory, can better be regarded as going from case to case than by way of general theories and their entailments.

It is convenient to begin the examination of confirmation relations in scientific theories by a discussion of conditions of adequacy given by Hempel which have been shown to give rise to paradoxes. We have already seen that the equivalence, entailment, and converse entailment conditions are intuitive criteria of adequacy for scientific inference and that they are all satisfied by Bayesian theory under certain conditions of nonzero probabilities. These are three of the conditions discussed by Hempel in his pioneering "Studies in the Logic of Confirmation,"²² in which they are initially presented, along with other suggested conditions, independently of whether confirmation should be explicated by a probabilistic or any other type of axiomatic system. Hempel's paper gave rise to the well-known "paradoxes of the raven," which will be discussed in the next section, but some of the other conditions of adequacy give rise to another paradox, which I shall discriminate as the "transitivity paradox." My terminology and notation will here be a little different from Hempel's.

Consider first the general question of how far the confirmation relation can be expected to be *transitive*, that is, should it satisfy the condition "If f confirms h , and h confirms g , then f confirms g "? In this general form transitivity is obviously not satisfied by the usual notions of confirmation or support. For example, the foundations of a house support the walls which support the roof, and it is true that the foundations support the roof, but although the pivot supports the seesaw, and the seesaw the child at one end, it is not in general true that the pivot supports the child. Again, if a man has disappeared, to find him shot dead confirms the hypothesis that he has been murdered, and that he has been murdered in itself confirms that there is a strangler at large, but that he has been shot does not confirm that there is a strangler.

²² C. G. Hempel, "Studies in the Logic of Confirmation," *Mind*, 54 (1945), 1 and 97, reprinted with a new postscript in *Aspects of Scientific Explanation* (New York: Free Press, 1965), p. 3. See also B. Skyrms, "Nomological Necessity and the Paradoxes of Confirmation," *Philosophy of Science*, 33 (1966), 230.

We have, however, already noted that in the case of the entailment relation $g \rightarrow h$, and $h \rightarrow f$, there is transitivity of confirmation as the converse of transitivity of entailment. There is also another special case in which transitivity seems both plausible and indispensable, in which just one of the confirmation relations of general transitivity is replaced by an entailment. Consider the condition Hempel calls special consequence:

If f confirms h , and $h \rightarrow g$, then f confirms g .

This condition seems to be required for the following sorts of cases which are typical of scientific inference:

(a) Positive instances f of a general law h together with given antecedents i are taken to confirm a prediction of the next instance g as well as confirming the law itself. In this case $h \& i \rightarrow f \& g$. Hence by converse entailment f confirms $h \& i$, and by special consequence f confirms g .

(b) In a more complex theoretical deductive system the condition might be stated in the form "If a hypothesis is well confirmed on given evidence, some further as yet unobserved consequences of the hypothesis ought to be predictable with high confirmation." Consider Newton's theory of gravitation, in which we have the theory confirmed by some of its entailments, for example the laws of planetary orbits and the law of falling bodies. The theory predicts other as yet unobserved phenomena, such as the orbits of the comets or of space satellites. We are not *given* the truth of the theory, only of its original data from planets and falling bodies, and yet we want to say that this evidence confirms predictions about comets and satellites. Special consequence would allow us to transmit confirmation in this way in virtue of the entailment of predictions by the theory. There is also an even stronger intuitive type of theoretical inference. Not only do we require theories to yield strong confirmation to their predictions before the predictions have been tested, but sometimes we use a theory actually to *correct* previously held evidential propositions, if the theory is regarded as sufficiently well confirmed by its other consequences. For example, Galileo's "observation" of constant acceleration of falling bodies and Kepler's third law of planetary motion (about the relation of the periodic times of the planets to the parameters of their orbits) were both corrected by Newton's theory in virtue of its supposed confirmation by other evidence. If the confirmation here is regarded as transmitted by converse entailment and special consequence to the corrected laws, it must have higher degree than was

originally ascribed to the pretheoretical laws on the basis of "observation," since the corrected laws are preferred to the laws previously confirmed by observation.

It seems clear that any adequate account of confirmation must be able to explicate our intuitions and practice in cases such as these. Unfortunately, however, it can easily be shown that the conditions of converse entailment and special consequence just described will not do so, because these conditions together yield further unacceptable consequences. To see this in terms of a specially constructed counterexample, put h equivalent to $f \& g$. Then f confirms h by converse entailment. Also h entails g , hence f confirms g by special consequence. But f and g may be any propositions whatever, and the example has shown that the two conditions together entail that f confirms g . Any relation of confirmation which allows any proposition to confirm any proposition is obviously trivial and unacceptable, for it contradicts the tacit condition that confirmation must be a selective relation among propositions. A paradox has arisen by taking together a set of adequacy conditions, all of which seem to be demanded by intuition; hence it is appropriate to call it the *transitivity paradox*.

That converse entailment and special consequence are not always satisfiable together can be made more perspicuous than in the artificial counterexample just given. If h entails f , and h consists of a conjunction of premises representable by $h_1 \& h_2$, of which only h_1 is necessary to entail f , it is clear that f may confirm only h_1 . If in addition h_2 is that conjunct of h which is necessary to entail g (for example, g may be h_2 itself) there is no reason why f should confirm either h_2 or g . What seems to be presupposed in the examples mentioned as cases of the special consequence condition is a more intimate relation between the parts of the hypothesis which entail f and h respectively than mere conjunction. We have to try to determine what this more intimate relation consists of.

Before doing so it is important to notice how general and powerful the transitivity paradox is. Nothing has been assumed about the relation of "confirmation" or "support" except conditions that seem to arise directly out of commonly accepted types of theoretical argument. As soon as the relation of theory to evidence is construed in deductive fashion the conditions seem natural presuppositions of any relation of confirmation within such a system. In particular, no assumption has yet

been made about the relation of *probability* to confirmation. Probability is the best explored measure of the confirmation relation, and many difficulties have been found in it. But that it cannot incorporate the intuitive conditions exemplified here in simple theoretical inferences cannot be held against probabilistic confirmation any more than against its rivals. For the conditions as they stand are inconsistent with any adequate confirmation theory, and this inconsistency may indeed drive us into deeper examination of the structure of theories than the disputes about probabilistic or nonprobabilistic theories of confirmation. We shall, however, first consider what might be done about the paradox in a Bayesian theory.

Expressed in Bayesian language the paradox is represented as follows. We are interested in the value of $p(e_2/e_1)$, where e_1 is an observed or otherwise given consequence of a theory h , and e_2 is an as yet unobserved or otherwise problematic further consequence of h , that is, an untested prediction made by h . Now $p(e_2/e_1)$ is a single-valued probability function of its arguments alone, and its value cannot depend on whether or not we are interested in some particular h from which e_2 and e_1 are deducible. Consequences e_1 and e_2 are in general deducible from a number of different sets of premises in the language, whether we have thought of them or not. Some of them are perhaps potential theories. But whether or not they turn out to be viable as theories, $p(e_2/e_1)$ has its unique value independent of them.

If this result still appears counterintuitive, we can make it a little more transparent in probabilistic terms. We want to show that $p(e_2/e_1) > p(e_2)$, which is the condition that e_1 and e_2 are not probabilistically independent of each other, but are related by positive relevance. We may be disposed to argue that if $h \rightarrow e_1 \& e_2$, then we should be able to show both that (1) $p(e_2/e_1) \geq p(h/e_1)$ since $h \rightarrow e_2$, and also that (2) $p(h/e_1) > p(e_2)$, since h can be imagined to be increasingly confirmed by a greater and greater amount of evidence e_1 until its posterior confirmation is greater than the initial confirmation of e_2 . Then it would follow immediately that $p(e_2/e_1) > p(e_2)$ as desired. But this intuitive argument is fallacious in general. (1) entails $p(e_1 \& e_2) \geq p(h \& e_1) = p(h)$ since $h \rightarrow e_1$. (2) entails $p(h) > p(e_2)p(e_1)$. Hence both conditions cannot be satisfied together unless initially $p(e_1 \& e_2) > p(e_1)p(e_2)$, which is equivalent to the result which was to be proved, namely, $p(e_2/e_1) > p(e_2)$.

The transitivity paradox was first noticed by Hempel and discussed by Carnap, and has been approached in various ways in the literature.²³ Carnap, and later Hempel, proposed to adopt the positive relevance criterion of confirmation, which entails converse entailment and hence entails abandoning special consequence, but neither Carnap nor Hempel considered the *prima facie* importance of special consequence in theoretical contexts, and therefore they did not discuss the implications of its abandonment. Other suggestions for resolving the paradox have included restricting the standard rule of entailment presupposed in the application of special consequence; replacing the positive relevance criterion of confirmation by the *k*-criterion; and supplementing confirmation by concepts of "acceptance" or "explanation," both of which stand in need of explication at least as badly as does confirmation itself. All these suggestions appear undesirably formal, for the paradox does seem to be a sufficiently fundamental and material consequence of our inductive intuitions to require a type of resolution suggested by concrete cases rather than by formal considerations.

In attempting to resolve the transitivity paradox in an illuminating way, I shall examine three types of inductive inference which can be regarded as special cases of the intuitions leading to the proposed special consequence condition. These are as follows: inferences concerned with *enumerative induction*, namely, general laws, instance confirmation, and De Finetti's concept of exchangeability; theoretical inference from *analogies* and *models*; and the requirement of *simplicity*. I shall argue that these can be interpreted in terms of a unified theory of Bayesian inference which evades the transitivity and raven paradoxes and results in reinterpretation of the function of general laws and theories.

VI. Generalizations, Analogy, and Simplicity

1. Positive Relevance as Clustering, and the Raven Paradox

Cases in which the special consequence condition seem to be required

²³ Carnap, *Logical Foundations*, p. 479; Hempel, *Aspects*, p. 33; B. A. Brody, "Confirmation and Explanation," *Journal of Philosophy*, 65 (1968), 282; H. Smokler, "Conflicting Conceptions of Confirmation," *Journal of Philosophy*, 65 (1968), 300; W. Todd, "Probability and the Theorem of Confirmation," *Mind*, 76 (1967), 260. A related problem of nontransitivity in connection with the concept of "indirect inductive support" is discussed by W. C. Salmon, "Verifiability and Logic," in P. K. Feyerabend and G. Maxwell, eds., *Mind, Matter, and Method* (Minneapolis: University of Minnesota Press, 1966), pp. 361ff.

are cases where it is desired to show that two consequences of a hypothesis are positively relevant to each other, and the transitivity paradox consists in the apparent irrelevance of the hypothesis to this dependence. This immediately suggests that Carnap's notion of *instance confirmation* of general laws²⁴ or Mill's *eduction* from particulars to particulars are paradigm cases of the sort of inference concerned here. Let us therefore first examine the problems surrounding confirmation of general laws.

Take a universal generalization h in the form "All P 's are Q " and consider first only the class of individuals $a_1, a_2, \dots, a_n$ which are P 's, thus evading for the moment the "raven paradox" to be discussed later, in which individuals which are neither P nor Q are considered as evidence for h . Call those individuals in the P -class which are also Q 's *positive instances* of h , and individuals which are P but not Q *negative instances*, and assume for simplicity Q and not Q are initially equally probable. A single negative instance falsifies h and sends its posterior probability to zero. It is usually assumed that a positive instance should confirm h , and an increasing number of positive instances should increasingly confirm h . A clause must of course be added to the effect that there is no other relevant evidence, to avoid confirmation of generalizations like "I wake up every morning at 7 o'clock" by evidence consisting of sixty year's worth of mornings on which I have woken up at 7 o'clock.

Assume the initial conditions $P(a_1), P(a_2), \dots, P(a_r), \dots$ to be given, and put their conjunction equivalent to i . Also put $e_r \equiv Q(a_r)$, and $E_r \equiv e_1 \& e_2 \dots e_r$ ($r = 1, 2, \dots, n$). Then, provided neither $p(h)$ nor $p(E_r)$ is zero, since $h \& i \rightarrow E_r$, and i is given, by converse entailment h is confirmed by E_r . It also follows that if $p(e_{r+1}/E_r) < 1$, all r , then h is increasingly confirmed as r increases. This increasing confirmation does not of course have limit 1 unless h is true, for if h is false a negative instance will eventually be reached and will send the posterior probability of h to zero.

So far the intuitive requirements of enumerative induction are satisfied, but at the cost of assuming $p(h) > 0$. It should be noticed at once, however, that if we are interested in predicting that the next instance of a P will also be a Q , increasing confirmation of h does not necessarily help us. For the argument that since $h \& i$ is confirmed by E_r , and $h \& i \rightarrow e_{r+1}$, therefore e_{r+1} is confirmed by E_r is a fallacious argument of

²⁴ Carnap, *Logical Foundations*, p. 571.

just the kind that has been exposed in connection with the special consequence condition. However, since $h \& i \rightarrow e_{r+1}$ but not conversely, and therefore $p(e_{r+1}/E_r) > p(h \& i/E_r)$, $p(e_{r+1}/E_r)$ may attain a high value even when $p(h) = 0$.

We have e_{r+1} confirmed by E_r given i if and only if

$$(10) \quad p(e_{r+1}/E_r \& i) > p(\sim e_{r+1}/E_r \& i)$$

or alternatively,

$$p(e_{r+1} \& E_r \& i) > p(\sim e_{r+1} \& E_r \& i).$$

Whether this is the case or not will depend only on the initial probabilities of the e_r 's and their conjunctions given i . We have already assumed $p(e_r) = p(\sim e_r)$, all r . Then two further conditions are jointly sufficient for (10). First it is generally assumed in the case of instances of universal generalizations that the tests for each instance being positive or negative are *symmetrical* or *exchangeable*,²⁵ that is, that the probability of obtaining r instances of a specific kind depends only on r and is independent of the order of the individual tests. In particular, the initial probability of a given outcome of a single test of a P -object for Q -ness or for not- Q -ness will respectively be the same for each individual P -object, that is, $p(e_1) = p(e_2) = \dots = p(e_r) = p(\sim e_1)$, and so on. The tests are in this sense random over the individuals.

Secondly, however, to satisfy (10) we still need the initial probability to be such that a conjunction of $r+1$ positive instances is more probable than a conjunction of r of them with one negative instance. This does not follow merely from exchangeability, which tells us only that it was equally likely that the negative instance might have turned up anywhere in the sequence, not that its coming at the end is less probable than a positive instance coming at the end. Exchangeability is concerned only with the order of tests, not with the specific outcomes of those tests. Thus the assumption that the world is more likely to be homogeneous than heterogeneous, in the sense that $r+1$ positive instances are initially more probable than r positive and one negative, must be explicitly built into the initial distribution and cannot be derived from the probability axioms and exchangeability alone. It follows, for example,

²⁵ Cf. Carnap, *Logical Foundations*, p. 484; B. De Finetti, "Foresight: Its Logical Laws, Its Subjective Sources," trans. in H. E. Kyburg and H. E. Smokler, eds., *Studies in Subjective Probability* (New York: John Wiley, 1964), p. 118.

that the counterinductive instruction to predict negative instances after some finite fixed number of positive instances can be represented by Bayesian methods just as well as enumerative inductive methods can. To give finite initial probability to universal generalizations in an infinite universe (which means giving vanishing probability to an infinity of possible statistical generalizations) is one way of giving a preference to homogeneity in the initial distribution. But it is not the only way, for a weaker requirement on the distribution is the strictly finite condition that, of r instances of P 's, it is more probable that none or all will be positive instances of "All P 's are Q 's" than that there will be any other proportion.²⁶ This assumption, it will be noticed, is directly contrary to the result that would be obtained if the P 's were supposed to consist of collection of independent Q 's and not Q 's equally weighted, for then the most probable composition of r P 's would be $r/2$ Q 's and $r/2$ not Q 's. In these matters of induction it must be assumed that God is not a coin-tosser.

I shall call the assumption of homogeneity the *clustering postulate*.²⁷ It has more general application than merely to enumerative induction, as we shall see presently. But first it will be illuminating to consider the raven paradox of confirmation from the point of view just developed. This paradox arises from the equivalence and converse entailment conditions for confirmation. The equivalence condition gives

If $h \equiv h'$, then if e confirms h , e confirms h' .

Consider $h \equiv$ "All ravens are black." This is confirmed by its entailed positive instances, black ravens. The generalization h is also equivalent to $h' \equiv$ "All non-black things are non-ravens," which is confirmed by its positive instances: white shoes, red herrings, blue books, and many other

²⁶ This is essentially the assumption underlying Carnap's c^* distribution. *Logical Foundations*, appendix. See also his discussion of the continuum of inductive methods in terms of their degrees of preference for more or less homogeneous universes. *Continuum*, sec. 22.

²⁷ This postulate is similar to J. M. Keynes's "principle of limited independent variety." *A Treatise on Probability* (London: Macmillan, 1921), ch. 22. The principle may be paraphrased, "If groups of properties are present together in a number of instances, in the absence of other evidence it is more probable than not that they will be present together or absent together in further instances." Attempts to incorporate a clustering postulate of this type into Carnap's theory are made in R. Carnap and W. Stegmüller, *Induktive Logik und Wahrscheinlichkeit* (Vienna: Springer-Verlag, 1959), appendix B, and M. Hesse, "Analogy and Confirmation Theory," *Philosophy of Science*, 31 (1964), 319.

objects. But by the equivalence condition anything which confirms h' confirms h , hence this motley collection of objects confirms "All ravens are black" just as black ravens do. This appears paradoxical, for we should not generally regard observations of red herrings, etc., as support of any kind for "All ravens are black."

Let us, however, along the lines previously suggested, forget the confirmation of universal generalizations, which may in any case be unobtainable, since $p(h)$ may be identically zero. The only finite concern we can have with h in a finite time is to confirm finite sets of predictions of the kind "a is a raven, hence a is black." Suppose we have observed, as we may well have, many black ravens and no nonblack ravens, and, as we certainly have, many nonblack nonravens. How much of this evidence supports the prediction "Raven a is black"? First, it is unlikely that we regard the test of raven a for blackness as *exchangeable* with color tests of herrings, polar bears, mailboxes, and the like, since there is no reason to suppose that the probability of obtaining a given proportion of black to nonblack objects in such a sequence of tests depends only on their number and not on their individual character as ravens, bears, etc. Indeed this is just a way of formulating part of the intuition which led to the paradox in the first place, namely that nonravens are *not* exchangeable for ravens in testing "All ravens are black." But now our assignment of probabilities explicitly depends on such intuitions and is not dictated by probability-equivalences of universal generalizations. Secondly, and more cogently, there is no reason whatever to assume that a motley collection which contains only one raven (the next instance) which is black is initially more probable than a collection of the same size and composition, except that it contains that raven and it is not black. No considerations of homogeneity of the universe dictate this, for homogeneity is more naturally represented by considering the properties of specified similar objects which are obvious candidates for exchangeability and for natural "clusters."

Suppose, however, we are concerned with a finite universe in which it is easy to conceive "All ravens are black" having a finite initial probability. It may then be thought that the paradox reappears, since positive instances now certainly confirm their respective generalizations. But the discussion of the abortive special consequence condition has shown that nothing follows from this about confirmation of *next* instances. Confirmation that the next raven is black cannot be obtained from the motley

collection unless the relevant initial probabilities satisfy exchangeability and clustering conditions as before, and there is nothing in the confirmation conditions or the probability axioms to force such conditions upon us in this case. So long as we restrict consideration to next instances, therefore, the paradox can be evaded by suitable choice of initial probabilities of particulars, whether the universal generalization has finite initial probability or not.²⁸

Whatever may be thought of these considerations as a resolution of the raven paradox, they do at least bring out the contrast between the orthodox view of scientific generalizations and the view advocated here. We are concerned here not with the relation of singular instances to generalizations, or of entailments to hypotheses, but with the relations subsisting among the instances or among the entailments themselves. So far the relations that have emerged as relevant are those of exchangeability and similarity or homogeneity. We shall now see how far these will carry us in considering inferences within theoretical systems.

2. Positive Relevance as Analogy

The examples that have been given of cases where the special consequence condition seems to be required can be seen to be cases where the clustering of instances of a generalization is replaced by the more general notion of an *analogical* relationship between two or more entities, systems, or empirical domains. The theory in these cases can be considered as confirmed by evidence in a given domain and as yielding predictions in another domain antecedently believed to be analogous to the first. Such predictions are, for example, derivation of the orbits of moon satellites from a Newtonian theory tested only in the neighborhood of the earth; prediction of the properties of a new chemical compound from knowledge of its elements and of other compounds; prediction of the properties of a full-sized object from experiments on a replica in, say, a wind tunnel; and prediction of the properties of micro-objects and events from those of macroscopic models.²⁹

²⁸ Methods of this kind will not suffice to resolve Goodman's "grue" paradox, which I believe to be a deeper paradox than those considered here, in the sense that it depends on choice of language basis rather than choice of initial probabilities for a given language. For my proposed resolution of "grue" see "Ramifications of 'Grue,'" *British Journal for the Philosophy of Science*, 20 (1969), 13.

²⁹ I have discussed some of these examples in more detail in "Consilience of Inductions," in I. Lakatos, ed., *The Problem of Inductive Logic* (Amsterdam: North

Predictions of next instances of universal generalizations are elementary special cases of this kind in which the notion of "analogy" between a model whose behavior is known in domain i_1 , and the predictions in domain i_2 , reduces to the notions of exchangeability and clustering of instances in the different initial conditions i_1 , i_2 . This suggests that the clustering postulate, which was invoked to provide an adequate initial probability distribution in the case of next instance confirmation, may be generalized to deal with analogical inference in theories also. Indeed De Finetti himself suggests such a generalization:

Let us consider, to begin with, a class of events (as, for example, the various tosses of a coin). We will say sometimes that they constitute the trials of a given phenomenon; this will serve to remind us that we are almost always interested in applying the reasoning that follows to the case where the events considered are events of the same type, or which have analogous characteristics, without attaching an intrinsic significance or a precise value to these exterior characteristics whose definition is largely arbitrary. Our reasoning will only bring in the events, that is to say, the trials, each taken individually, the analogy of the events does not enter into the chain of reasoning in its own right but only to the degree and in the sense that it can influence in some way the judgment of an individual on the probabilities in question.³⁰

In other words, according to De Finetti, the judgment that the events of a given class are exchangeable is a subjective judgment made prior to and independently of the probability calculations and is capable only of influencing our initial beliefs. One point in De Finetti's remarks needs clarification here: although a relation of analogy between events may suggest that they are exchangeable, exchangeability does not imply that they are analogous, and to obtain next instance predictions in cases of ravens or tossed coins, we need the clustering as well as the exchangeability postulate, as we saw in connection with counterinduction and the raven paradox.

If we consider the kind of theoretical inference of which examples have been given as a species of analogical argument, we must conclude that the judgment of analogy is itself a judgment that the theory in question applies to diverse domains, that the differences between the do-

Holland, 1968), pp. 232, 254, and in "An Inductive Logic of Theories," in M. Radner and S. Winokur, eds., *Minnesota Studies in the Philosophy of Science*, vol. 4 (Minneapolis: University of Minnesota Press, 1970), p. 164.

³⁰ "Foresight: Its Logical Laws, Its Subjective Sources," p. 120.

mains are probabilistically independent of their similarities; and that their similarities are sufficient to justify regarding experiments in any domain of the theory as typical of all its domains. Samples of planets and falling bodies are exchangeable with and analogous to the tests of any other consequences of the gravitational law; macroscopic experiments confirming, say, Maxwell's theory, are analogous to tests of that theory in microphysics; and in general analogical arguments are justified when sufficient analogy is recognized between events.

This approach to the confirmation of theoretical inference has its attractions even though it is certainly incomplete. It will appeal to those who regard the adoption of a theory as a "way of looking at" phenomena rather than as an inference based upon phenomena. It evades the practically insurmountable problem of specifying total evidence which afflicts all global confirmation theories of Carnap's type, for the background evidence is here simply assimilated into the judgment that a set of events is analogous in certain respects and hence probabilistically independent of the rest of the evidence which thereafter need not be explicitly mentioned. From this point of view theories which specify and summarize analogy judgments may be regarded as simplifications of the mass of our total experience which are essential to allow us to operate with it. The significant content of the theory is then seen to be, not so much the assertion of a set of universal generalizations, but rather the assertion that these generalizations describe the essential character of a certain finite set of domains which are held to be analogous in respect of that theory. If such a theory were said to constitute an "explanation," this would not be in the deductive sense, but in the sense of specifying what are the significant causal relations in a given domain and distinguishing them from those that are irrelevant and accidental with respect to that theory. It would be to assimilate all the domains to which the theory is applied to a particular explanatory model which is already found to be pervasive in experience (as for example mechanism in seventeenth-century science), or which is regarded as significant in more limited domains (for example, natural selection or field theories). Recognition of similarities and differences in observable instances would be primitive, prior to theories, and ultimately unanalyzable in terms of theories.

It should be remarked in passing that this view of theories does not commit us to instrumentalism, although its emphasis on observable data and predictions may seem to do so. But the theory has an essential func-

tion in relation to data and predictions, for theory does have a material content, which consists in the assertion that certain observable events, objects, or systems judged analogous are really analogous in respects that permit successful prediction from one system to another. The universality which is usually held to be an essential constituent of theories is seen in this view as rather a convenient method of summarizing case-by-case analogy judgments. But there are no conclusive grounds for associating lack of universality with instrumentalism. Real relations of analogy can subsist between events or objects without the events or objects constituting infinite classes.

On the other hand, this analogical view of explanation seems to preclude the possibility of asking, within the context of confirmation, what justifies adoption of the theory, or, what amounts to the same thing, what justifies us in regarding all features of a domain, other than those specifically mentioned in the theory, as independent of the theory and hence discardable. Surely this judgment is itself based on evidence, and hence is in principle capable of inclusion in the process of confirmation? It is clearly not sufficient to regard it as, in De Finetti's words, "largely arbitrary." We must try to specify in further detail the presuppositions of the independence judgments involved, which will then involve appropriate restrictions upon the choice of initial probability distributions.

The clustering postulate directs us to regard as more probable that state of the world which is more homogeneous. As between f & g and f & $\sim g$, this means that state of the world which exhibits greatest analogy between the systems described by f and by g on the one hand, or by f and by $\sim g$ on the other. This analogy is what would be expressed by a general theory claiming that in certain features systems f and g are identical, that is, they are instances of the same causal laws. It is not, however, the generality of the theory which is in question here, but rather the specific application of it to f and g as analogous instances. But since "analogy" is not yet a very clear notion, it still remains to answer the question: What precisely constitute analogous instances in cases more complex than the recognition that certain observable particulars are sufficiently similar to count as substitution instances of the same generalization? I shall now attempt to answer this question in terms of a more general and universally recognized set of criteria for good theories, namely various kinds of simplicity criteria. These will be shown to yield criteria of positive relevance both in cases of analogical argument from familiar

and observable models and in more general examples where there is no analogical relationship with an observable model.

3. Positive Relevance as Simplicity

There are two major contending concepts for the task of explicating the simplicity of hypotheses, which may be described respectively as the concepts of content and economy.³¹ First, the theory is usually required to have high power or content; to be at once general and specific; and to make precise and detailed claims about the state of the world; that is, in Popper's terminology, to be *highly falsifiable*. This, as Popper maintains against all probabilistic theories of induction, has the consequence that good theories should be in general improbable, since the more claims a theory makes on the world, other things being equal, the less likely it is to be true. On the other hand, as would be insisted by inductivists, a good theory is one that is more likely than its rivals to be true, and in particular it is frequently assumed that simple theories are preferable because they require fewer premises and fewer concepts, and hence would appear to make fewer claims than more complex rivals about the state of the world, and hence to be more probable. Since Popper has annexed the notion of simplicity to the first of these sets of conflicting requirements,³² it has been difficult to do proper justice to the second.

There are, however, two irreducibly different concepts operating here which ought to be considered separately. That this is so can be made obvious in the following example. Suppose the explanandum j is a conjunction of empirical laws of the form

$$j \equiv (x) (PQx \supset Sx \& .QRx \supset Sx)$$

and take k, k' to be theories both entailing j , where

$$\begin{aligned} k &\equiv (x) (Qx \supset Sx), \\ k' &\equiv (x) (Px \supset Sx \& .Rx \supset Sx). \end{aligned}$$

Assuming that expressions of the form $(x) (Qx \supset Sx)$ all have the same initial probability and are logically independent, the probability ordering

³¹ I am indebted to Elliott Sober for useful discussions on the notion of simplicity, although he will not agree with my proposed explication.

³² *Logic of Scientific Discovery*, ch. 7. See, however, the countersuggestion in N. Goodman, "Safety, Strength, Simplicity," *Philosophy of Science*, 28 (1961), 150.

is $p(j) > p(k) > p(k')$, and this is the order of the number of states of the world in which the respective expressions can be true, and hence the inverse order of content or power. Thus if content is the measure of simplicity, simpler theories are less probable, as Popper asserts. But the ordering of intuitive simplicity s here is neither this nor its converse, but rather $s(k) > s(k') > s(j)$. And k would surely be preferred to k' as an explanatory theory for j on grounds of economy of predicates and simplicity of logical form. The greater content of k' does not in this case count in its favor, since the explanation it gives of j depends on two logically unrelated laws, and it gives no insight into what may be the common cause of the two laws conjoined in j .

This example is enough to show that the orderings of content, preferability, and economy are not directly correlated and that content does not exhaust the desirable characteristics of a theory. However, in the situation outlined at the end of the last section, we are not concerned directly with the content of theories, but with comparing two states of the universe $f \& g$ and $f \& \sim g$, where it has been assumed that these two state-descriptions have the same content, that is, they imply conflicting results of the test for g or $\sim g$, each result being assumed to have the same initial probability. We are therefore free in this situation to consider the comparison of $f \& g$ with $f \& \sim g$ only for economy.

Take first the clustering postulate which has already been invoked as the presupposition of analogy arguments. This is also in a sense an economy postulate. Consider evidence consisting of individuals $a_1, a_2, \dots, a_n$, all of which have properties P, Q , and R . Now consider an individual b with properties $\sim P \& Q$. Does b have R or not? If nothing else is known, the clustering postulate will direct us to predict Rb , since, *ceteris paribus*, the universe is to be postulated to be as homogeneous as possible consistently with the data, and a world in which $PQRa$, where a is one of $a_1, a_2, \dots, a_n$, and $\sim PQRb$, is more homogeneous than one in which $PQRa$ and $\sim PQ\sim Rb$. But this is also the prediction that would be made by taking the most economical general law which is both confirmed by the data and of sufficient content to make a prediction about b . For $h \equiv (x) (Qx \supset Rx)$ is certainly more economical than the conflicting

$$h' \equiv (x) (PQx \supset Rx) \& (x) (\sim PQx \supset \sim Rx),$$

or even than a weaker particular hypothesis

$$h'' \equiv (x) (PQx \supset Rx) \ \& \ \sim PQ \sim Rb.$$

Notice that h and h' have the same content, namely all Q 's, whereas h'' has lower content as well as being less economical than h .

Here the clustering postulate gives the same result as adoption of the most economical general law, but the probabilistic justification for choosing the prediction Rb must be given in terms of clustering of individuals rather than economy of the law, since the probability of the law in itself is untransferable to its consequences in virtue of nontransitivity.

Many examples of economy do, however, seem to involve hypotheses directly. A straight line hypothesis is preferred to a higher-order curve when both are consistent with the evidence, and, in general, hypotheses introducing fewer extralogical predicates and fewer arbitrary numerical parameters are preferred to those that introduce more. Popper and H. Jeffreys³³ have both given accounts of the simplicity of curves that depend essentially on counting the number of arbitrary parameters and assigning simplicity inversely as this number. Let us see how this general proposal can be applied to the predictive inference from data points to another specific point satisfying the hypothesis of a straight line rather than a higher-order curve.

Let f be the assertion that two data points (x_1, y_1) , (x_2, y_2) are obtained from an experiment. These points can be written equivalently $(x_1, a + bx_1)$, $(x_2, a + bx_2)$, where a , b are calculated from x_1, y_1, x_2, y_2 . The two points are consistent with the hypothesis $y = a + bx$, and also of course with an indefinite number of other hypotheses of the form $y = a_0 + a_1x + a_2x^2 \dots$ where the values of $a_0, a_1, \dots$ are not determined by x_1, y_1, x_2, y_2 . What is the most economical prediction of the y -value of a further point g , where the x -value of g is x_3 ? Clearly it is that prediction which uses only the information already contained in f , that is, the calculable values of a , b , rather than a prediction which assigns arbitrary values to parameters $a_0, a_1, \dots$ of a higher-order hypothesis. Hence the most economical prediction is $g = (x_3, a + bx_3)$, which is also the prediction given by the "simplest" hypothesis on almost all accounts of the simplicity of curves. Translated into probabilistic language, that is to say that to conform to intuitions about economy we should assign higher initial probability to the assertion that points

³³ H. Jeffreys, *Scientific Inference*, 2nd ed. (Cambridge: At the University Press, 1957), pp. 37, 60; *Theory of Probability* (Oxford: Clarendon, 1961), ch. 1.

$(x_1, a + bx_1)$, $(x_2, a + bx_2)$, $(x_3, a + bx_3)$ are satisfied by the experiment, than to that in which the third point is inexpressible in terms of a and b alone. In this formulation economy is a function of finite descriptive lists of points rather than general hypotheses, and the relevant initial probability is that of a universe containing these points rather than that of a universe in which the corresponding general law is true (which probability may in any case be zero).

A similar interpretation can be given to theories which are held to be simple or economical in virtue of exhibiting invariance between different systems. For example, Einstein was originally convinced of the preferability of relativity theory over the extended classical theory of Lorentz on grounds of this kind of economy. In the period immediately following Einstein's 1905 paper Lorentz's theory was consistent with all existing experimental results, including the Michelson-Morley experiment, and even had greater content than Einstein's theory, in that it contained a micro-theory of electrons which Einstein did not attempt.³⁴ With regard to Lorentz invariance and the consequent radical revision of space-time concepts in Einstein's theory, however, Lorentz's theory involved postulation of an entity, the ether, which Einstein abandoned, and a parameter, the velocity of a system within the ether, which became meaningless in Einstein's theory. Invariance to Lorentzian transformations in space-time was considered by Einstein (and in 1909 by Lorentz himself) as an argument for adoption of Einstein's theory. Such invariance can be interpreted as economy in the description of systems in relatively moving space-time frameworks, since where the two theories conflict, prediction of the result by Lorentz's theory involved the velocity

³⁴ For discussion of "simplicity" comparisons of Einstein's and Lorentz's theories see A. Einstein, "Autobiographical Note," in P. A. Schilpp, ed., *Albert Einstein: Philosopher-Scientist* (Evanston, Ill.: Library of Living Philosophers, 1949), p. 23; G. Holton, "On the Origins of the Special Theory of Relativity," *American Journal of Physics*, 28 (1960), 627; "Mach, Einstein, and the Search for Reality," *Daedalus*, 97 (1968), 636; "Einstein, Michelson, and the 'Crucial' Experiment," *Isis*, 60 (1969), 133; A. Grünbaum, *Philosophical Problems of Space and Time* (New York: Knopf, 1963), ch. 12; K. F. Schaffner, "The Lorentz Electron Theory and Relativity," *American Journal of Physics*, 37 (1969), 498; "Outline of a Logic of Comparative Theory Evaluation with Special Attention to Pre- and Post-Relativistic Electrodynamics," in R. Steuwer, ed., *Minnesota Studies in the Philosophy of Science*, vol. 5 (Minneapolis: University of Minnesota Press, 1970), p. 311. For a useful analysis of simplicity and related concepts of "invariance" and "optimal coding," see H. R. Post, "Simplicity in Scientific Theories," *British Journal for the Philosophy of Science*, 11 (1960), 32, "A Criticism of Popper's Theory of Simplicity," *ibid.*, 12 (1961), 328.

of systems relative to the ether, whereas prediction by Einstein's theory did not.

In probabilistic terms, finite conjunctions of invariant systems should therefore be assigned higher initial probabilities than those not exhibiting invariances. This is the justification in terms of a probabilistic confirmation theory of the preference for general theories entailing invariance as more economical, even though the theories themselves may have to be assigned zero probability if they are expressed in strictly universal form.

4. Conclusion

The claims made here for Bayesian methods in confirmation theory are at once more specific than is usual in the literature and also more modest. They are more modest in that I do not claim these methods are adequate for a global inductive theory for scientists or any other learning organisms. The theory developed here is an explication of some elementary inductive and theoretical inferences applicable when the language basis remains sufficiently stable. In spite of much recent discussion of "meaning variance," I presuppose that language, as well as its meanings, remains locally invariant throughout such pieces of inference, for if it does not it is impossible to understand how communication, let alone rational argument, is possible either in science or in everyday life. I have not attempted to identify the elements of the language or to assign probability weightings to them, since I assume that once a model of inference is given which explicates adequately enough our inductive intuitions, its application to actual cases will suffice to indicate sufficiently well what the language basis in those cases is. To the charge of possible circularity here, I would reply that this is the virtuous circularity of the process of explication itself, in which different parts of the exercise should be mutually adjustable. Confirmation theory should not in any case be so rigorous and inflexible that it becomes open to the opposite charge of dictating the methods of science *a priori*.

I have not suggested stronger constraints on initial distributions of probability than are required for explication of certain familiar types of inference. Consequently nothing like numerical assignments of probability are provided by these constraints, but only comparisons of probability in cases where comparisons of reliability of hypotheses and predic-

tions are normally assumed to be possible. To the objection that such weak constraints on initial probabilities only repeat these intuitive assumptions, it must be replied that the analysis, if correct, has had far-reaching consequences for the deductive model of science. Discussion of generalizations and theories has led to the proposal that confirmation in science should be seen primarily as an application of inference from case to case rather than as confirmation of universal laws and theories. I suggest, however, that far from being an inadequacy in the concept of confirmation, this consequence is highly illuminating for the general problem of theoretical inference in science. In particular it provides resolutions of some so-called "paradoxes of confirmation," and it also enables a unified inductive account to be given of enumerative induction, analogical inference, and the criterion of simplicity of theories. This inductive account differs in important respects from the usual deductive model of science, but its very systematic character provides an argument in its favor, for the deductive model has no such systematic account of the process of confirmation and related desiderata for theories such as analogy and simplicity. A modest Bayesian confirmation theory therefore has revisionary as well as descriptive implications for our understanding of the structure of scientific theories.