8

RECENT PROBLEMS OF INDUCTION

Carl G. Hempel

> It is true that from truths we can conclude only truths; but there are certain falsehoods which are useful for finding the truth.
>
> —Leibniz, Letter to Canon Foucher (1692)

THE CLASSICAL PROBLEM OF INDUCTION

In the philosophical discussion of induction, one problem has long occupied the center of the stage—so much so, indeed, that it is usually referred to as *the* problem of induction. That is the problem of justifying the way in which, in scientific inquiry and in our everyday pursuits, we base beliefs and assertions about empirical matters on logically inconclusive evidence.

This classical problem of justification, raised by Hume and made famous by his skeptical solution, is indeed of great philosophical importance. But more recent studies, most of which were carried out during the past two or three decades, have given rise to new problems of induction, no less perplexing and important than the classical one, which are logically prior to it in the sense that the classical problem cannot even be clearly stated—let alone solved—without some prior clarification of the new puzzles.

In this paper, I propose to discuss some of these recent problems of induction.

Induction may be regarded as effecting a transition from some body of empirical information to a hypothesis which is not logically implied by it, and for this reason it is often referred to as nondemonstrative *inference*. This characterization has to be taken with a grain of salt; but it is suggestive and convenient, and in accordance with it, I will therefore sometimes

From Robert G. Colodny (ed.), *Mind and Cosmos, Essays in Contemporary Science and Philosophy* (University of Pittsburgh Press, 1966), pp. 112–134. Reprinted by permission of the University of Pittsburgh Press.

refer to the sentences specifying the evidence as the *premises* and to the hypothesis based on it as the conclusion of an "*inductive inference*."

Among the simplest types of inductive reasoning are those in which the evidence consists of a set of examined instances of a generalization, and the hypothesis is either the generalization itself or a statement about some unexamined instances of it. A standard example is the inference from the evidence statement that all ravens so far observed have been black to the generalization that all ravens are black or to the prediction that the birds now hatching in a given clutch of raven eggs will be black or to the retrodiction that a raven whose skeleton was found at an archeological site was black. As these examples show, induction does not always proceed from the particular to the general or from statements about the past or present to statements about the future.

The inductive procedures of science comprise many other, more complex and circumstantial, kinds of nondemonstrative reasoning, such as those used in making a medical diagnosis on the basis of observed symptoms, in basing statements about remote historical events on presently available evidence, or in establishing a theory on the basis of appropriate experimental data.

However, most of the problems to be considered here can be illustrated by inductions of the simple kind that proceed from instances of a generalization, and in general I will use these as examples.

THE NARROW INDUCTIVIST VIEW OF SCIENTIFIC INQUIRY

It should be stressed at the outset that what we have called inductive inference must not be thought of as an effective method of discovery, which by a mechanical procedure leads from observational data to appropriate hypotheses or theories. This misconception underlies what might be called the narrow inductivist view of scientific inquiry, a view that is well illustrated by the following pronouncement:

> If we try to imagine how a mind of superhuman power and reach, but normal so far as the logical processes of its thought are concerned . . . would use the scientific method, the process would be as follows: First, all facts would be observed and recorded, *without selection* or *a priori* guess as to their relative importance. Second, the observed and recorded facts would be analyzed, compared, and classified, *without hypothesis or postulates* other than those necessarily involved in the logic of thought. Third, from this analysis of the facts, generalization would be inductively drawn as to the relations, classificatory or causal, between them. Fourth, further research would be deductive as well as inductive, employing inferences from previously established generalizations.[1]

[1] A. B. Wolfe, "Functional Economics," *The Trend of Economics*, ed. R. G. Tugwell (New York: Knopf, 1924), p. 450 (author's italics).

It need hardly be argued in detail that this conception of scientific procedure, and of the role induction plays in it, is untenable; the reasons have been set forth by many writers. Let us just note that an inquiry conforming to this idea would never go beyond the first stage, for—presumably to safeguard scientific objectivity—no initial hypotheses about the mutual relevance and interconnections of facts are to be entertained in this stage, and as a result, there would be no criteria for the selection of the facts to be recorded. The initial stage would therefore degenerate into an indiscriminate and interminable gathering of data from an unlimited range of observable facts, and the inquiry would be totally without aim or direction.

Similar difficulties would beset the second stage—if it could ever be reached—for the classification or comparison of data again requires criteria. These are normally suggested by hypotheses about the empirical connections between various features of the "facts" under study. But the conception just cited would prohibit the use of such hypotheses, and the second stage of inquiry as here envisaged would again lack aim and direction.

It might seem that the quoted account of inductive scientific procedure could be rectified by simply adding the observation that any particular scientific investigation is aimed at solving a specified problem, and that the initial selection of data should therefore be limited to facts that are relevant to that problem. But this will not do, for the statement of a problem does not generally determine what kinds of data are relevant to its solution. The question as to the causes of lung cancer does not by itself determine what sorts of data would be relevant—whether, for example, differences in age, occupation, sex, or dietary habits should be recorded and studied. The notion of "relevant" facts acquires a clear meaning only when some specific answer to the problem has been suggested, however tentatively, in the form of a hypothesis: an observed fact will then be favorably or unfavorably relevant to the hypothesis according as its occurrence is by implication affirmed or denied by the hypothesis. Thus, the conjecture that smoking is a potent causative factor in lung cancer affirms by implication a higher incidence of the disease among smokers than among nonsmokers. Data showing for a suitable group of subjects that this is the case or that it is not would therefore constitute favorably relevant (confirming) or unfavorably relevant (disconfirming) evidence for the hypothesis. Generally, then, those data are relevant and need to be gathered which can support or disconfirm the contemplated hypothesis and which thus provide a basis for testing it.

Contrary to the conception quoted above, therefore, hypotheses are put forward in science as tentative answers to the problem under investigation. And contrary to what is suggested by the description of the third stage of inquiry above, such answers in the form of hypotheses or theories cannot be inferred from empirical evidence by means of some set of mechanically applicable rules of induction. There is no generally applicable mechanical routine of "inductive inference" which leads from a given set of data to a corresponding hypothesis or theory somewhat in the way in which the familiar routine of multiplication leads from any two given integers, by a finite number of mechanically performable steps, to the corresponding product.

To be sure, mechanical induction routines can be specified for certain special kinds of cases, such as the construction of a curve, and of an analytic expression for the corresponding function, which will fit a finite set of points. Given a finite set of measurements of associated values of temperature and volume for a given body of gas under constant pressure, this kind of procedure could serve mechanically to produce a tentative general law connecting temperature and volume of the gas. But for generating scientific theories, no such procedure can be devised.

Consider, for example, a theory, such as the theory of gravitation or the atomic theory of matter, which is introduced to account for certain previously established empirical facts, such as regularities of planetary motion and free fall, or certain chemical findings such as those expressed by the laws of constant and of multiple proportions. Such a theory is formulated in terms of certain concepts (those of gravitational force, of atom, of molecule, etc.) which are novel in the sense that they had played no role in the description of the empirical facts which the theory is designed to explain. And surely, no set of induction rules could be devised which would be generally applicable to just any set of empirical data (physical, chemical, biological, etc.) and which, in a sequence of mechanically performable steps, would generate appropriate novel concepts, functioning in an explanatory theory, on the basis of a description of the data.[2]

Scientific hypotheses and theories, then, are not mechanically inferred from observed "facts": *They are invented by an exercise of creative imagination.* Einstein, among others, often emphasized this point, and more than a century ago William Whewell presented the same basic view of induction. Whewell speaks of scientific discovery as a "process of invention, trial, and acceptance or rejection" of hypotheses and refers to great scientific advances as achieved by "Happy *Guesses*," by "felicitous and inexplicable strokes of inventive talent," and he adds: "No rules can ensure to us similar success in new cases; or can enable men who do not possess similar endowments, to make like advances in knowledge."[3] Similarly, Karl Popper has characterized scientific hypotheses and theories as conjectures, which must then

[2] This argument does not presuppose a fixed division of the vocabulary of empirical science into observational and theoretical terms; it is quite compatible with acknowledging that as a theory becomes increasingly well established and accepted, certain statements couched in terms of its characteristic concepts may come to be qualified as descriptions of "observed facts."

[3] William Whewell, *The Philosophy of the Inductive Sciences*, 2d ed. (London: John W. Parker, 1847), II, 41 (author's italics).

be subjected to test and possible falsification.[4] Such conjectures are often arrived at by anything but explicit and systematic reasoning. The chemist Kékulé, for example, reports that his ring formula for the benzene molecule occurred to him in a reverie into which he had fallen before his fireplace. Gazing into the flames, he seemed to see snakes dancing about; and suddenly one of them moved into the foreground and formed a ring by seizing hold of its own tail. Kékulé does not tell us whether the snake was forming a *hexagonal* ring, but that was the structure he promptly ascribed to the benzene molecule.

Although no restrictions are imposed upon the *invention* of theories, scientific objectivity is safeguarded by making their *acceptance* dependent upon the outcome of careful tests. These consist in deriving, from the theory, consequences that admit of observational or experimental investigation, and then checking them by suitable observations or experiments. If careful testing bears out the consequences, the hypothesis is accordingly supported. But normally a scientific hypothesis asserts more than (i.e., cannot be inferred from) some finite set of consequences that may have been put to test, so that even strong evidential support affords no conclusive proof. It is precisely this fact, of course, that makes inductive "inference" nondemonstrative and gives rise to the classical problem of induction.

Karl Popper, in his analysis of this problem, stresses that the inferences involved in testing a scientific theory always run deductively from the theory to implications about empirical facts, never in the opposite direction; and he argues that therefore "Induction, i.e., inference based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor one of scientific procedure";[5] and it is essentially this observation which, he holds, "solves . . . Hume's problem of induction."[6] But this is surely too strong a claim, for although the procedure of empirical science is not inductive in the narrow sense we have discussed and rejected, it still may be said to be *inductive in a wider sense,* referred to at the beginning of this paper: While scientific hypotheses and theories are not *inferred* from empirical data by means of some effective inductive procedure, they are *accepted* on the basis of observational or experimental findings which afford no deductively conclusive evidence for their truth. Thus, the classical problem of induction retains its import: What justification is there for accepting hypotheses on the basis of incomplete evidence?

[4] See, for example, Popper's essay, "Science: Conjectures and Refutations," in his book, *Conjectures and Refutations* (New York and London: Basic Books, 1962).

[5] Karl Popper, "Philosophy of Science: A Personal Report," *British Philosophy in the Mid-Century*, ed. C. A. Mace (London: Allen and Unwin, 1957), pp. 155–91, quotation from p. 181.

[6] Popper, "Philosophy of Science," p. 183.

The search for an answer to this question will require a clearer specification of the procedure that is to be justified; for while the hypotheses and theories of empirical science are not deductively implied by the evidence, it evidently will not count as inductively sound reasoning to accept a hypothesis on the basis of just any inconclusive evidence. Thus, there arises the logically prior problem of giving a more explicit characterization and precise criteria of what counts as sound inductive reasoning in science.

It may be instructive briefly to consider the analogue to this problem for deductive reasoning.

DEDUCTION AND INDUCTION; DISCOVERY AND VALIDATION

Deductive soundness, of course, is tantamount to deductive validity. This notion can be suggestively although imprecisely characterized by saying that an argument is deductively valid if its premises and its conclusion are so related that if all the premises are true, then the conclusion cannot fail to be true as well.[7]

As for *criteria* of deductive validity, the theory of deductive logic specifies a variety of forms of inference which are deductively valid, such as, for example, *modus ponens:*

$$\begin{array}{l} p \supset q \\ p \\ \hline \quad q \end{array}$$

or the inference rules of quantificational logic. Each of these represents a sufficient but not necessary condition of deductive validity. These criteria have the important characteristic of being expressible by reference to the syntactical structure of the argument, and thus without any reference to the meanings of the extralogical terms occurring in premises and conclusion. As we will see later, criteria of inductive soundness cannot be stated in purely syntactical terms.

We have already noted that whatever the rules of induction may be, they cannot be expected to specify mechanical routines leading from empirical evidence to appropriate hypotheses. Are the rules of deductive inference superior in this respect? Consider their role in logic and mathematics.

A moment's reflection shows that no interesting theorem in these fields is discovered by a mechanical application of the rules of deductive inference. Unless a putative theorem has first been put forward, such application

[7] Precise general characterizations of deductive validity, for arguments in languages of certain specified forms, will be found, e.g., in W. V. O. Quine, *Methods of Logic,* rev. ed. (New York: Holt, Rinehart & Winston, 1959).

would lack direction. Discovery in logic and mathematics, no less than in empirical science, *calls for imagination and invention;* it does not follow any mechanical rules.

Next, even when a putative theorem has been proposed, the rules of deduction do not, in general, provide a mechanical routine for proving or disproving it. This is illustrated by the famous arithmetical conjectures of Goldbach and of Fermat, which were proposed centuries ago but have remained undecided to this day. Mechanical routines for proving or disproving any given conjecture can be specified only for systems that admit of a decision procedure; and even for first-order quantificational logic and for elementary arithmetic, it is known that there can be no such procedure. In general, then, the construction of a proof or a disproof for a given logical or mathematical conjecture requires ingenuity.

But when a putative theorem has been proposed and a step-by-step argument has been offered as a presumptive proof for it, then the rules of deductive logic afford a means of establishing the validity of the argument: If each step conforms to one of those rules—a matter which can be decided by mechanical check—then the argument is a valid proof of the proposed theorem.

In sum, the formal rules of deductive inference are not rules of discovery leading mechanically to correct theorems or even to proofs for conjectured theorems which are in fact provable; rather, they provide criteria of soundness or of validity for proposed deductive proofs.

Analogously, rules of inductive inference will have to be conceived, not as canons of discovery, but as criteria of validation for proposed inductive arguments; far from generating a hypothesis from given evidence, they will *presuppose* that, in addition to a body of evidence, a hypothesis has been put forward, and they will then serve to appraise the soundness of the hypothesis on the basis of the evidence.

Broadly speaking, inductive arguments might be thought of as taking one of these forms:

$$\begin{array}{c} e \\ \hline\hline h \end{array} \quad \text{(i.e., evidence } e \text{ supports hypothesis } h\text{)}$$

$$\begin{array}{c} e \\ \hline\hline h \end{array} [r] \quad \text{(i.e., evidence } e \text{ supports hypothesis } h \text{ to degree } r\text{)}$$

Here, the double line is to indicate that the relation of e to h is not that of full deductive implication but that of partial inductive support.

The second of these schemata incorporates the construal of inductive support as a quantitative concept. Rules of induction pertaining to it would provide criteria determining the degree of support conferred on certain kinds of hypotheses by certain kinds of evidence sentences; these criteria might even amount to a general definition assigning a definite value of r to any given e and h; this is one objective of Carnap's inductive logic.[8]

The first schema treats inductive support or confirmation as a qualitative concept; the corresponding inference rules would specify conditions under which a given evidence sentence supports, or confirms, a given hypothesis.[9]

The formulation of rules of these or similar kinds will be required to explicate the concept of inductive inference in terms of which the classical problem of justification is formulated. And it is in this context of explication that the newer problems of induction arise. We now turn to one of those problems; it concerns the qualitative concept of confirmation.

THE PARADOXES OF QUALITATIVE CONFIRMATION

The most familiar rules of induction concern generalizations of the simple form "All F are G." According to one widely asserted rule, a hypothesis of this kind receives support from its positive instances—i.e., from cases of F that have been found also to be G. For example, the hypothesis "All ravens are black," or

$$(x)\ (Rx \supset Bx) \qquad (h)$$

is supported, or confirmed, by any object i such that

$$Ri \cdot Bi \qquad (I)$$

or, as we will say, by any evidence sentence of the form "$Ri \cdot Bi$." Let us refer to such instances as *positive instances of type I for h*. Similarly, h is disconfirmed (invalidated) by any evidence sentence of the form $Ri \cdot -Bi$. This criterion was explicitly discussed and advocated by Jean Nicod;[10] I will therefore call it Nicod's criterion.

[8] See especially the following publications by Rudolf Carnap: *Logical Foundations of Probability*, 2d ed. (Chicago: University of Chicago Press, 1962); "The Aim of Inductive Logic," *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, eds. E. Nagel, P. Suppes, and A. Tarski (Stanford: Stanford U. Press, 1962), pp. 303–18.

[9] It seems to me, therefore, that Popper begs the question when he declares: "But it is obvious that this rule or craft of 'valid inducton'... simply does not exist. No rule can ever guarantee that a generalization inferred from true observations, however often repeated, is true" ("Philosophy of Science," p. 181). That inductive reasoning is not *deductively* valid is granted at the outset; the problem is that of constructing a concept of *inductive* validity.

[10] Jean Nicod, *Foundations of Geometry and Induction* (New York: Harcourt, Brace & World, 1930), p. 219. Nicod here speaks of "truths or facts," namely, "the presence or absence of B in a case of A," as confirming or invalidating "the law *A entails B*" (author's italics). Such confirmatory and disconfirmatory facts can be thought of as described by corresponding evidence sentences. Nicod remarks about his criterion: "We have not seen it stated in any explicit manner. However, we do not think that anything ever written on induction is incompatible with it" (p. 220). Whether Nicod regards the specified conditions as necessary and sufficient or merely as sufficient for confirmation or invalidation is not entirely clear, although he does say: "It is conceivable that we have here the only two direct modes in which a fact can influence the probability of a law" (p. 219). We will construe his criteria simply as *sufficient* conditions of confirmation and invalidation.

Now, the hypothesis h is logically equivalent to, and thus makes exactly the same assertion as, the statement that all nonblack things are nonravens, or

$$(x)\ (-Bx \supset -Rx) \qquad (h')$$

According to Nicod's criterion, this generalization is confirmed by *its* instances—i.e., by any individual j such that

$$-Bj \cdot -Rj \qquad (II)$$

But since h' expresses exactly the same assertion as h, any such individual will also confirm h. Consequently, such things as a yellow rose, a green caterpillar, or a red herring confirm the generalization "All ravens are black," by virtue of being nonblack nonravens. I will call such objects *positive instances of type II for h*.

Next, the hypothesis h is logically equivalent also to the following statement:

$$(x)\ [(Rx \vee -Rx) \supset (-Rx \vee Bx)] \qquad (h'')$$

in words: Anything that is a raven or not a raven—i.e., anything at all—either is not a raven or is black. Confirmatory instances for this version which I will call *positive instances of type III for h*, consist of individuals k such that

$$-Rk \vee Bk \qquad (III)$$

This condition is met by any object k that is not a raven (no matter whether it is black) and by any object k that is black (no matter whether it is a raven). Any such object, then, affords a confirmatory instance in support of the hypothesis that all ravens are black.

On the other hand, the hypothesis h can be equivalently expressed by the sentence

$$(x)\ [(Rx \cdot -Bx) \supset (Rx \cdot -Rx)] \qquad (h''')$$

for which nothing can possibly be a confirmatory instance in the sense of Nicod's criterion, since nothing can be both a raven and not a raven.

These peculiarities, and some related ones, of the notion of confirmatory instance of a generalization have come to be referred to as the *paradoxes of* *confirmation.*[11] And indeed, at first glance they appear to be implausible and perhaps even logically unsound. But on further reflection one has to conclude, I think, that they are perfectly sound, that it is our intuition in the matter which leads us astray, so that the startling results are paradoxical only in a psychological, but not in a logical sense.

To see this, let us note first that the results in question follow deductively from two simple basic principles, namely: (*A*) A generalization of the form "All *F* are *G*" is confirmed by its positive instances—i.e., by cases of *F* that have been found also to be cases of *G.* (*B*) Whatever confirms a hypothesis also confirms any logically equivalent one.

Principle (*A*) is, in effect, part of Nicod's criterion, of which Nicod himself remarks that it "cannot claim the force of an axiom. But it offers itself so naturally and introduces such great simplicity, that reason welcomes it without feeling any imposition."[12] We will encounter some surprising exceptions to it in Sections 5 and 6, but it does indeed seem very reasonable in cases of the kind we have considered so far—i.e., in reference to generalizations of universal conditional form containing exclusively property terms (one-place predicates).

Principle (*B*) may be called the equivalence condition. It simply reflects the idea that whether given evidence confirms a hypothesis must depend only on the content of the hypothesis and not on the way in which it happens to be formulated.

And once we accept these principles, we must also accept their surprising logical consequences.

Let us look at these consequences now from a different point of view, which will support the claim that they are sound. Suppose we are told that in the next room there is an object i which is a raven. Our hypothesis h then tells us about i that it is black, and if we find that this is indeed the case, so that we have $Ri \cdot Bi$, then this must surely count as bearing out, or confirming, the hypothesis.

Next, suppose we are told that in the adjoining room there is an object j that is not black. Again, our hypothesis tells us something more about it, namely, that it is not a raven. And if we find that this is indeed so—i.e., that $-Bj \cdot -Rj$, then this bears out, and thus supports, the hypothesis.

Finally, even if we are told only that in the next room there is an object k, the hypothesis still tells us something about it, namely, that either it is no raven or it is black—i.e., that $-Rk \vee Bk$; and if this is found to be the case, it again bears out the hypothesis.

[11] These paradoxes were first noted in my essay "Le problème de la vérité," *Theoria* (Göteborg), 3 (1937), 206–46 (see especially p. 222) and were discussed in greater detail in my articles "Studies in the Logic of Confirmation," *Mind,* 54 (1945), 1–26, 97–121, and "A Purely Syntactical Definition of Confirmation," *The J. of Symbolic Logic,* 8 (1943), 122–43.

[12] Nicod, *Geometry and Induction,* pp. 219–20.

Thus, our three types of positive instance must indeed be counted as confirmatory or supporting evidence for the generalization that all ravens are black.

Finally, the fact that the formulation h''' of our generalization admits of no confirming instances in the sense of Nicod's criterion presents no serious problem if, as here has been done, that criterion is stated as a sufficient but not necessary condition of confirmation.

But why does it seem implausible or paradoxical in the first place that positive instances of types *II* and *III* should be confirmatory for the generalization h? One important reason seems to lie in the assumption that the hypothesis "All ravens are black" is a statement about ravens and not about nonravens, let alone about all things in general. But surely, such a construal is untenable; anyone who accepts h would be bound to accept also the sentences h' and h'', which by the same token would have to be viewed as statements about nonravens and about all things, respectively. The use made of some statements of the form "All *F* are *G*" illustrates the same point. The Wassermann test, for example, is based, roughly speaking, on the generalization that any person infected with syphilis has a positive Wassermann reaction; but in view of its diagnostic implications for cases yielding a negative test result, this generalization surely cannot be said to be about syphilitically infected persons only.

To say that positive instances of types *I, II,* and *III* all confirm the hypothesis h is not to say, however, that they confirm the generalization to the same extent. Indeed, several writers have argued that the different types differ greatly in this respect and that, in particular, a positive instance of type *I,* i.e., a black raven, lends much stronger support to our generalization than a positive instance of type *II,* i.e., a nonblack object that is not a raven; and they have suggested that this is the objective basis for the first impression that instances of type *I* alone can count as confirmatory for our hypothesis.

This view can be made plausible by the following suggestive but imprecise consideration: Let k be the hypothesis "All marbles in this bag are red," and suppose that there are twenty marbles in the bag. Then the generalization k has twenty instances of type *I,* each being provided by one of the marbles. If we had checked each of the twenty objects that are marbles in the bag, we have exhaustively tested the hypothesis. And roughly speaking we might say that if we have examined one of the marbles and found it red, we have shown one twentieth of the total content of the hypothesis to be true.

Now consider the contrapositive of our generalization—i.e., the statement, "Any object that is not red is not a marble in this bag." Its instances are provided by all nonred objects. There are a large number of these in the world—perhaps infinitely many of them. Examining one of them and averring that it is not a marble in the bag is therefore to check, and corroborate, only a tiny portion of all that the hypothesis affirms. Hence, a positive finding of type *II* would indeed support our generalization, but only to a very small extent.

Analogously in the case of the ravens. If we may assume that there are vastly more nonblack things than there are ravens, then the observation of one nonblack thing that is not a raven would seem to lend vastly less support to the generalization that all ravens are black than would the observation of one raven that *is* black.

This argument might serve to mitigate the paradoxes of confirmation.[13] But I have stated it here only in an intuitive fashion. A precise formulation would require an explicit quantitative theory of degrees of confirmation or of inductive probability, such as Carnap's. Even within the framework of such a theory, the argument presupposes further assumptions, and the extent to which it can be sustained is not fully clear as yet.

Let us now turn to another perplexing aspect of induction. I will call it Goodman's riddle, because it was Nelson Goodman who first called attention to this problem and proposed a solution for it.[14]

GOODMAN'S RIDDLE: A FAILURE OF CONFIRMATION BY "POSITIVE INSTANCES"

One of the two basic principles from which we deduced the paradoxes of confirmation stated that a generalization of the form "All *F* are *G*" is confirmed, or supported, by its positive instances of type *I*—i.e., by objects which are *F* and also *G*. Although this principle seems entirely obvious, Goodman has shown that there are generalizations that derive no support at all from their observed instances. Take for example the hypothesis

All ravens are blite (*h*)

where an object is said to be blite if it is either examined before midnight tonight and is black or is not examined before midnight and is white.

Suppose now that all the ravens examined so far have been found to be black; then, by definition, all ravens so far examined are also blite. Yet this latter information does not support the generalization *h*, for that generalization implies that all ravens examined after midnight will be white—and

[13] It was first offered by Janina Hosiasson-Lindenbaum in her article "On Confirmation," *The J. of Symbolic Logic,* 5 (1940), 133–48. Similar ideas were proposed by, among others, D. Pears, "Hypotheticals," *Analysis,* 10 (1950), 49–63; I. J. Good, "The Paradoxes of Confirmation," Pts. I and II, *The British J. for the Philosophy of Science,* 11 (1960), 145–48; 12 (1961) 63–64. A detailed and illuminating study of qualitative confirmation and its paradoxes is offered in sec. 3, Pt. I of Israel Scheffler, *The Anatomy of Inquiry* (New York: Knopf, 1963).

[14] Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge: Harvard U. Press, 1955); 2d, rev. ed. (Indianapolis: Bobbs-Merrill, 1965).

surely our evidence must be held to militate against this forecast rather than to support it.

Thus, some generalizations do derive support from their positive instances of type *I;* for example, "All ravens are black," "All gases expand when heated," "In all cases of free fall from rest, the distance covered is proportional to the square of the elapsed time," and so forth; but other generalizations, of which "All ravens are blite" is an example, are not supported by their instances. Goodman expresses this idea by saying that the former generalizations can, whereas the latter cannot, be *projected* from examined instances to as yet unexamined ones.

The question then arises how to distinguish between projectible and nonprojectible generalizations. Goodman notes that the two differ in the character of the terms employed in their formulation. The term "black," for example, lends itself to projection; the term "blite" does not. He traces the difference between these two kinds of term to what he calls their *entrenchment*—i.e., the extent to which they have been used in previously projected hypotheses. The word "blite," for example, has never before been used in a projection, and is thus much less entrenched than such words as "black," "raven," "gas," "temperature," "velocity," and so on, all of which have served in many previous inductive projections—successful as well as unsuccessful ones. What Goodman thus suggests is that our generalizations are chosen not only in consideration of how well they accord with the available evidence, but also in consideration of how well entrenched are their constituent extralogical terms.

By reference to the relative entrenchment of those terms, Goodman then formulates criteria for the comparison of generalizations in regard to their projectibility, and he thus constructs the beginnings of a theory of inductive projection.

I cannot enter into the details of Goodman's theory here, but I do wish to point out one of its implications which is, I think, of great importance for the conception of inductive inference.

As we noted earlier, the standard rules of deductive inference make reference only to the syntactical form of the sentences involved; the inference rules of quantification theory, for example, apply to all premises and conclusions of the requisite form, no matter whether the extralogical predicates they contain are familiar or strange, well entrenched or poorly entrenched. Thus,

All ravens are blite

and

r is a raven

deductively implies

r is blite

no less than

All ravens are black

and

r is a raven

deductively implies

r is black

But on Goodman's conception of projectibility, even elementary rules of induction cannot be similarly stated in purely syntactical terms. For example, the rule that a positive instance confirms a generalization holds only for generalizations with adequately entrenched predicates; and entrenchment is neither a syntactical nor even a semantic property of terms, but a pragmatic one; it pertains to the actual use that has been made of a term in generalizations projected in the past.

A FURTHER FAILURE OF CONFIRMATION BY "POSITIVE INSTANCES"

Goodman's riddle shows that Nicod's criterion does not offer a generally adequate sufficient condition of confirmation: Positive instances do not confirm nonprojectible hypotheses.

But the criterion fails also in cases of a quite different kind, which do not hinge on the use of predicates such as "blite." Consider the hypothesis, "If for any two persons x,y it is not the case that each likes the other, then the first likes the second, but not vice versa"; in symbolic notation:

$$(x)\,(y)[-(Lxy \cdot Lyx) \supset (Lxy \cdot -Lyx)] \qquad (h)$$

Let e be the information that a,b are two persons such that a likes b but not vice versa, i.e. that

$$Lab \cdot -Lba \qquad (e)$$

This information can equivalently be stated as follows:

$$-(Lab \cdot Lba) \text{ and } (Lab \cdot -Lba) \qquad (e')$$

for the first of these two sentences is a logical consequence of the second one. The sentence e' then represents a positive instance of type I for h; hence, on Nicod's criterion, e' should confirm h.[15]

[15] Nicod does not explicitly deal with hypotheses which, like h, contain relational terms rather than only property terms such as "raven" and "black"; but the application here suggested certainly seems to be in full accord with his basic conception.

But e' is equivalent to

$$-(Lba \cdot Lab) \text{ and } (-Lba \cdot Lab) \qquad (e'')$$

and this, on Nicod's criterion, disconfirms h. In intuitive terms, the preceding argument is to this effect: If a is counted as the first person and b as the second, then the information provided by e shows that, as e' makes explicit, a and b satisfy both the antecedent and the consequent of h and thus confirm the hypothesis; but if b is counted as the first person and a as the second one, then by virtue of the same information, b and a satisfy the antecedent but not the consequent of h, as is made explicit in e''. Thus, on Nicod's criterion, e constitutes both confirming and invalidating evidence for h.

Incidentally, h can be thrown into the form

$$(x)\,(y)\,(Lxy \cdot Lyx)\,, \qquad (h')$$

which makes it obvious that the evidence e logically contradicts the given hypothesis; hence, the same is true of e', although Nicod's criterion qualifies e' as confirming h.[16]

Hypotheses of the form illustrated by h can be formulated in terms of well-entrenched predicate expressions, such as "x likes y" and "x is soluble in y"; the difficulty here illustrated does not, therefore, spring from the use of ill-behaved predicates of the Goodmanian variety.

The difficulty rather shows that the intuition which informs the Nicod criterion simply fails when the hypotheses under consideration include relational terms rather than only property terms. If one considers, in addition, that the Nicod criterion it limited to hypotheses of universal conditional form, then it becomes clear that it would be of great interest to develop a general characterization of qualitative confirmation which (1) affords a full definition rather than only partial criteria for the confirmation of a hypothesis h by an evidence sentence e, (2) is applicable to any hypothesis, of whatever logical form, that can be expressed within a specified language, and (3) avoids the difficulties of the Nicod criterion which have just been pointed out.

An explicit definition of this kind for the concept "h qualitatively confirms e" has in fact been constructed for the case where h and e are formulated in a formalized language that has the structure of a first-order functional calculus without identity; h may be any sentence whatsoever in such a language, and e may be any consistent sentence containing no quantifiers. The concept thus defined demonstrably avoids the difficulties encountered by the Nicod criterion in the case of hypotheses with relational predicates; and it implies the Nicod criterion in reference to those hypoth-

[16] This further paradox of qualitative confirmation was briefly noted in my article, "Studies in the Logic of Confirmation," p. 13.

eses of universal conditional form which contain only property terms. It has been argued, however, that the concept thus arrived at is not fully satisfactory as an explication of the vague idea of qualitative confirmation because it fails to capture certain characteristics which might plausibly be attributed to the relation of qualitative confirmation.[17]

THE AMBIGUITY OF INDUCTION

I now turn to a further basic problem, which I will call the problem of inductive ambiguity. This facet of induction, unlike those we have considered so far, is not a recent discovery; both the problem and a possible solution of it have been recognized, if not always very explicitly, by several writers on probability, past as well as contemporary. But certain aspects of the problem are of special interest in the context of our discussion, and I will therefore consider them briefly.

Suppose that we have the following information:

> Jones, a patient with a sound heart, has just had an appendectomy, and of all persons with sound hearts who underwent appendectomy in the past decade, 93% had an uneventful recovery. (e_1)

This information, taken by itself, would clearly lend strong support to the hypothesis

> Jones will have an uneventful recovery. (h_1)

But suppose that we also have the information:

> Jones is a nonagenarian with serious kidney failure; he just had an appendectomy after his appendix had ruptured; and in the past decade, of all cases of appendectomy after rupture of the appendix among nonagenarians with serious kidney failure only 8% had an uneventful recovery. (e_2)

This information by itself lends strong support to the contradictory of h_1:

> Jones will not have an uneventful recovery. $(-h_1)$

[17] The general definition is developed in "A Purely Syntactical Definition of Confirmation"; the gist of it is presented in sec. 9 of my article essay, "Studies in the Logic of Confirmation." The objections in question were raised especially by R. Carnap in *Logical Foundations of Probability*, secs. 86–88. Briefly, Carnap's principal objection is to the effect that under an adequate definition of qualitative confirmation, e should confirm h only if, in the sense of inductive probability theory, e raises the prior probability of h; and my definition of confirmation is not compatible with such a construal.

But e_1 and e_2 are logically compatible and may well both be part of the information available to us and accepted by us at the time when Jones' prognosis is being considered. In this case, our available evidence provides us with a basis for two rival arguments, both of them inductively sound, whose "conclusions" contradict each other. This is what I referred to above as the ambiguity of inductive reasoning: Inductively sound reasoning based on a consistent, and thus possibly true, set of "premises" may lead to contradictory "conclusions."

This possibility is without parallel in deductive reasoning: The consequences deducible from any premises selected from a consistent set of sentences form again a consistent set.

When two sound inductive arguments thus conflict, which conclusion, if any, is it reasonable to accept, and perhaps to act on? The answer, which has long been acknowledged, at least implicitly, is this: If the available evidence includes the premises of both arguments, it is irrational to base our expectations concerning the conclusions exclusively on the premises of one or the other of the arguments; the credence given to any contemplated hypothesis should always be determined by the support it receives from the *total* evidence available at the time. (Parts may be omitted if they are irrelevant in the sense that their omission leaves the inductive support of the contemplated hypothesis unchanged.) This is what Carnap has called the *requirement of total evidence*. According to it, an estimate of Jones' prospects of recovery should be based on all the relevant evidence at our disposal; and clearly, a physician trying to make a reasonable prognosis will try to meet this requirement as best he can.

What the requirement of total evidence demands, then, is that the credence given to a hypothesis h in a given knowledge situation should be determined by the inductive support, or confirmation, which h receives from the total evidence e available in that situation. Let us call this confirmation $c(h,e)$. Now for some brief comments on this maxim.

(1) In the form just stated, the requirement presupposes a quantitative concept of the degree, $c(h,e)$, to which the evidence e confirms or supports the hypothesis h. This raises the question how such a concept might be defined and whether it can be characterized so generally that $c(h,e)$ is determined for *any* hypothesis h that might be proposed, relative to *any* body of evidence e that might be available. This issue has been much discussed in recent decades. Carnap, in his theory of inductive logic, has developed an explicit and completely general definition of the concept for the case where e and h are any two sentences expressible in one or another of certain formalized languages of relatively simple logical structure.[18] Others have ar-

[18] See especially the following publications: "On Inductive Logic," *Philosophy of Science*, 12 (1945), 72–97; *Logical Foundations of Probability; The Continuum of Inductive Methods* (Chicago: U. of Chicago Press, 1952).

gued that the concept in question can be satisfactorily defined at best for certain special types of hypotheses and of evidential information. For example, if the total relevant evidence consists just of the sentences e_1 and e_2 listed above, certain analysts would hold that no probability or degree of confirmation can be significantly assigned to the hypothesis, "Jones will have an uneventful recovery," since the evidence provides no information about the percentage of uneventful recoveries among nonagenarians with sound hearts but seriously defective kidneys who undergo appendectomy after rupture of the appendix.

(2) Next, let us note that while the requirement of total evidence is a principle concerning induction, it is not a rule of inductive inference or, more precisely, of inductive support, for it does not concern the question whether, or how strongly, a given hypothesis is supported by given evidence. The requirement is concerned rather with the rational use, or application, of inductive reasoning in the formation of empirical beliefs. This observation suggests a distinction between two kinds of rules pertaining to inductive reasoning:

(*a*) *Rules of inductive support, or of valid inductive inference.* These would encompass, for example, all criteria concerning the qualitative confirmation or disconfirmation of generalizations by positive or negative instances; criteria determining degrees of confirmation; and also all general principles connecting degrees of confirmation with each other, such as the law, that the degrees of confirmation of a hypothesis and of its contradictory on the same evidence add up to unity.

(*b*) *Rules of application.* These concern the use of rules of the former kind in the rational formation of empirical beliefs. The requirement of total evidence is one such rule of application, but not the only one, as will soon be seen.

The distinction between rules of inference and rules of application can be made also in reference to deductive reasoning. The rules of inference, as we noted earlier, provide criteria of deductive validity; but they qualify as deductively valid many particular arguments whose conclusions are false, and they do not concern the conditions under which it is reasonable to believe, or to accept, the conclusion of a deductively valid argument. To do so would be the task of rules for the rational application of deductive inference.

One such rule would stipulate, for example, that if we have accepted a set of statements as presumably true, then any logical consequence of that set (or, perhaps rather, any statement that is known to be such a consequence) should equally be accepted as presumably true.

The two kinds of rules for deduction call for quite different kinds of justification. An inference rule such as *modus ponens* might be justified by showing that when applied to true premises it will invariably yield a true conclusion—which is what is meant by the claim that an argument conforming to the rule is deductively valid.

But in order to justify a rule of application, we will have to consider what ends the acceptance or rejection of deductive conclusions is to serve. For example, if we are interested in accepting a set of statements, or of corresponding beliefs, which will afford us an emotionally reassuring or esthetically satisfying account of the world, then it will not always be reasonable to accept, or to believe, the logical consequences of what we have previously accepted. If, on the other hand, truth is what we value in our accepted statements, and if we are accordingly concerned to give credence to all statements that are true as far as our information enables us to tell, then indeed we have to accept all the consequences of previously accepted statements; thus, justification of our rule of application requires reference to the objectives, or the values, that our acceptance procedure is meant to achieve.

INDUCTION AND VALUATION

Similarly, if we wish to devise rules for the rational application of valid inductive reasoning, or if we wish to appraise or justify such rules, we will have to take into account the objectives to be achieved by the inductive acceptance procedure, or the values or disvalues of the consequences that might result from correct or from incorrect acceptance decisions. In this sense, the construction and the justification of inductive acceptance rules for empirical statements presupposes judgments of value.

This is especially obvious when we wish to decide whether a given hypothesis is to be accepted in the strong sense of being relied on as a basis for practical action. Suppose, for example, that a new vaccine has been developed for immunization against a serious infectious disease that can afflict humans as well as chimpanzees. Let h be the hypothesis that the vaccine is both safe and effective in a sense specified by suitable operational criteria, and suppose that the hypothesis has been tested by examining a number of samples of the vaccine for safety and effectiveness. Let e be the evidence thus obtained.

Our rules of inductive support may then tell us how strongly the hypothesis is confirmed by the evidence; but in deciding whether to act on it we will have to consider, besides the strength of confirmation, also the kind of action that is contemplated, and what benefits might result from a correct decision, what harm from a mistaken one. For example, our standards of acceptance are likely to differ according as humans or chimpanzees are to be treated with the vaccine; and it may well happen that *on the same evidence* the given hypothesis is accepted as a basis of action in one case but rejected in the other.

Inductive decisions of this kind have been extensively studied in the mathematical theory of testing and decision-making. This theory deals in particular with the case where the values or disvalues attached to the possible consequences of the available decisions are expressible in numerical terms as so-called utilities. For such situations, the theory has developed a number of specific decision rules, which are rules of application in our sense. These rules—maximin, maximax, maximizing the expectable utility of the outcome, and others—make the acceptance or the rejection of the hypothesis contingent on the utilities assigned to the different possible consequences of acceptance or rejection; and when a measure for the evidential support of the hypothesis is available, that support is likewise taken into consideration.[19] In this fashion, the inductive decision rules combine empirical considerations with explicitly valuational ones.

That rules for the acceptance or rejection of empirical hypotheses thus presuppose valuational considerations has been emphasized by several writers. Some of these have made the stronger claim that the values in question are ethical values. Thus, Churchman asserts that "the simplest question of fact in science requires for even an approximation, a judgment of value," and that "the science of ethics . . . is *basic* to the meaning of any question the experimental scientist raises."[20] And in the context of a detailed study of the logic of testing statistical hypotheses, Braithwaite asserts, in a similar vein: "To say that it is 'practically certain' that the next 1000 births in Cambridge will include the birth of at least one boy includes a hedonic or ethical assessment."[21]

But while it is true that the justification of rules of acceptance for statements of fact requires reference to judgments of preference or of valuation, the claim that the values concerned are ethical values is, I think, open to question. Our argument about valuational presuppositions has so far been concerned only with the acceptance of hypotheses as a basis of specific *actions,* and in this case the underlying valuations may indeed be ethical in character. But what standards will govern the acceptance and rejection of hypotheses for which no practical application is contemplated? Braithwaite's statement about male births in Cambridge might well belong in that category, and surely so do the hypotheses examined in pure, or basic, scientific research; these might concern, for example, the rate of recession of distant galaxies or the spontaneous creation of hydrogen atoms in empty space. In such cases, it seems, we simply wish to decide, in consideration of the available evidence, whether to believe a proposed hypothesis; whether to record it, so to speak, in our book of tentative scientific knowledge, without envisaging any technological application. Here, we cannot relevantly base our decisions on any utilities or disutilities attached to practical consequences of acceptance or rejection and, in particular, ethical considerations play no part.

[19] A lucid account of these rules and of their theoretical use will be found in R. D. Luce and H. Raiffa, *Games and Decisions* (New York: Wiley, 1957).

[20] C. W. Churchman, *Theory of Experimental Inference* (New York: Macmillan, 1948), pp. vii, viii (author's italics).

[21] R. B. Braithwaite, *Scientific Explanation* (Cambridge: Cambridge U. Press, 1953), p. 251.

What will have to be taken into account in constructing or justifying inductive acceptance rules for pure scientific research are the objectives of such research or the importance attached in pure science to achieving certain kinds of results. What objectives does pure scientific research seek to achieve? Truth of the accepted statements might be held to be one of them. But surely not truth at all costs. For then, the only rational decision policy would be never to accept any hypothesis on inductive grounds since, however well supported, it might be false.

Scientific research is not even aimed at achieving very high probability of truth, or very strong inductive support, at all costs. Science is willing to take considerable chances on this score. It is willing to accept a theory that vastly outreaches its evidential basis if that theory promises to exhibit an underlying order, a system of deep and simple systematic connections among what had previously been a mass of disparate and multifarious facts.

It is an intriguing but as yet open question whether the objectives, or the values, that inform pure scientific inquiry can all be adequately characterized in terms of such theoretical desiderata as confirmation, explanatory power, and simplicity and, if so, whether these features admit of a satisfactory combination into a concept of purely theoretical or scientific utility that could be involved in the construction of acceptance rules for hypotheses and theories in pure science. Indeed, it is by no means clear whether the conception of basic scientific research as leading to the provisional acceptance or rejection of hypotheses is tenable at all. One of the problems here at issue is whether the notion of accepting a hypothesis independently of any contemplated action can be satisfactorily explicated within the framework of a purely logical and methodological analysis of scientific inquiry[22] or whether, if any illuminating construal of the idea is possible at all, it will have to be given in the context of a psychological, sociological, and historical study of scientific research.[23]

[22] For a fuller discussion and bibliographic references concerning these issues, see, e.g., sec. 12 of C. G. Hempel, "Deductive-Nomological *vs.* Statistical Explanation" in *Scientific Explanation, Space, and Time,* eds. H. Feigl and G. Maxwell, Minnesota Studies in the Philosophy of Science, III (Minneapolis: U. of Minnesota Press, 1962), 98–169. Some of the basic issues are examined in R. B. Braithwaite's paper, "The Role of Values in Scientific Inference," and especially the discussion of that paper in *Induction: Some Current Issues,* eds. H. E. Kyburg, Jr., and E. Nagel (Middletown, Conn.: Wesleyan U. Press, 1963), pp. 180–204.

[23] Such an alternative conception is represented, e.g., by T. S. Kuhn's work, *The Structure of Scientific Revolutions* (Chicago: U. of Chicago Press, 1962).

To conclude with a summary that centers about the classical problem of induction: For a clear statement of the classical problem of justification, two things are required. First, the procedure to be justified must be clearly characterized—this calls for an explication of the rules governing the inductive appraisal of hypotheses and theories; second, the intended objectives of the procedure must be indicated, for a justification of any procedure will have to be relative to the ends it is intended to serve. Concerning the first of these tasks, we noted that while there are no systematic mechanical rules of inductive discovery, two other kinds of rule have to be envisaged and distinguished, namely, rules of support and rules of application. And in our discussion of the objectives of inductive procedures we noted certain connections between rational belief on one hand and valuation on the other.

Whatever insights further inquiry may yield, the recognition and partial exploration of these basic problems has placed the classical problem of induction into a new and clearer perspective and has thereby advanced its philosophical clarification.

9

DE PRINCIPIIS NON DISPUTANDUM . . . ?

[ON THE MEANING AND THE LIMITS OF JUSTIFICATION]

Herbert Feigl

Arguments purporting to justify beliefs or evaluations often proceed from specific to more general issues. Opposition and challenge tend to provoke critical reflection; through various dialectical moves higher levels of justification are reached and made explicit. Argument usually terminates with appeals to principles which are considered indisputable, at least by those who invoke them. But, notoriously, initial disagreements cannot always be removed by what is called "rational argument." Frequently enough, initial disagreement can be traced back to disagreement in basic presuppositions. It is a characteristic of those modern cultures which endorse freedom of thought that they countenance divergencies in religious, political, or economic positions. "It is all a matter of one's ultimate presuppositions"—this phrase and its variants indicate that enlightened common sense is aware of the limits of argument and justification. But on the other hand there is also the deep-rooted wish to be *right,* absolutely right, in one's basic outlook. When the disagreement concerns mere gastronomical matters, we are quite willing to reconcile ourselves with the saying, "*De gustibus non est disputandum.*" Art critics and aestheticians, however, do not unreservedly extend such tolerance to all issues of aesthetic evaluation. Most people, including the majority of philosophers, are still more reluctant to grant any relativity to the basic standards of moral evaluation. There is, at least in this age of sci-

From Max Black (ed.), *Philosophical Analysis, A Collection of Essays* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963), pp. 113–116 and 123–131. Reprinted by permission of Max Black and the author.