24

Hilary Putnam

"DEGREE OF CONFIRMATION" AND INDUCTIVE LOGIC

I

CARNAP'S attempt to construct a *symbolic inductive logic,* fits into two major concerns of empiricist philosophy. On the one hand, there is the traditional concern with the formulation of Canons of Induction; on the other hand, there is the distinctively Carnapian concern with providing a formal reconstruction of the language of science as a whole, and with providing precise meanings for the basic terms used in methodology.

Of the importance of continuing to search for a more precise statement of the inductive techniques used in science, I do not need to be convinced; this is a problem which today occupies mathematical statisticians at least as much as philosophers.

But this general search need not be identified with the particular project of defining a *quantitative* concept of "degree of confirmation". I shall argue that this last project is misguided.

Such a negative conclusion needs more to support it than "intuition"; or even than plausible arguments based on the methodology of the developed sciences (as the major features of that method may appear evident to one). Intuitive considerations and plausible argument might lead one to the conclusion that it would not be a good investment to spend ones *own* time trying to "extend the definition of degree of confirmation"; it could hardly justify trying to, say, convince Carnap that this particular project should be abandoned. But that is what I shall try to do here: I shall argue that one can *show* that no definition of degree of confirmation can be adequate or can attain what any reasonably good inductive judge might attain *without* using such a concept. To do this it will be necessary (a) to state precisely the condition of adequacy that will be in question; (b) to show that no inductive method based on a "measure function"[1] can satisfy it; and (c) to show that *some* methods (which can be precisely stated) *can* satisfy it.

From this we have a significant corollary: not every (reasonable) inductive method can be represented by a "measure function". Thus, we might also state what is to be proved here in the following form: the actual inductive procedure of science has features which are incompatible with being represented by a "measure function" (or, what is the same thing, a quantitative concept of "degree of confirmation").

II

Let us begin with the statement of the condition of adequacy. The first problem is the *kind of language* we have in mind.

What we are going to suppose is a language rich enough to take account of the *space-time-arrangement* of the individuals. Languages for which d.c. (degree of confirmation) has so far been defined are not this rich: we can express the hypothesis that five individuals are black and five red, but not the hypothesis that ten *successive* individuals are *alternately* black and red. Extension of d.c. to such a language is evidently one of the next steps on the agenda; it would still be far short of the final goal (definition of d.c. for a language rich enough for the formalization of empirical science as a whole).

In addition to supposing that our language, L, is rich enough to describe spatial relations, we shall suppose that it possesses a second sort of richness; we shall suppose that L contains elementary number theory. The problem of defining d.c. for a language which is rich enough for elementary number theory (or more broadly, for classical mathematics) might seem an insuperable one, or, at any rate, much more difficult than defining d.c. for a language in which the individuals have an "order". But such is not the case. I have shown elsewhere[2] that any measure function defined for an (applied) first order functional calculus can be extended to a language rich enough for Cantorian set theory; hence certainly rich enough for number theory, and indeed far richer than needful for the purposes of empirical science. The difficult (I claim: *impossible*) task is not the "extension to richer languages" in the *formal* sense (i.e. to languages adequate for larger parts of logic and mathematics) but the

[1]This is Carnap's term for an a priori probability distribution. Cf. Carnap's book *Logical Foundations of Probability* (Chicago: Univ. of Chicago Press, 1950); and for an excellent explanation of leading ideas, *vide* also the paper by Kemeny in this volume.

[2]"A Definition of Degree of Confirmation for Very Rich Languages," *Philosophy of Science*, XXIII, 58-62.

extension to languages richer in a *physical* sense (i.e. adequate for taking account of the fact of *order*).

In short, we consider a language rich enough for
(a) the description of space-time order.
(b) elementary number theory.

The purpose of the argument is to show that d.c. *cannot* be adequately defined for such a language. This is independent of whether or not the particular method of "extending to richer languages" used in the paper mentioned is employed. But by combining the argument of that paper with the present argument we could get a stronger result: it is not possible to define d.c. adequately in a language satisfying just (a).

To state our condition of adequacy, we will also need the notion of an *effective hypothesis* (a deterministic law).

Informally, an effective hypothesis is one which says of each individual whether or not it has a certain molecular property M; and which does so effectively in the sense that it is possible to *deduce* from the hypothesis what the character of any individual will be. Thus an effective hypothesis is one that we can *confront* with the data: one can deduce what the character of the individuals will be, and then see whether our prediction agrees with the facts as more and more individuals are observed. Formally, an hypothesis h will be called an effective hypothesis if it has the following properties:

(i) h is expressible in L.
(ii) if it is a consequence of h that $M(x_i)$ is true[3] (where M is a molecular predicate of L and x_i is an individual constant), then $h \supset M(x_i)$ is *provable* in L.
(iii) h is equivalent to a set of sentences of the forms $M(x_i)$ and $\sim M(x_i)$; where M is some molecular predicate of L, and x_i runs through the names of all the individuals.

The notion of an effective hypothesis is designed to include the hypotheses normally regarded as expressing putative universal laws. For example, if a hypothesis implies that each individual satisfies the molecular predicate[4] $P_1(x) \supset P_2(x)$, we require that (for each i) $(P_1(x_i) \supset P_2(x_i))$ should be deducible from h in L, for h to count as effective.

We can now state our condition of adequacy:

I. If h is an effective hypothesis and h is true, then the *instance confirmation* of h (as more and more successive individuals are examined) approaches 1 as limit.

[3]Logical formulas are used in this paper only as names of themselves; never in their object-language use.

[4]I.e., the predicate $P_1(\ldots) \supset P_2(\ldots)$; we use the corresponding open sentence to represent it.

We may also consider *weaker* conditions as follows:

I′. If h is an effective hypothesis and h is true, then the *instance confirmation* of h eventually becomes and remains greater than .9 (as more and more successive individuals are examined).

I″. (Same as I′, with '.5' in place of '.9'.)

Even the weakest of these conditions is violated—*must* be violated—by every measure function of the kind considered by Carnap.

III

In I and its variants we have used the term "instance confirmation"[5] introduced by Carnap. The instance confirmation of a universal hypothesis is, roughly speaking, the degree of confirmation that the next individual to be examined will conform to the hypothesis.

It would be more natural to have "degree of confirmation" in place of "instance confirmation" in I, I′, and I″. However, on Carnap's theory, the degree of confirmation of a universal statement is always zero. Carnap does not regard this as a defect; he argues[6] that when we refer to a universal statement as amply confirmed all we really mean is that the instance confirmation is very high. I shall make two remarks about this contention:

(1) This proposal is substantially the same as one first advanced by Reichenbach[7] and criticized by Nagel.[8] The criticism is simply that a very high confirmation in this sense (instance confirmation) is *compatible with any number of exceptions.*

(2) The whole project is to define a concept of degree of confirmation which underlies the scientist's "qualitative" judgments of "confirmed", "disconfirmed", "accepted", "rejected", etc. much in the way that the quantitative magnitude of *temperature* may be said to underlie the qualitative distinctions between "hot" and "cold", "warm" and "cool", etc. But a universal statement *may* be highly confirmed (or even "accepted") as those terms are actually used in science. Therefore it must have a high degree of confirmation, if the relation of "degree of confirmation" to "confirmed" is as just described. To say that it only has a high *instance* confirmation is to abandon the analogy "degree of confirmation is to confirmed as temperature is to hot". But this analogy explains what it *is* to try to "define degree of confirmation".

[5]*Logical Foundations of Probability*, 571ff.

[6]*Ibid.*, 572.

[7]*The Theory of Probability* (Berkeley, 1949). See the work cited in n. 8 for an exposition.

[8]*Principles of the Theory of Probability, International Encyclopedia of Unified Science*, I, no. 6 (Chicago: Univ. of Chicago Press, 1939), 63f.

(Carnap's reply is to maintain the analogy, and deny that a universal statement is ever really confirmed; what is really confirmed, on his view, is that no exceptions will be found in, say, our lifetime, or the lifetime of the human race, or *some* specifiable space-time region (which must be finite).)

IV

Before we proceed to the main argument, let us consider the possibility of obviating the entire discussion by *rejecting* I (and its weaker versions). To do this is to be willing to occupy the following position: (a) one accepts a certain system of inductive logic, based on a function c for "degree of confirmation", as *wholly adequate*; (b) one simultaneously admits that a certain effective hypothesis h is such that if it be true, we will never discover this fact by our system.

Such a position might indeed be defended by maintaining that certain effective hypotheses are *unconfirmable in principle.* For instance, suppose that we have an ordered series of individuals of the same order-type as the positive integers. Let h be the hypothesis that every individual in the series with a prime-numbered position is red and every individual with a composite position is black (count x_1 as "composite").

In other words, $x_1, x_4, x_6, x_8, x_9, x_{10}$, etc. are all black; $x_2, x_3, x_5, x_7, x_{11}$, etc. are all red.

Someone might reason as follows:

The arithmetic predicates "prime" and "composite" do not appear in a single known scientific law; therefore such a "hypothesis" is not a legitimate scientific theory, and it is only these that we require to be confirmable. In short, it is not a defect of the system if the hypothesis h just described cannot be confirmed (if its instance confirmation does not eventually exceed, and remain greater than, .9 or even .5).

But this reasoning does not appear particularly plausible; one has only to say—

"Of course the situation described by h has not so far occurred in our experience (as far as we know); but *could we find it out* if it did occur"?

I think the answer is clearly "yes"; existing inductive methods are capable of establishing the correctness of such a hypothesis (provided someone is bright enough to suggest it), and so must be any adequate "reconstruction" of those methods.

Thus, suppose McBannister says:

"You know, I think this is the rule: the prime numbers are occupied by red!"

We would first check the data already available for consistency with McBannister's hypothesis. If McBannister's hypothesis fit the first thousand or so objects, we might be impressed, though perhaps not enough to "accept". But if we examined another thousand, and then a million, and then ten million objects and McBannister's suggestion "held-up"—does anyone want to suggest that a reasonable man would *never* accept it?

A similar argument may be advanced if instead of the predicate "prime" we have any recursive predicate of positive integers. It may take a genius, an Einstein or a Newton, to *suggest* such a hypothesis (to "guess the rule", as one says) ; but once it has been suggested any reasonably good inductive judge can verify that it is true. One simply has to keep examining new individuals until any other, antecedently more plausible, hypotheses that may have been suggested have all been ruled out.

In short, if someone rejects I (and its several versions) he must be prepared to offer one of the following "defenses":

(a) I know that if h is true I won't find it out; but I am "gambling" that h is false.

(b) If h turns out to be true, I will *change my inductive method.*

Against the first "defense" I reply that such a "gamble" would be justifiable only if we could show that *no* inductive method will find it out if h is true (or at least, the *standard* inductive methods will not enable me to accomplish this). But in the case of McBannister's hypothesis about the prime-numbered objects and similar hypotheses, this cannot be urged. Against the second "defense" I reply that this defense *presupposes that one can find out* if h "turns out to be true." But, from the nature of h, the only way to find out would be *inductively*. And if one has an inductive method that will accomplish this, then one's definition of degree of confirmation is evidently not an adequate reconstruction of that inductive method.

V

To simplify the further discussion, we shall suppose that there is only *one* dimension, and not four, and that the series of positions is discrete and has a beginning. Thus we may name the positions $x_1, x_2, x_3, \ldots$ etc. (Following a suggestion of Carnap's we will identify the positions and the individuals. Thus "x_1 is red" will mean "the position x_1 is occupied by something red" or "red occurs at x_1"). The modification of our argument for the actual case (of a four-dimensional space-time continuum) is simple.[9]

[9]Thus we may suppose that $x_1, x_2, \ldots$ are a subsequence of observed positions from the whole four-dimensional continuum; and that the hypotheses under consideration differ only with respect to these.

Next we suppose a function c for degree of confirmation to be given. Technically, c is a function whose arguments are sentences h and e, and whose values are real numbers, $0 \leq c(h,e) \leq 1$. The numerical value of c(h,e) is supposed to measure the extent to which the statement expressed by h is confirmed by the statement expressed by e; thus c(h,e) may conveniently be read "the degree of confirmation of h on evidence e".

Admissible functions c for degree of confirmation are required by Carnap to fulfill several conditions. One of these conditions is that the degree of confirmation of $M(x_i)$ should converge to the relative frequency of M in the sample, as more and more individuals other than x_i are examined. This requirement can no longer be maintained in this form in the case of an *ordered* set of individuals; but the following weaker version must still be required:

II. For every n (and every molecular property M) it must be possible to find an m such that, if the next m individuals (the individuals $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$) are all M, then the d.c. of the hypothesis $M(x_{n+m+1})$[10] is greater than .5, regardless of the character of the first n individuals.

If n is 10, this means that there must be an m, say 10,000,000 such that we can say: if the individuals $x_{11}, x_{12}, \ldots, x_{10,000,000}$ are all red, then the probability is more than one-half that $x_{10,000,001}$ will be red (whether or not some of $x_1, x_2, \ldots, x_{10}$ are non-red).

What is the justification of II? Let us suppose that II were violated. Then there must be an n (say, 10) and a property M (say, "red") such that, for some assignment of "red" and "non-red" to $x_1, x_2 \ldots, x_{10}$ (say, x_1,x_2,x_3 are red; $x_4,x_5,\ldots,x_{10}$ are non-red) it holds that no matter how many of $x_{11}, x_{12}, \ldots$, are red, the d.c. that the *next* individual will be red does not exceed .5. Therefore the hypothesis h: x_1,x_2,x_3 are red; $x_4,x_5,\ldots x_{10}$ are non-red; x_{11} and all subsequent individuals are red—violates I (and in fact, even I''). For no matter how many successive individuals are examined, it is not the case that the instance confirmation of h (this is just the probability that the *next* individual will be red) becomes and remains greater than .5.

Thus I entails II. But II is independently justifiable: if II were violated, then there would be a hypothesis of an exceptionally simple kind such that we could never find it out if it were true; namely a hypothesis which says *all* the individuals (with a specified finite number of exceptions) are M. For we would know that h above is true if we knew that "all the individuals with seven exceptions are red", once we had observed $x_1,x_2,\ldots,x_{10}$. Thus if we want hypotheses of the simple form

[10]Relative to a complete description with respect to M of the individuals $x_1, x_2, \ldots x_{n+m}$. A similar "evidence" will be understood in similar cases.

"all individuals, with just n exceptions, are M" to be confirmable (to have an instance confirmation which eventually exceeds .5), we must accept II.

One more point: c cannot be an *arbitrary* mathematical function. For example, if the value of c were *never* computable, it would be no use to anybody. All the c-functions so far considered by Carnap and other workers in this field have very strong properties of computability. For instance, the d.c. of a singular hypothesis relative to singular evidence is always computable. However this will not be assumed here (although it would materially simplify the argument); all I will assume is the very weak condition: the "it must be possible to find" in II means *by an effective process*. In other words,[11] for each n (say, 10) there is some m (say, 10,000,000) such that one can *prove* (in an appropriate metalanguage M_L) that if $x_{11}, x_{12}, \ldots, x_{10,000,000}$ are "red", then the d.c. that the *next* individual will be "red" is greater than one-half.

If this is not satisfied, then (by an argument parallel to the above) there is some hypothesis of the simple form "all the individuals, with just n exceptions, are M" such that we *cannot prove* at any point (with a few exceptions "at the beginning") that it is more likely than not that the next individual will conform.

E.g. even if we have seen only "red" things for a very long time (except for the seven "non-red" things "at the beginning"), we cannot prove that the d.c. is more than .5 that the next individual will be red.

We can now state our result:

Theorem: there is no definition of d.c. which satisfies II (with the effective interpretation of "it is possible to find") and also satisfies I.

The following proof of this theorem proceeds *via* what mathematical logicians call a "diagonal argument".

Let C be an infinite class of integers $n_1, n_2, n_3, \ldots$ with the following property: the d.c. of $\mathrm{Red}(x_{n_1})$ is greater than .5 if all the preceding individuals are red; the d.c. of $\mathrm{Red}(x_{n_2})$ is greater than .5 if all the preceding individuals *after* x_{n_1} are red; and, in general, the d.c. of $\mathrm{Red}(x_{n_j})$ is greater than .5 if all the preceding individuals *after* $x_{n_{j-1}}$ are red.

The existence of a class C with this property is a consequence of II. For (taking n = 0) there must be an n_1 such that if the first $n_1 - 1$ individuals are red, the d.c. is greater than one-half that x_{n_1} is red. Choose such an n_1: then there must be an m such that if the individuals $x_{n_1+1}, x_{n_1+2}, \ldots, x_{n_1+m}$ are all red, the d.c. is more than one-half that x_{n_1+m+1} is red; call n_{1+m+1} "n_2": . . . . etc.

Moreover, if we assume the "effective" interpretation of "it must be possible to find" in II, there exists a *recursive* class C with this property.

[11]What follows "in other words" entails the existence of such an effective process, because it is effectively possible to enumerate the *proofs* in M_L.

(A class is "recursive" if there exists a mechanical procedure for determining whether or not an integer is in the class.) We shall therefore asume that our chosen class C is "recursive".

A predicate is called "arithmetic" if it can be defined in terms of polynominals and quantifiers.[12] For instance, the predicate "n is square" can be defined by the formula $(\exists m)\ (n = m^2)$, and is therefore arithmetic.

Now, Gödel has shown that every recursive class is the extension of an arithmetic predicate.[13] In particular, our class C is the extension of some arithmetic predicate P. So we may consider the following hypothesis:

(1) An individual x_n is red if and only if $\sim P(n)$.

Comparing this with McBannister's hypothesis:

(2) An individual x_n is red if and only if n is prime.

We see that (1) and (2) are of the same form. In fact, the predicate "is prime" is merely a particular example of a recursive predicate of integers.

Thus the hypothesis (1) is *effective*. It is expressible in L, because P is arithmetic; it satisfies condition (ii) in the definition of "effective" (see above), because P is recursive; and (iii) is satisfied, since (1) says for each x_i either that $\mathrm{Red}(x_i)$ or $\sim\mathrm{Red}(x_i)$.

But the hypothesis (1) violates I. In other words, a scientist who uses c would never discover that (1) is true, even if he were to live forever (and go on collecting data forever). This can be argued as follows: However we interpret "discover that (1) is true", a scientist who has discovered that (1) is true should reflect this in his behavior to this extent: he should be willing to bet at even money that the next individual will be non-red whenever (1) *says* that the next individual will be non-red (the more inasmuch as the a priori probability of "non-red" is greater than "red"). But, by the definition of C, the scientist will bet at *more* than even money (when he has examined the preceding individuals) that each of the individuals $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$, is *red*. Thus he will make infinitely many mistakes, and his mistakes will show that he has never learned that (1) is true.

Finally, it is no good replying that the scientist will be right more often than not. The aim of science is not merely to be right about particular events, but to discover general laws. And a method that will not *allow* the scientist to accept the law (1), even if someone *suggests* it, and even if no exception has been discovered in ten billion years, is unacceptable.

[12]This usage is due to Gödel.

[13]"Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I" *Monatshefte für Mathematik und Physik*, XXXVIII, 173-198. Cf. Kleene's *Introduction to Mathematics*, (New York: Van Nostrand, 1952), Theorem X., 292, and Theorem I, 241.

VI

One might suspect that things are not so black as they have just been painted; perhaps it is the case that *every* formalized system of inductive logic suffers from the difficulty just pointed out, much as every formalized system of arithmetic suffers from the incompleteness pointed out by Gödel. It is important to show that this is not so; and that other approaches to induction—e.g. that of Goodman,[14] or that of Kemeny[15] are not necessarily subject to this drawback.

Many factors enter into the actual inductive technique of science. Let us consider a technique in which as few as possible of these factors play a part: to be specific, only the direct factual support[16] (agreement of the hypothesis with the data) and the previous acceptance of the hypothesis.[17] Because of the highly over-simplified character of this technique, it is easily formalized. The following rules define the resulting inductive method (M):

1. Let $P_{t,M}$ be the set of hypotheses considered at time t with respect to a molecular property M. I.e. $P_{t,M}$ is a finite set of effective hypotheses, each of which specifies, for each individual, whether or not it is M.
2. Let $h_{t,M}$ be the effective hypothesis on M *accepted* at time t (if any). I.e. we suppose that, at any given time, various incompatible hypotheses have been actually suggested with respect to a given M, and have not yet been ruled out (we require that these should be consistent with the data, and with accepted hypotheses concerning other predicates). In addition, one hypothesis may have been accepted at some time prior to t, and may not yet have been abandoned. This hypothesis is called the "accepted hypothesis at the time t". So designating it is not meant to suggest that the other hypotheses are not considered as serious candidates for the post of accepted hypotheses on M" at some later t.
3. (Rule I:) At certain times t_1 t_2 t_3 . . . initiate an *inductive test with respect to M*. This proceeds as follows: the hypotheses in P_{t_i},M at this time t_i are called the *alternatives*. Calculate the character (M or not-M) of the next individual on the basis of each alternative. See which alternatives succeed in predicting this. Rule out those that fail. Continue until (a) all alternatives but one have failed; or (b) all alternatives have failed; (one or the other must eventually happen). In case (a) *accept* the alternative that does not fail. In case (b) reject all alternatives.
4. (Rule II:) hypotheses suggested in the course of the inductive test are taken as alternatives (unless they have become inconsistent with the data) in the *next* test. I.e. if h is proposed in the course of the test begun at t_3, then h belongs to $P_{t_4,M}$ and not to $P_{t_3,M}$.
5. (Rule III:) if $h_{t,M}$ is accepted at the conclusion of any inductive test, then $h_{t,M}$ continues to be accepted as long as it remains consistent with the data. (In particular, while an inductive test is still going on, the previously accepted hypothesis continues to be accepted, for all practical purposes.)

[14]*Fact, Fiction & Forecast* (Cambridge, Mass.: Harvard Univ. Press, 1955).

[15]"The Use of Simplicity in Induction," *Philosophical Review*, LXII, 391-408.

[16]This term has been used in a related sense by Kemeny and Oppenheim, "Degree of Factual Support," *Philosophy of Science*, XIX, 307-324.

[17]This factor has been emphasized by Conant, *On Understanding Science* (New Haven, Conn.: Yale Univ. Press, 1947).

Ridiculously simple as this method M is, it has some good features which are not shared by any inductive method based on a "measure function". In particular:

III. If h is an effective hypothesis, and h is true; then, using method M, one will eventually accept h if h is ever proposed.

The method M differs from Carnap's methods, of course, in that the acceptance of a hypothesis depends on which hypotheses are actually proposed, and also on the *order* in which they are proposed. But this does not make the method informal. Given a certain sequence of sentences, (representing the suggested hypotheses and the order in which they are suggested), and given the "time" at which each hypothesis is put forward (i.e. given the *evidence* at that stage: this consisting, we may suppose, of a complete description of individuals $x_1, x_2, \ldots, x_t$ for some t); and given, finally, the "points" (or evidential situations) at which inductive tests are begun; the "accepted" hypothesis at any stage is well defined.

That the results a scientist gets, using method M, depend on (a) what hypotheses he considers at any given stage, and even (b) at what points he chooses to inaugurate observational sequences ("inductive tests") is far from being a *defect* of M: these are precisely features that M shares with ordinary experimental and statistical practice. (Carnap sometimes seems to say[18] that he is looking for something *better* than ordinary experimental practice in these respects. But this is undertaking a task far more ambitious, and far more doubtful, than "reconstruction".)

In comparing the method M with Carnap's methods, the problem arises of correlating the essentially qualitative notion of "acceptance" with the purely quantitative notion of "degree of confirmation". One method is this (we have already used it): say that a hypothesis is *accepted* if the instance confirmation is greater than .5 (if one is willing to bet

[18]*Logical Foundations of Probability*, 515-520; see esp. the amazing paragraph at the bottom of p. 518!

at more than even money that the next individual will conform). In these terms, we may say: using Carnap's methods one will, in general, accept an effective hypothesis sooner or later if it is true, and in fact one will accept it infinitely often. But one won't *stick* to it. Thus these methods lack *tenacity*.

Indeed, we might say that the two essential features of M are

i) *corrigibility:* if h is inconsistent with the data, it is abandoned; and

ii) *tenacity*: if h is once accepted, it is not subsequently abandoned *unless* it becomes inconsistent with the data.

It is the first feature that guarantees that any effective hypothesis will eventually be accepted if true; for the other alternatives in the set $P_{t_1,M}$ to which it belongs must all be false and, for this reason, they will all eventually be ruled out while the true hypothesis remains. And it is the second feature that guarantees that a true hypothesis, once accepted, is not subsequently rejected.[19]

It would, of course, be highly undesirable if, in a system based on "degree of confirmation" one had "tenacity" in quite the same sense. If we are willing to bet at more than even money that the next individual will conform to h, it does not follow that if it *does* conform we should *then* be willing to bet that the next individual in turn will conform. For instance, if we are willing to bet that the next individual will be red, this means that we are betting that it will conform to the hypothesis that all individuals are red; and also that it will conform to the hypothesis that all individuals up to and including it are red, and all those thereafter green.[20] If it *does* conform to both these hypotheses, we cannot go on to bet that the next individual in turn will conform to both, for this would involve betting that it will be both red and green.[21] But we can say this: for any effective hypothesis h, there should come a point (if h continues to be consistent with the data) at which we shall be willing to bet that the next individual will conform; and if the next individual conforms, we shall be willing to bet that the next in turn will conform; and so on. To say this is merely to say again: if it is true we ought *eventually* to accept it. And it is to this simple principle that M conforms, while the Carnapian methods do not.

[19]It is of interest to compare III with the "pragmatic justification of induction" given by Feigl, "De Principiis non Disputandum . . . ?" in *Philosophical Analysis*, ed. by M. Black (Ithaca, 1950).

[20]The difficulty occasioned by pairs of hypotheses related in this way was first pointed out by Goodman. *Vide* "A Query on Confirmation," *Journal of Philosophy*, XLIII, 383-385.

[21]This raises a difficulty for Reichenbach's "Rule of Induction"; the use of the rule to estimate the relative frequency of "green" and "grue" (see below) is another case in which contradictory results are obtained.

Moreover, that the method M has the desirable property III is closely connected with a feature which is in radical disagreement with the way of thinking embodied in the "logical probability" concept: the acceptance of a hypothesis depends on *which* hypotheses are actually proposed. The reader can readily verify that it is this feature (which, I believe, M shares with the actual procedure of scientists) that blocks a "diagonal argument" of the kind we used in the preceding section. In short, M is *effective*, and M is able to discover any true law (of a certain simple kind); but this is because what we will predict "next", using M, does not depend *just* on the evidence. On the other hand, it is easily seen that any method that shares with Carnap's the feature: what one will predict "next" depends *only* on what has so far been observed, will also share the defect: either what one should predict will not in practice be *computable*,[22] or some law will elude the method altogether (one is *in principle* forbidden to accept it, no matter how long it has succeeded).

This completes the case for the statement made at the beginning of this paper: namely, that a good inductive judge can do things, provided he does *not* use "degree of confirmation", that he could not *in principle* accomplish if he *did* use "degree of confirmation". As soon as a scientist announces that he is going to use a method based on a certain "c-function", we can exhibit a hypothesis (in fact, one consistent with the data so far obtained, and hence possibly true) such that we can say: if this is true *we* shall find it out; but you (unless you abandon your method) will never find it out.

Also, we can now criticize the suggested analogy between the "incompleteness" of Carnap's systems, and the Godelian incompleteness of formal logics. A more correct analogy would be this: the process of *discovery* in induction is the process of suggesting the correct hypothesis (and, sometimes, a suitable language for its expression and a mathematical technique that facilitates the relevant computation).

But once it has been suggested, the inductive checking, leading to its eventual acceptance, is relatively straightforward. Thus the suggestion of a hypothesis in induction is analogous to the *discovery of a proof* in formal logic; the inductive verification (however protracted, and however many "simpler" hypotheses must first be ruled out) is analogous to the *checking* of a formal proof (however tedious). Thus one might say: the incompleteness we have discovered in Carnap's system is analogous to the "incompleteness" that would obtain if there were no mechanical way of *checking* a proof, once discovered, in a formal logic. (Most logi-

[22]Even in the case of induction by simple enumeration; i.e., there will be hypotheses of the simple form "all individuals from x_n on are red," such that one will not be able to prove that one should accept them, no matter how many "red" things one sees.

cians[23] would hesitate at applying the word "proof" in such a case.) On the other hand, in the system M, it may take a genius to *suggest* the correct hypothesis; but if it *is* suggested, we can verify it.

VII

The oversimplified method M ignores a great many important factors in induction. Some of these, like the reliability of the evidence, are also ignored by Carnap's methods. In addition there is the simplicity of the hypothesis (e.g. the data may be consistent with McBannister's hypothesis, and also with the simpler hypothesis "no individual is red"); the "entrenchment" of the various predicates and laws in the language of science;[24] etc.

Also, the method M is only a method for selecting among deterministic hypotheses. But we are often interested in selecting from a set of statistical hypotheses, or in choosing between a deterministic hypothesis and a statistical hypothesis (the use of the "null hypothesis"[25] is a case in point). This is, in fact, the normal case: a scientist who considers the hypothesis "all crows are black" is not likely to have in mind an alternative deterministic hypothesis, though he might (i.e. all the crows in such-and-such regions are black; all those in such-and-such other regions are white, etc.); he is far more likely to choose between this hypothesis and a statistical hypothesis that differs reasonably from it (e.g. "at most 90% of all crows are black").

It is not difficult to adapt the method M to the consideration of statistical hypotheses. A statistical hypothesis is ruled out when it becomes statistically inconsistent with the data at a pre-assigned confidence level. (A statistical hypothesis, once ruled out, may later "rule itself back in"; but a deterministic hypothesis, as before, is ruled out for good if it is ruled out at all). This involves combining the method M with the standard method of "confidence intervals". If a statistical hypothesis is true, we cannot guarantee that we shall "stick to it": this is the case because a statistical regularity is compatible with arbitrarily long finite stretches of any character whatsoever. But the *probability* that one will stick to the true hypothesis, once it has been accepted, converges to 1. And if a deterministic hypothesis is true, we will eventually accept it and "stick to it" (if someone suggests it).[26]

[23] E.g. Quine, in *Methods of Logic*, 245.

[24] *Fact, Fiction & Forecast*, 95.

[25] (The hypothesis that the character in question is randomly distributed in the population.)

[26] The above is only a sketch of the method employed in extending M to statistical hypotheses. For statistical hypotheses of the usual forms, this method can be fully elaborated.

Another approach, with a feature very similar to III above, has been suggested by Kemeny.[27] This method rests on the following idea: the hypotheses under consideration are assigned a *simplicity order*. This may even be arbitrary; but of course we would like it to correspond as well as possible to our intuitive concept of simplicity. Then one selects the simplest hypothesis consistent with the data (at a pre-assigned confidence level).

Thus, if we have three incompatible hypotheses h_1, h_2, h_3, we have to wait until at most one remains consistent with the data, if we use the method M. And this may take a very long time. Using Kemeny's method, one will, in general, make a selection much more quickly.

On the other hand, Kemeny's method does not make it unnecessary to take into account *the hypotheses that have in fact been proposed*, as one might imagine. (E.g. one might be tempted to say: choose the simplest hypothesis *of all those in the language.)* For one cannot effectively enumerate all the effective hypotheses on a given M in the language.[28] However, we may suppose that a scientist who suggests a hypothesis shows that it is effective (that it does effectively predict the relevant characteristic); and shows that it does lead to different predications than the other hypotheses. Then with respect to the class $P_{t_i,M}$ of hypotheses belonging to the inductive test we may apply the Kemeny method; since every hypothesis in the class is effective, and no two are equivalent. For instance, one might simply take the hypothesis with the fewest symbols as the simplest (i.e. a 10-letter hypothesis is simpler than a 20-letter); but this would be somewhat crude. But even a *very* crude method such as this represents an improvement on the method M above, and a closer approximation to actual scientific practice.

It is instructive to consider the situation in connection with an oversimplified example. The following excellent example is due to Goodman:[29]

(1) All emeralds are green.

(2) All emeralds are green prior to time t; and blue subsequently.

We might object to (2) on the ground that it contains the name of a specific time-point (t). This does not appear to me to be a good objection. The hypothesis that the first 100 objects produced by a certain machine will be red; the next 200 green; the next 400 red; etc. mentions a particular individual (the machine) and a particular time-point (the point at which the machine starts producing objects). But a scientist who is forbidden to open the machine or investigate its internal construction

[27] "The Use of Simplicity in Induction," *Philosophical Review*, LXII, 391-408.

[28] This is a consequence of Gödel's theorem.

[29] *Fact, Fiction & Forecast*, 74.

might "behavioristically" acquire a considerable inductive confidence in this hypothesis.

Moreover, Goodman has shown how to rephrase (2) so that this objection is avoided. Define "grue" as applying to green objects prior to t; and to blue objects subsequently. Then (2) becomes:

(2′) All emeralds are grue.

What interests us about the hypotheses (1) and (2) (or (1) and (2′)) is this: if time t is in the future and all emeralds so far observed are green, both are consistent with the data. But in some sense (2) is less simple than (1). Indeed, if the language does not contain "grue", (1) is simpler than (2) by the "symbol count" criterion of simplicity proposed above. How do these hypotheses fare under the inductive methods so far discussed?

Under the method M, there are three relevant possibilities: (2) may be suggested at a time when no one has thought of (1) (highly implausible); or (1) and (2) may be suggested at the same time (slightly more plausible); or (1) may be advanced long before anyone even thinks of (2) (much more plausible, and historically accurate). In the last (and actual) case what happens is this: (1) is compared with, say

(3) All emeralds are red.

and (1) is accepted. Much later someone (Goodman, in fact) suggests (2). Then (1) is *still* accepted, in accordance with the principle of "tenacity", until and unless at time t (2) turns out to be correct.

In the case that (2) is suggested first we would, of course, accept (2) and refuse to abandon it in favor of the simpler hypothesis (1) until experimental evidence is provided in favor of (1) over (2) at time t. As Conant has pointed out[30] this is an important and essential part of the actual procedure of science: a hypothesis once accepted is not easily abandoned, even if a "better" hypothesis appears to be on the market. When we appreciate the connection between tenacity and the feature III of our inductive method, we may see one reason for this being so.

In the remaining case, in which (1) and (2) are proposed at the same time, *neither* would be accepted before time t. This is certainly a defect of method M.

Now let us consider how these hypotheses fare under Kemeny's method (as here combined with some features of method M). If (1) is suggested first, everything proceeds as it did above, as we would wish. If (2) is suggested first, there are two possibilities: we may have a rule of tenacity, according to which a hypothesis once adopted should not be abandoned until it proves inconsistent with the data. In this case things

[30]*On Understanding Science*, chap. 3.

will proceed as with the method M. Or, we may adopt the rule that we shift to a simpler hypothesis if one is suggested, provided it is consistent with the data. In this case we must be careful that only a finite number of hypotheses are simpler than a given hypothesis under our simplicity-ordering; otherwise we may sacrifice the advantages of the principle of tenacity (i.e. one might go on "shifting" forever). Then we would adopt (1) when it is suggested, even if we have previously accepted (2) and (2) is still consistent with the data. Lastly, if (1) and (2) are suggested at the same time, we will accept (1) (as soon as the "null hypothesis" is excluded at the chosen confidence level).[31]

Thus the method incorporating Kemeny's proposal has a considerable advantage over M; it permits us to accept (1) long before t even if (2) and (2′) are also available. In general, this method places a premium on simplicity, as M does not.

Another suggestion has been made by Goodman. Goodman rejects (2) as an inductive hypothesis as *explicitly* mentioning a particular time-point. This leaves the version (2′), however. So the notion of *entrenchment* is introduced. A predicate is better entrenched the more often it (or any predicate coextensive with it) has been used in inductive inferences. Under this criterion it is clear that "green" is a vastly better-entrenched predicate than the weird predicate "grue". So in any conflict of this kind, the data are regarded as confirming (1) and not (2′).

Goodman's proposal might be regarded as a special case of Kemeny's. Namely, we might regard the ordering of hypotheses according to "entrenchment" as but one of Kemeny's simplicity-orders. On the other hand, we may desire to have a measure of simplicity as distinct from entrenchment. (Under most conceptions, simplicity would be a *formal* characteristic of hypotheses, whereas entrenchment is a *factual* characteristic.) In this case we might order hypotheses according to some weighted combination of simplicity and entrenchment (assuming we can decide on appropriate "weights" for each parameter).

What has been illustrated is that the aspects of simplicity and entrenchment emphasized by Kemeny and Goodman (and any number of further characteristics of scientific hypotheses) can be taken into consideration in an inductive method without sacrificing the essential characteristics of *corrigibility* and *tenacity* which make even the method M, bare skeleton of an inductive method though it may be, superior as an inductive instrument to any method based on an a priori probability distribution.

[31]It is desirable always to count the null hypothesis as simplest; i.e., not to accept another until this is ruled out.

VIII

At the beginning of this paper I announced the intention to present a precise and formal argument of a kind that I hope may convince Carnap. I did this because I believe (and I am certain that Carnap believes as well) that one should never abandon a constructive logical venture because of *merely* philosophical arguments. Even if the philosophical arguments are well taken they are likely to prove *at most* that the scope or significance of the logical venture has been misunderstood. Once the logical venture has succeeded (if it does succeed), it may become important to examine it philosophically and eliminate conceptual confusions; but the analytical philosopher misconstrues his job when he advises the logician (or any scientist) to stop what he is doing.

On the other hand, it is not the part of wisdom to continue what one is doing no matter what relevant considerations may be advanced against it.

If the venture is logical, so must the considerations be. And in the foregoing sections we have had to provide strict proof that there are features of ordinary scientific method which cannot be captured by any "measure function". (Unless one wants to try the doubtful project of investigating measure functions which are not effectively computable, *even for a finite universe*.[32] And then one sacrifices other aspects of the scientific method as represented by M; its *effectiveness* with respect to what hypothesis one should select, and hence what prediction one should make.)

In short, degree of confirmation is supposed to represent (quantitatively) the judgments an ideal inductive judge would make. But the judgments an ideal inductive judge would make would presumably have this character: if a deterministic law (i.e. an effective hypothesis) h is true, and someone suggests it, and our "ideal judge" observes for a *very* long time that h yields only successful prediction, he will eventually base his predictions on it (and continue to do so, as long as it does not fail). But this very simple feature of the inductive judgments he makes is represented by no measure function whatsoever. Therefore, *the aim of representing the inductive policy of such a "judge" by a measure function represents a formal impossibility.*

Now that the formal considerations have been advanced, however, it becomes of interest to see what can be said on less formal grounds about the various approaches to induction. In the present section, let us see what can be said about the *indispensibility of theories* as instruments of prediction on the basis of the inductive methods we have considered.

[32]If a particular measure-function is computable for finite universes, the d.c. of a singular prediction on singular evidence is computable for *any* universe.

We shall find that the method M and the method incorporating Kemeny's idea "make sense" of this; the Carnapian methods give a diametrically opposite result.

To fix our ideas, let us consider the following situation: prior to the first large scale nuclear explosion various directly and indirectly relevant observations had been made. Let all these be expressed in a single sentence in the observation vocabulary, e. Let h be the prediction that, when the two subcritical masses of uranium 235 are "slammed together" to produce a single super-critical mass, there will be an explosion. It may be formulated without the theoretical expression "uranium 235", namely as a statement that when two particular "rocks" are quickly "slammed together" there will be "a big bang". Then h is also in the observation vocabulary. Clearly, good inductive judges, given e, did in fact expect h. And they were right. But let us ask the question: is there any *mechanical rule* whereby given e one could have found out that one should predict h?

The example cited is interesting because there was not (or, at any rate, we may imagine there was not) any *direct* inductive evidence from the standpoint of induction by simple enumeration, to support h. No rock of this kind had ever blown up (let us suppose). Nor had "slamming" two such rocks together ever had any effect (critical mass had never been attained). Thus the direct inductive inference *a la* Mill would be: "slamming two rocks of this kind (or any kind) together does not make them explode." But a *theory* was present; the theory had been accepted on the basis of *other* experiments; and the theory *entailed* that the rocks would explode if critical mass were attained quickly enough (assuming a coordinating definition according to which "these rocks" are U-235). Therefore the scientists were willing to make this prediction in the face of an utter lack of direct experiential confirmation.[33]

(Incidentally, this is also a refutation—if any were needed—of Bridgman's view of scientific method. According to Bridgman, a theory is a summary of experimental laws; these laws should be explicitly formulated, and should be accepted only insofar as they are directly confirmed (apparently, by simple enumerative induction). Only in this way shall we avoid unpleasant "surprises".[34])

But, if this view is accepted, then the scientists in the experiment described above were behaving most irrationally; they were willing to accept, at least tentatively (and advise the expenditure of billions of dollars on the basis of) an experimental law that had never been tested

[33]The physics in this example is slightly falsified, of course; but not essentially so.

[34]This seems the only possible reading of a good many passages in *The Logic of Modern Physics* (New York, 1927).

once, simply because it was deduced from a theory which entailed *other* experimental laws which had been verified.

I believe that we should all want to say that even the most "ideal inductive judge" could not have predicted h on the basis of e unless someone had suggested the relevant theories. The theories (in particular, quantum mechanics) are what connect the various facts in e (e.g. the fact that one gets badly burned if he remains near one of the "rocks") with h. Certainly it appears implausible to say that there is a *rule* whereby one can go from the observational facts (if one only had them all written out) to the observational prediction without any "detour" into the realm of theory. But this is a consequence of the supposition that degree of confirmation can be "adequately defined"; i.e. defined in such a way as to agree with the actual inductive judgments of good and careful scientists.

Of course, I am not accusing Carnap of believing or stating that such a rule exists; the existence of such a rule is a *disguised* consequence of the assumption that d.c. can be "adequately defined", and what I hope is that establishing this consequence will induce Carnap, as it has induced me, to seek other approaches to the problem of inductive logic.

Thus let O be the observational language of science, and let T be a formalization of the full-fledged language of science, including both observational and theoretical terms. O we may supppose to be an applied First Order Functional Calculus; and we may suppose it contains only (qualitative) predicates like "Red" and no functors. T, on the other hand, must be very rich, both physically and mathematically. Then we state: *if d.c. can be adequately defined for the language O, then there exists a rule of the kind described.*

Incidentally, it is clear that the possibility of defining d.c. for T entails the existence of a rule which does what we have described (since all the relevant theories can be expressed in T). But this is not as disturbing, for the creative step is precisely the invention of the theoretical language T.[35] What one has to show is that the possibility of defining d.c. just for O has the same consequence.

Carnap divides all inductive methods into two kinds. For those of the first kind, *the d.c. of h on e must not depend on the presence or absence in the language of predicates not occurring in either h or e.* Since h and e do not mention any theoretical terms, the d.c. of h on e must be the same, in such a method, whether the computation is carried out in T or O! In short, if we have a definition of d.c. in O, what we have is nothing less than a definition of *the best possible prediction in any evidential situation,* regardless of what laws scientists of the future may

[35]This has been remarked by Kemeny, in his paper in the present volume.

discover. For if the degree of confirmation of h on e is, say, .9 in the complete language T, then it must be .9 in the sub-language O.

For inductive methods of the second kind, the d.c. of h on e depends, in general on κ (the number of strongest factual properties). But, with respect to the actual universe, each method of the second kind coincides with some method of the first kind (as Carnap points out).[36] Thus, if there is any adequate method of the second kind (for the complete language T) there is also some adequate method of the first kind.

If we recall that the degree of confirmation of a singular prediction is effectively computable relative to singular evidence, we get the further consequence that it is possible in principle to build an electronic computer such that, if it could somehow be given all the observational facts, it would always make the best prediction—i.e. the prediction that would be made by the best possible scientist if he had the best possible theories. *Science could in principle be done by a moron* (or an electronic computer).[37]

From the standpoint of method M, however, the situation is entirely different. The prediction one makes will depend on what laws one accepts. And what laws one accepts will depend on what laws are proposed. Thus M does not have the counter-intuitive consequence just described. If two "ideally rational" scientists both use M, and one thinks of quantum mechanics and the other not, the first may predict h given e while the second does not. Thus theories play an indispensible role.

This feature is intrinsic to M. We cannot take the class $P_{t_1,M}$ to be infinite; for the proof that each inductive test will terminate depends on it being finite. Also there is no *effective* way to divide all hypotheses into successive finite classes $P_{t_1,M}, P_{t_2,M}, P_{t_3,M}, \ldots$ in such a way that a) every class contains a finite number of mutually incompatible effective hypotheses, and b) every effective hypothesis is in some class.[38] M cannot be transformed into an effective method for selecting the best hypothesis from the class of *all* hypotheses expressible in the language (as opposed to the hypotheses in a given finite class). Thus science *cannot* be done by a moron; or not if the moron relies on the method M, at any rate.

The situation is even more interesting if one uses the Kemeny method. For the simplicity of hypotheses with the same observational consequences may vary greatly (even by the "symbol count" criterion).

[36]*The Continuum of Inductive Methods* (Chicago: Univ. of Chicago Press, 1952), 48. For a lengthier discussion of the plausibility of making d.c. dependent on κ, see "On the Application of Inductive Logic," *Philosophy and Phenomenological Research*, VIII, 133-148; particularly 144.

[37]Readers familiar with Rosenbloom's *Elements of Mathematical Logic* (Dover, 1950), will recognize the identification of the computer with the "moron".

[38]This is a consequence of Gödel's theorem, as remarked above (n. 28).

A way of putting it is this: call two hypotheses "essentially the same" if they have the same observational consequences. Then the relative simplicity of hypotheses that are "essentially the same" may vary greatly depending on the language in which they are couched. (For instance, Craig has shown[39] that every hypothesis can be "essentially" expressed in O, in this sense; but the axiomatization required is infinite if the original hypothesis contains theoretical terms, so there would be infinite complexity.) Thus the hypothesis a scientist will accept, using a method which includes a simplicity order, will depend not only on what hypotheses he has been able to think of, but on the theoretical language he has constructed for the expression of those hypotheses. Skill in constructing theories within a language and skill in constructing theoretical languages both make a difference in prediction.

IX

There are respects in which all the methods we have considered are radically oversimplified: for instance, none takes account of the reliability of the data. Thus, Rule I of method M is unreasonable unless we suppose that instrumental error can be neglected.[40] It would be foolish, in actual practice, to reject a hypothesis because it leads to exactly one false prediction; we would rather be inclined to suppose that the prediction might not really have been false, and that our instruments may have deceived us. Again there is the problem of assigning a proper weight to *variety* of evidence, which has been emphasized by Nagel. But my purpose here has not been to consider all the problems which might be raised. Rather the intention has been to follow through one line of inquiry: namely, to see what features of the scientific method can be represented by the method M and related methods, and to show that crucial features cannot be represented by any "measure function".

Again, I have not attempted to do any philosophic "therapy"; to say what, in my opinion, are the mistaken conceptions lying at the base of the attempt to resuscitate the "logical probability" concept. But one such should be clear from the foregoing discussion. The assumption is made, in all work on "degree of confirmation", that there is such a thing as a "fair betting quotient", that is, the odds that an ideal judge would assign if he were asked to make a fair bet on a given prediction. More precisely, the assumption is that *fair odds* must exist in any evidential situation, and *depend only on the evidence*. That they must depend on the evidence is clear; the odds we should assign to the prediction "the next thing will be red" would intuitively be quite different (in the

[39]"Replacement of Auxiliary Expressions," *Philosophical Review*, LXV, 38-53.

[40]I am indebted to E. Putnam for pointing this out.

absence of theory!) if 50% of the individuals examined have been red, and if all have been. But, I do not believe that there exists an abstract "fairness of odds" independent of *the theories available to the bettors.* To suppose that there does is to suppose that one can define the best bet *assuming that the bettors consider the best possible theory;* or (what amounts to the same thing) assuming they consider all possible theories.

Such a concept appears to be utterly fantastic from the standpoint of the actual inductive situation; hence it is not surprising that any definition would have to be so non-effective as not to be of any use to anybody.

Since this assumption underlies the work of De Finetti,[41] and the "subjective probability" approach of Savage,[42] I am inclined to reject all of these approaches. Instead of considering science as a monstrous plan for "making book", depending on what one experiences, I suggest that we should take the view that science is a method or possibly a collection of methods for *selecting a hypothesis,* assuming languages to be given and hypotheses to be proposed. Such a view seems better to accord with the importance of the hypothetico-deductive method in science, which all investigators have come to stress more and more in recent years.

HILARY PUTNAM

DEPARTMENT OF PHILOSOPHY
PRINCETON UNIVERSITY

[41]"Sul significato suggestivo della probabilita," *Fundamenta mathematicae*, XVII, 298-329.

[42]*The Foundations of Statistics* (New York: Wiley, 1954).