WESLEY C. SALMON

HANS REICHENBACH'S VINDICATION OF INDUCTION

Hans Reichenbach believed that he had solved Hume's problem of the justification of induction, but his arguments have not proved persuasive to most other philosophers. The majority of those who addressed the problem held, for one reason or another, that it is a pseudo-problem. In a number of articles during the 1950s and 1960s I tried to refute this position.[1] I still believe it is incorrect – that the problem of justification of induction is a genuine and profoundly important problem – but I shall not rehearse that issue here. In this article I shall first discuss Reichenbach's justification and the problems confronting it. I shall then consider other attempts at vindication that in one way or another pursue a similar goal. In the end, I shall maintain, Reichenbach's program can succeed, given certain additional considerations, a crucial one of which is found in his own writings.

1. REICHENBACH'S JUSTIFICATION

One of the key steps in Reichenbach's solution to the problem of induction was his recognition that what is needed is the justification of a rule, not the proof of a factual proposition such as the uniformity of nature.[2] He realized, in addition, that it is impossible to justify the rule in question by proving that it will always, or even sometimes, yield true conclusions, given true premises. He argued, nevertheless, that his *rule of induction*[3] should be adopted because one has everything to gain and nothing to lose by employing it. Although successful prediction cannot be guaranteed, he argued, if any method works the rule of induction will work. The structure of his argument is rather similar to Pascal's wager; it seeks not to justify belief in a proposition but rather to justify a practice. For that reason, Reichenbach characterized his justification as *pragmatic*.

Herbert Feigl, who was strongly sympathetic to Reichenbach's approach, drew a basic distinction between two kinds of justification, namely, *validation* and *vindication* (1950). One validates a rule or proposition by deriving it from a more fundamental principle. In deductive

Erkenntnis **35**: 99–122, 1991.

TABLE I

	Nature is uniform	Nature is not uniform
Induction is used	Success	Failure
Induction is not used	Success or failure	Failure

logic, for example, the rule of conditional proof is validated by means of the deduction theorem, which shows that any conclusion deduced through use of conditional proof can be deduced without appeal to that rule. Obviously, the most fundamental rules cannot be validated. If they can be justified at all, it must be by vindication. One vindicates a rule by showing that its use is well suited to the achievement of some aim we have. The rules of propositional logic can be vindicated by showing them to be truth-preserving. Their use fulfills our desire to avoid deriving false conclusions from true premises.

Since Reichenbach's justification of induction does not consist in the derivation of the rule of induction from a more fundamental rule, his pragmatic justification qualifies as a vindication. Our goal is to make correct predictions of future events (or more generally, to make correct inferences from observed phenomena to as yet unobserved phenomena). He argues roughly as follows (see Table I).[4] As Hume has shown, we cannot know whether nature is uniform or not. If we are fortunate and nature is uniform then the use of induction will fulfill our goal. This does not mean that every prediction will be correct, but we will be successful on the whole. If we are unlucky and nature turns out not to be uniform we may fail miserably. Perhaps there will be a few lucky guesses, but overall we will suffer failure in our attempts at prediction.

Suppose, instead, that we do not use induction. This might happen in either of two ways. In the first place, we might simply refuse to make any inferences at all. This alternative obviously fails whether nature is uniform or not. Nothing ventured, nothing gained. In the second place, we might try some different method for making predictions, for example, making wild guesses, consulting a crystal gazer, or believing what is found in Chinese fortune cookies. If nature exhibits uniformities, any of these methods might work, but there is no guarantee of success. If nature is uniform, then, it seems clear that induction is the best method, for it is bound to work on the whole, whereas the others may or may not be successful.

But what if nature is not uniform? In that case, no method can yield

consistent success, for the *consistent* success of any noninductive method would be a uniformity, contary to the hypothesis that nature is not uniform. Moreover, if such a uniformity did transpire, that regularity could be exploited inductively. If, for example, the crystal gazer achieves a good predictive record – consistently predicting the outcomes of future horse races – we could use induction to predict that such predictions will continue to be accurate. It would *not* be foolish to use such information for the placing of bets. Therefore, Reichenbach concludes, if nature is not uniform the user of inductive reasoning is no worse off than anyone who uses any other method. Consequently, we have everything to gain and nothing to lose by employing inductive methods.

Although the foregoing argument gives some of the flavor of Reichenbach's vindication, it suffers from excessive vagueness. How uniform must nature be to qualify as uniform? What sorts of uniformities are important? What exactly do we mean by "success" and "failure"? How are the various kinds of methods to be precisely characterized?

Reichenbach was well aware of these difficulties, and he offered a more precise version of the foregoing pragmatic argument (1949, Sec. 91). Whether he knew of Bishop Butler's famous aphorism or not, he concurred fully that *probability is the very guide of life*. The fundamental goal in terms of which induction is to be vindicated is the acquiring of knowledge of probabilities. Inasmuch as he advocated the limiting frequency interpretation of probability, the goal is the ascertainment of the values of limits of relative frequencies in potentially infinite sequences of events. Given any such sequence of events, and any attribute that can meaningfully be predicated of its members, we can attempt to ascertain its limiting frequency in that sequence. There is, of course, no a priori guarantee that the limit in question exists. If it does, in fact, exist, then nature is uniform in that respect. The probability constitutes a statistical regularity. As we try to establish the value of the probability we are concerned with precisely that uniformity. If the frequency with which the attribute occurs in the sequence does not converge to any limit, nature fails to be uniform in the pertinent respect.

In order to make the notion of using induction more precise, Reichenbach offers his *rule of induction*. which may be formulated as follows:

> If an observed initial section consisting of *n* members of a sequence of As contains *m* elements with the attribute B

TABLE II

	Sequence has a limit	Sequence has no limit
Induction is used	Success	Failure
Induction is not used	Success or failure	Failure

POSIT THAT the limit of the relative frequency of B in A lies within the interval $m/n \pm \delta$.

It is to be understood that nothing is known about the probability of B within A beyond the observed frequency of B in the specified initial section of the sequence. Moreover, the rule is to be used repeatedly as larger and larger initial sections of the sequence are observed. The size of δ is determined by pragmatic aspects of the context – the degree of precision required in that situation. Looking at the problem in terms of these more precise concepts, we may offer a refined version of Table I (see Table II).

Given the precise formulation of the rule of induction, Reichenbach points out, it is an immediate consequence of the mathematical definition of the limit of a sequence that, if the limit exists, repeated application of that rule will lead sooner or later to posits that are accurate to any desired degree of approximation; moreover, further posits, based upon larger and larger observed initial sections of that sequence, will continue to be at least that accurate. Thus, he argues, it is an analytic truth that *if success in ascertaining a limit is possible his rule of induction will yield success*. If no limit exists, obviously no method will succeed. This is his argument to establish the first row of Table II.

Various objections have been raised against this part of Reichenbach's argument. Some authors have questioned the presumption that ascertainment of limits of relative frequencies can seriously be considered a goal of human inquiry.[5] The first question to ask, it seems to me, is whether knowledge of the objective physical probability relations that obtain in the world is an aim of our endeavor. With Bishop Butler's aphorism in mind, I think the answer must be affirmative. The next question concerns the appropriate interpretation of physical probabilities. Reichenbach, of course, adopted the limiting frequency interpretation. Nowadays the so-called propensity interpretation is more popular,

but it has severe drawbacks, including the fact that it is not an admissible interpretation of the probability calculus.[6] Since the controversy over these interpretations is beyond the scope of this paper, I shall adopt Reichenbach's position for purposes of argument herein.

Another objection arises from the fact that, with respect to any given sequence, we cannot predict how far we must go before reaching posits that are accurate to any particular degree. Moreover, even if we have arrived at that point, we have no way of knowing that we are there. For all we can ever know, a sequence might not begin to converge until an initial section containing billions upon billions of members had elapsed. One can only hope that convergence occurs reasonably rapidly, for convergence that occurs too slowly for the benefit of human investigators is no better than complete lack of convergence. Reichenbach, and John Venn[7] before him, stated clearly that the infinite sequence is a mathematical idealization of a very large finite class, much as the Euclidean plane is an idealization of a large surface that is approximately flat. Reichenbach (1949, 347–48, 447–48) called specific attention to this issue by introducing the concept of the *practical limit*.

There are obviously many ways of attempting to ascertain the value of the limiting frequency of a sequence – assuming it exists – that may or may not yield approximately correct results. One could, for instance, write a lot of rational fractions between zero and one on slips of paper, place them in a hat, mix them well, and draw one out. One could then posit that the number drawn is within δ of the actual limit. Still assuming that the sequence has a limit, there is no proof that this method will not work, but there is no proof that it will. Another example is the *counter-inductive rule* – discussed, but not advocated, by Max Black (1954) – according to which one posits that the relative frequency of non-B in the observed sample, $(n - m)/n$, is approximately equal to the limiting frequency. In any case in which the limiting frequency is not near 1/2, the persistent use of this method is guaranteed to yield posits that do not approximate the limiting frequency. In my view, Reichenbach's convergence argument is sufficient to show the superiority of his rule of induction to either of these noninductive methods. The counter-inductive rule can, incidentally, be rejected on other grounds as well, for it yields radically incoherent sets of probability values.[8]

Unfortunately for Reichenbach's attempted pragmatic justification, there exists a nondenumerably infinite set of rules each of which shares

TABLE III

	Sequence has a limit	Sequence has no limit
Rule of induction used	Success	Failure
Other asymptotic rule used	Success	Failure
Nonasymptotic method used	Success or failure	Failure

the convergence property of his rule of induction. He was fully aware of this set; he called them *asymptotic rules*. These rules can be characterized as follows:

> If an observed initial section consisting of n members of a sequence of As contains m elements with the attribute B **POSIT THAT** the limit of the relative frequency of B in A lies within the interval $(m/n + c_n) \pm \delta$, where $c_n \rightarrow 0$ as $n \rightarrow \infty$.

We may think of c_n as a 'corrective term' that modifies the observed frequency for the sake of a 'better posit'. Because the sequence of posits endorsed by any rule of this type converges to the sequence of posits endorsed by the rule of induction, both sequences will converge to the actual limiting frequency provided that such a limit exists. The situation is shown in Table III.

In view of this circumstance, Reichenbach cannot claim that his rule of induction is the *only* rule that is guaranteed to succeed if any method can succeed; any of the asymptotic rules will succeed if success is possible. He realized, moreover, that it is impossible to show that his rule of induction will yield faster convergence than any others of the asymptotic rules.[9] If his rule of induction is to be justified, some adequate reason must be given to prefer it to the other asymptotic rules. He was fully aware of that fact, and he offered a justification that he considered sufficient. From the entire class of asymptotic rules, he said, we select the rule of induction on grounds of descriptive simplicity (1949, p. 447).

Reichenbach had distinguished two types of simplicity: inductive and descriptive (1938, Sec. 42). Suppose we have two hypotheses, one simpler, the other more complex, both of which are compatible with all of our observations up to the present. They are, however, factually distinct; indeed, they are mutually incompatible. Faced with a choice between them, we sometimes select the simpler because we believe it

is more likely to be true. This is *inductive simplicity*. In such cases there is evidence, not in our possession as yet, to undermine at least one of them. They are *not* observationally equivalent.

Descriptive simplicity comes into play when we have two or more theories that *are* observationally equivalent. The most vivid example arises in his theory of space and geometry. He maintained that the physical space of our universe can be described equally adequately by Euclidean geometry augmented by a suitable set of universal forces or by a non-Euclidean geometry without universal forces. Given such a pair of descriptions, either both are true or both are false. There is no empirical or factual difference between them. In company with Einstein, he maintains, we choose the description that eschews universal forces, but it is a matter of aesthetics or intellectual economy. Truth or falsity does not enter into the choice.

Reichenbach claimed that, since all of the asymptotic rules, including his rule of induction, converge to the same limits in the long run, they are empirically equivalent. We are free to choose the rule of induction because it is the descriptively simplest rule in the set of asymptotic rules. It appears, however, that a serious lapse has occurred in his argument. We noted at the outset that he placed great emphasis upon the fact that what stands in need of justification is not a statement but, rather, a rule. Clearly, both inductive and descriptive simplicity apply only to selections among statements, not to selections among rules. If Reichenbach's claim about the equivalence of all asymptotic rules has any merit at all, it must refer to equivalence "in the long run". According to a famous aphorism of J. M. Keynes, *in the long run we will all be dead*. If we look at the asymptotic rules in terms of human application, they are as radically nonequivalent as any rules could be.

Reichenbach's attempt to vindicate his rule of induction cannot be considered successful.

2. SALMON'S ATTEMPT TO FILL THE GAP

When I realized that the counter-inductive rule runs into incoherence, it occurred to me that the same consideration imposed a serious constraint on Reichenbach's 'corrective term' c_n. Any asymptotic rule in which c_n is a function of n alone will lead to the same sort of incoherence (Salmon, 1956). After imposing a suitable coherence condition – which I first called *regularity*, but later referred to as *normalizing conditions*

– on asymptotic rules, I was able to show that, even after disqualifying the irregular asymptotic rules, an infinity remained. This collection is so broad that it contains rules to license any posit whatever regarding limiting frequencies. More explicitly, one may arbitrarily choose any positive integer n as the size of the observed sample (initial section) of the sequence A, any nonnegative integer $m \leq n$ to represent the number of elements of the sample having the property B, and any real number $p(0 \leq p \leq 1)$ as the value of the limit of the frequency of B in A. Then, there exists among the regular asymptotic rules some rule that directs one to posit p as the limit on the basis of the observed frequency m/n (Salmon, 1957a). If, for example, a million members of A have been observed, all of which have possessed the attribute B, there is an asymptotic rule that licenses the posit that the limiting frequency of B within A is zero. Although the set of regular asymptotic rules is convergent, it is *nonuniformly* convergent. That means that there is no finite integer N representing a sample size at which all of the regular asymptotic rules begin to converge. For purposes of human prediction, this set of rules is as divergent – as empirically nonequivalent – as it could possibly be. Descriptive simplicity is not a suitable criterion for making a selection from that class. If simplicity is to be invoked for purposes of justifying an inductive rule, it must be a different sort of simplicity. Since it would be applicable to rules, I suggest that it be called *methodological simplicity*. I shall return to that concept in Section 4.

Having noted the foregoing difficulty regarding regular asymptotic rules, and having taken into account the fact that Reichenbach's 'corrective term' c_n cannot be a function of n alone, I began looking at other variables on which it might depend. The search was facilitated by consideration of Carnap's continuum of inductive methods (1952). All of these methods, with the exception of the straight rule, are dependent upon the language in which the evidence and hypotheses are stated. It seemed to me at the time – and it still does – that our inductive rules should be invariant across languages. For example, a switch from metric units to English units should make no difference to the inductive relationship between the hypothesis and the experiment. Similarly, if a hypothesis can be articulated in German and in English, and if the evidence can also be described in both languages, then the degree to which the hypothesis is supported or undermined should be the same

in both cases. It would be strange, indeed, if a native speaker of English – displeased with the outcome of an experiment with respect to a hypothesis – could find a more pleasing inductive result by translating the experimental outcome and the hypothesis into German.

In an attempt to bring these considerations to bear on the problem at hand, I imposed – in addition to the regularity requirement or normalizing conditions – a *requirement of linguistic invariance*. I then offered a rather tedious mathematical argument to show that these two requirements constrained c_n to be identically zero, thus establishing Reichenbach's rule of induction as the only acceptable rule (1963a). Regrettably – at least from my standpoint – the argument was flawed.[10] As Ian Hacking pointed out, I had failed to take into account the fact that we often know, not only the relative frequency of the attribute in the sample, but also the order in which the Bs occur within the sample. It is possible to construct asymptotic rules that violate neither of my requirements but are not identical with Reichenbach's rule of induction (Hacking, 1968, esp. pp. 57–59).

Hacking did more; he proved a general theorem, showing that the fulfillment of three conditions – (1) *consistency*, (2) *symmetry*, and (3) *invariance* – is necessary and sufficient to select the Reichenbach rule of induction. The first condition, consistency, is unproblematic from my standpoint. It is closely related to the normalizing conditions (regularity requirement). It is somewhat stronger, but it is demonstrably satisfied by limits of relative frequencies. The third condition, invariance, is related to my criterion of linguistic invariance, but is considerably stronger. It includes a condition of *statistical invariance* over and above linguistic invariance. In a commentary on Carnap's inductive logic, I had already argued for Hacking's stronger condition (1967). As we shall see, it still appears to be defensible.

Hacking's second condition, symmetry, is closely related to what personalists call *exchangeability*. It says, in effect, that for a given relative frequency in an initial section of a sequence, the posited value for the limiting frequency must be the same regardless of the order in which the members of the sample occur. I shall return to this requirement in Sections 6–7. For 25 years, it has seemed to me an insurmountable obstacle to the kind of vindication I had hoped to provide.

My attempt to vindicate Reichenbach's rule of induction cannot be considered successful.

3. SELLARS: INDUCTION AS VINDICATION

More than a quarter of a century ago, Wilfrid Sellars published "Induction as Vindication" (1964) in which he offered his resolution of the problem of justifying induction. At the very heart of his argument, and of the article (paragraph 55), there occurs a brief subsidiary argument that deserves careful scrutiny.

Let me begin by setting the stage. The context in which the crucial argument occurs is one in which Sellars is concerned with inferences from a finite observed sample K of population X to an unobserved finite sample ΔK of that population with respect to an attribute Y. The practice to be justified is to infer that the relative frequency of Y in ΔK is approximately the same as it is in K. Sellars maintains quite explicitly that he is trying to justify a practice rather than attempting to establish any sort of lawful statistical generalization. More precisely, he wants to establish "the state of being able to draw inferences concerning the composition with respect to a given property Y of unexamined finite samples (ΔK) of a kind, X, in a way which also provides an explanatory account of the composition with respect to Y of the total examined sample, K, of X" (1964, p. 215). As I understand it, Sellars is making a distinction between (1) simply drawing an inference from a frequency in an observed sample to the frequency in an unobserved sample of the same population and (2) drawing an inference from something that *explains* the frequency in the observed sample to the frequency in an unobserved sample of the same population. We can now consider the crucial argument; I shall quote it in full:

> . . . to give an explanatory account of the composition of the class K of *examined* Xs one must, logically, assert that the composition in question is the most statistically probable composition on the basis of the finite population (P) of Xs which are known to exist but of which only the members of K have been examined. If we take the finite unexamined remainder of P as ΔK, so that
>
> $$P = K + \Delta K$$
>
> then, since the statistically probable composition of a random sample approximates that of the population, the above condition logically requires the acceptance of
>
> $$rf(Y, P) \approx rf(Y, \Delta K) \quad [\text{'}\approx\text{' means 'approximates to'}]$$
>
> which, in turn, logically requires the acceptance of
>
> $$rf(Y, \Delta K) \approx rf(Y, K)$$
>
> and hence that the proportion [mentioned above] be specified as (approximately) the proportion of Ys in the examined sample K. [p. 216]

This argument involves, I think, a serious ambiguity.

When Sellars refers to "the statistically probable composition of a random sample", he seems to be referring to the combinatorial fact that, *given a finite population* P, in most subsets K (of reasonable size m) of P the relative frequency of Y in K is approximately equal to the relative frequency of Y in P. This fact is not in dispute; it is a consequence of the Bernoulli theorem. If you pick *randomly* one subset K from the set of all subsets of specified size of P, you will probably get a representative sample. (We shall consider the meaning of the concept of randomness and its role in Sellars's arguments below.)

It is crucial to realize, however, that the foregoing fact does not imply that, *given an observed set* K *of size m containing n elements with property* Y, it very probably comes from a population P in which the relative frequency is approximately *n*/*m*. This probability must be computed by means of Bayes's theorem, and to do so requires prior probabilities. No conclusion can be drawn concerning the relative frequency of Y in P unless we are given a distribution of prior probabilities for the logically possible relative frequencies of that attribute in that population.

This point can be illustrated by a simple example. Suppose we have a sack containing a million pennies, all but one of which are standard fair coins. The exceptional one is two-headed. Someone draws a coin at random from this sack and proceeds to flip it ten thousand times. We have no opportunity to examine the coin, nor to witness all of the tosses, but we do have the opportunity to learn the outcomes of ten tosses selected at random. All of the ten tosses resulted in heads. It would obviously be unwarranted to infer that the relative frequency of heads in the set of ten thousand tosses is approximately one. It would obviously be unwarranted to infer that the relative frequency of heads in another (nonoverlapping) randomly selected subset of ten would be approximately one.

Now Sellars has, in the foregoing argument, insisted on the importance of being able to *explain* the frequency in the observed sample, and the principle of explanation he adopts is essentially of a maximum likelihood type. Recall his assertion that "to give an explanatory account of the composition of the class K of *examined* Xs one must, logically, assert that the composition in question is the most statistically probable composition on the basis of the finite population (P)". In spite of Sellars's use of such expressions as "must, logically", maximum likelihood explanations are not uncontroversial. For instance, it would be rash in the extreme to offer as an explanation of the ten heads in the foregoing example that the coin being tossed has two heads.[11] A much more plausible explanation of the set of ten heads in this context would be that it was a fairly improbable (1/1024) chance occurrence that came about as a result of flipping a fair coin. And, from a practical standpoint, it would surely be unwise to take a number close to one as a betting quotient for purposes of wagering that the next toss will be a head.

In the above-quoted argument, Sellars appears to make use of two transitivity relations, both of which are illegitimate. In the first place, he claims that we are logically required to accept

(1) $$rf(\mathrm{Y}, \Delta\mathrm{K}) \approx rf(Y, P)$$

"which, in turn, logically requires the acceptance of

(2) $$rf(\mathrm{Y}, \Delta\mathrm{K}) \approx rf(Y, K)\text{''}.$$

How could (1) logically require the acceptance of (2)? As far as I can see, the justification for accepting the second approximation, given the first, is the assumption that

(3) $$rf(\mathrm{Y}, \mathrm{P}) \approx rf(\mathrm{Y}, \mathrm{K}).$$ [12]

If "$\approx$" designated a transitive relation, (2) would follow from (1) and (3); however, it is well-known that relations of this sort (approximate matching relations) are not transitive (Salmon, 1984, pp. 78–79). But this misuse of transitivity is relatively innocuous compared to the second.[13]

Sellars is not claiming, after all, that the frequency makeup of the sample *must* approximate that of the population, even that of population P as he defines it. In order to make his argument go through, Sellars is apparently relying on something like the transitivity of probabilistic support. The argument would seem to go as follows: Given that the frequency of Y in K = r, it is highly probable that the frequency of Y in P = $r \pm \delta$, and given that the frequency of Y in P = $r \pm \delta$, it is highly probable that the frequency in ΔK = $r \pm \epsilon$ (for suitably chosen δ and ϵ).[14] My favorite counterexample to probabilistic transitivity at the time

Sellars's article was published was this: Given that x is a scientist, it is highly probable that x is alive,[15] and given that x is alive, it is highly probable that x is a micro-organism. The reader can complete the syllogism.

In subsequent paragraphs (81–82) Sellars addresses the issue of inferences from samples to populations: ". . . given an identificatory ordering of the classes K_i which are members of $^2\mathrm{K}$ [the class of all subclasses of P having m members]

$$\mathrm{K}_1, \mathrm{K}_2, \ldots, \mathrm{K}_\mu$$

formulating the class of questions

Does K_i match the B composition of P within ϵ?

and answering them all in the affirmative, a majority of the answers, for properly chosen ϵ, will be true". This much is certainly correct (at least for samples and populations of reasonable size). But Sellars continues, ". . . we can argue

I shall accept all the affirmative answers to the question 'Does K_i match P in B within ϵ?'
Therefore since S is a random sample of P having m members, it is identical with one of the members K_i of $^2\mathrm{K}$
Therefore I shall accept 'S matches [P] in B within ϵ'
But the B composition of S is n/m
Therefore I shall accept 'the B composition of P is within ϵ of m/n'".

Seductive as this reasoning may be, it is unsound. Sellars begins with the true claim that, if we examine *each and every m*-size sample of P *exactly once*, we shall be right in the majority of cases if we assert that the sample approximately matches the population.[16] But that is not the situation in which we find ourselves for the most part. We ordinarily observe one or a few samples of the given population. If we are to have any basis for claiming that these observed frequencies match the population, we must, as Sellars seems to realize, assume that the observed samples are random samples.

A good deal of trouble is caused in Sellars's arguments by his use of the term "random". Unfortunately, it is seriously ambiguous. The meaning Sellars adopts, in a footnote to the quoted passage, is "a sample concerning which *nothing else is known* relevant to its matching o[r] not matching the composition of P". Another standard meaning is *a sample selected by a method under which all possible samples have an* *equal probability of being drawn*. Now Sellars's discussion of the numbers of samples of a given composition from a given population is pertinent if the proportion of such samples in the population reflects the frequency with which they are drawn. Ignorance of bias is not, however, an adequate basis for concluding that there is absence of bias. For example, if the members of the population are presented sequentially, and if the probability that an element possesses an attribute is *not* independent of the possession of that attribute by its predecessor, then sequential sampling is not random. Consider the probability that a day is warm and sunny. In Pittsburgh, for example, it is far more probable that a warm sunny day follows a warm sunny day than that it follows a cold snowy day. You will not get a random sample with respect to warmth and sunshine by visiting Pittsburgh for a month, whichever month you choose. The samples nature gives us are often far from random; frequently we must contrive cleverly to find random samples.

The randomness issue arises again near the conclusion of Sellars's essay. In paragraph 86 he deals with inferences regarding an attribute B from the composition of a finite population P to the composition of a *random sample* S drawn from that population. Citing the combinatorial facts discussed above, he concludes, "I shall accept 'S matches the B composition of P within ϵ'". The problem is, how are we to know whether an actual sample drawn from an actual population is random or not? Sellars's answer seems to be that we are entitled to consider any sample random if we have no knowledge to the contrary (1964, p. 225, n. 16). This answer constitutes an appeal to Laplace's *principle of indifference*: two events are equally probable if we have no reason to prefer one to the other. In this context Sellars is saying that any particular m-member sample of the population in question is just as likely to be drawn as any other sample of the same size. The principle of indifference has been widely criticized, and I am convinced that it cannot be sustained in the face of the well-known objections, the chief among which is that its unbridled use leads to outright logical contradiction (Salmon, 1967a, pp. 65–68). In his theory of confirmation, Carnap attempted to preserve what he considered "the valid core" of the principle, but even a limited adoption seems to me to entail intolerable a priorism. I cannot see that Sellars has escaped these well-known difficulties.

Sellars's attempt to show that induction is vindication cannot be considered successful.

4. CLENDINNEN'S APPEAL TO SIMPLICITY

In *Experience and Prediction* (1938, p. 355) Reichenbach remarked that the adoption of any asymptotic method in which the 'corrective term' c_n is not identically zero would be *arbitrary*, but he did not elaborate.[17] Instead, he argued that the adoption of any such rule would carry a greater risk of error than would adoption of his rule of induction. Later, in his *Theory of Probability*, he saw that this argument is unfounded, and he abandoned it in favor of an appeal to descriptive simplicity.

In his (1982), F. John Clendinnen develops an argument for a vindication of induction that hinges crucially on the notion of arbitrariness. The point of departure is a claim that *it is irrational to believe in a proposition that we have arrived at by guessing*. He readily concedes that there are circumstances in which guessing is perfectly appropriate. Suppose I am on my way to Damascus, and I come to a fork in the road. I have no idea which fork leads to Damascus – no evidence to which I can appeal. Unfortunately, there are no inveterate liars, truth-tellers, or anyone else for me to interrogate. If I stay at the fork in the road I am likely to die of thirst before anyone shows up. I must choose one way or the other. I could flip a coin or just make a wild guess. That would be better than no choice at all, for the arbitrary choice gives me some chance of getting to Damascus, while the absence of choice will surely prevent me from reaching my goal. It would be rational, in these circumstances, to make an arbitrary choice, and to *hope* that it is the right choice. It would, however, be irrational to *believe* that the road I choose will lead to Damascus, or even to believe it *more likely* than the road I reject to lead to Damascus.[18]

How does all of this apply to the problem of induction? Clendinnen offers the following suggestion:

> In outline the argument is that non-inductively based prediction is, if not itself a mere guess, based on a procedure which includes a purely arbitrary step. As such, non-inductive predictions are no better than guesses. And it is irrational to place any more reliance on one guess than any of the other guesses which might have been made (p. 3).

Suppose that we have observed an initial section consisting of n members of a sequence A and that we have found m of them to be B. Suppose also that we have no further evidence concerning the probability that an A is a B. If we use Reichenbach's rule of induction we will posit that the limit of the relative frequency is approximately equal to m/n. If, instead, we use a different asymptotic rule, we must add a (possibly negative) quantity c_n to the observed frequency. As we have already noticed, there is a vast plethora of asymptotic rules which furnish a superabundance of 'corrective terms'; hence, the decision to use one value instead of another is simply a blind guess. To guess when you don't have to is irrational.

It might be objected that even the use of Reichenbach's rule of induction is arbitrary. As Hume's skeptical arguments show, it might be said, any prediction we happen to make is just as arbitrary as any other. Clendinnen would deny this claim. Reichenbach's convergence argument shows that the frequency in the observed initial section of a sequence is relevant evidence regarding the probability of a given attribute in that sequence.[19] We are, after all, investigating the behavior of a sequence of relative frequencies, and the observed frequencies are constitutive of that very sequence. Rationality requires that we utilize all of the available relevant evidence; it requires, further, that we not introduce unnecessary arbitrary elements. The quantities we get by adding a 'corrective term' to the observed frequency may or may not occur anywhere in the sequence of relative frequencies. Hence, to use Reichenbach's rule of induction constitutes use of the available evidence. However, to modify that rule by adding a 'corrective term' is to add an element that is arbitrary. To a sound inductive procedure it adds an irrational guess.

The outcome of Clendinnen's argument is a principle that warrants designation as *the principle of methodological simplicity*:

> Adopt the simplest system of predicting rules which are compatible with, and exemplified in, the set of known facts (p. 20).

It is manifestly more appropriate to appeal to this concept of simplicity, which pertains to the selection of rules, than to appeal either to Reichenbach's concept of descriptive simplicity or to his concept of inductive simplicity, both of which pertain to the selection of statements or propositions. Moreover, Clendinnen's principle, prohibiting reliance on guesses, seems to me eminently sound.

Although Clendinnen's essay does not, in my opinion, furnish a fully articulated vindication of induction, it does provide a significant step in the right direction.

5. CARNAP'S REJECTION OF THE STRAIGHT RULE

Given what strikes many philosophers as the reasonableness of the straight rule, it is incumbent upon us to consider Carnap's reasons for rejecting it. His argument is actually quite simple. In any case in which all observed As are B, or in which no observed As are B, Reichenbach's rule of induction would have us posit that the limit of the relative frequency, and hence the probability, is one or zero respectively. In Carnap's view, a major function of probabilities is to serve as betting quotients, and it is easy to see that zero and one are unsuitable values for them. A betting quotient of one would mandate a wager of a million dollars to a penny on the next event in the sequence. Such bets would obviously be absurd.

The source of Carnap's difficulty is his claim that statements of probability$_1$ (degree of confirmation) are, if true, analytic, or, if false, contradictory. One can be as certain of the correctness of the value as one is of any other truths of arithmetic. When one has a probability value of unity, one can be certain that it is the correct betting quotient. In consequence, to protect against foolhardy betting quotients, Carnap builds in what might be called a "safety factor" which keeps various probabilities from assuming the offending values. It is the safety factor rather than an arbitrary guess that makes Carnap's probabilities differ from those the straight rule would give. Users of the straight rule – Reichenbach's rule of induction – are willing to posit probabilities having the extreme values, but they recognize full well that they *cannot* be confident that the posited values are truly the limiting values of the relative frequencies. A posit is *by defnition* something about which we cannot be certain. Rather than building the safety factor into their inductive rules, they would offer such practical advice as to avoid making large bets at unfavorable odds on the basis of probabilities whose values are not known with great confidence.

6. TWO REICHENBACHIAN DISTINCTIONS

In Section 2 I mentioned the three criteria – consistency, symmetry, and invariance – shown by Hacking to be necessary and sufficient for justification of Reichenbach's rule of induction. Consistency poses no problems; it amounts to probabilistic coherence and is demonstrably satisfied by limits of relative frequencies. The invariance requirement, as I remarked above, corresponds to two invariance criteria I have proposed, namely, linguistic invariance and statistical invariance.[20] These two invariance requirements are, I think, fully justified by Clendinnen's principle of methodological simplicity; a violation of either would be a case of proscribed arbitrariness. The criterion of symmetry is another story.

The reason symmetry is required, according to Hacking, is that we often have knowledge, not only of the relative frequency of a given attribute in a given sample, but also of the order in which the elements occur. Symmetry requires that the order have nothing to do with our posit regarding the limiting frequency. One can devise asymptotic rules that satisfy both consistency and invariance, as Hacking has shown (1968, pp. 50–51), but that differ from Reichenbach's rule of induction. This particular example can be ruled out by Clendinnen's methodological simplicity principle, but I do not think all rules that violate symmetry can be handled quite that easily.

Depending on one's philosophical proclivities, there are various ways of dealing with the symmetry issue. A subjective Bayesian can simply invoke exchangeability, meaning that his or her subjective probability with respect to an A being a B would not be any different regardless of the order of B and non-B in the sample. Reichenbach, as a dedicated frequentist, would not have admitted subjective probabilities.

Another approach would be to apply statistical estimation methods. In stating his rule of induction, Reichenbach consistently used the term "posit", meaning something like a bet or a wager. It is not entirely clear whether he intended this term in the sense of an inference or in the sense of an estimate, but the fact that he included a margin of error δ in the formulation of the rule suggests that it can properly be construed as an estimate. On this construal, we can circumvent the symmetry requirement by adopting the observed frequency S_n as our estimator. It has the sorts of virtues we would seek; it is convergent, additive, unbiased, and has minimal variance. The observed frequency provides a sufficient statistic; the order of occurrence of the constituents is irrelevant.

Reichenbach's objection to this approach would be, I believe, to call attention to the fact that it requires factual assumptions about the

nature of the population under investigation.[21] In his treatment of induction, he distinguished *primitive knowledge* from *advanced knowledge*. In primitive knowledge we have no previous inductive results upon which to depend; in advanced knowledge the results of previous inductive inferences are available. His rule of induction is a method for primitive knowledge, and this is what he was attempting to justify. Thus, he would argue, since we have no results of previous inductions to establish these factual assumptions, we are not entitled to make them.

From a *psychological* standpoint, the distinction between primitive and advanced knowledge is probably unfounded; it is doubtful that any such psychological state as primitive knowledge exists. We can, nevertheless, make a useful *logical* distinction between primitive inductive rules and advanced inductive methods. Primitive rules are employed where no previous inductive results are available – the permissible inputs for such rules are statements of observed facts, but no inductive generalizations are allowed. According to Reichenbach, all inductions, primitive or advanced, are carried out by means of the rule of induction and the probability calculus. The theorems of Bernoulli and Bayes play prominent parts in the development of advanced induction. On the frequency interpretation of probability, he argues, all of the axioms and theorems are logically necessary. The rule of induction provides the only nondeductive element in inductive logic. Relying on no inductive assumptions, it provides the probability values to plug into the formulas of the mathematical calculus in order to get the whole inductive enterprise off the ground. It should be noted, by the way, that Reichenbach (1949, Sec. 86) offers a treatment of induction by enumeration in advanced knowledge. In this context he is free to make use of any of the tools of mathematical statistics, provided there is inductive evidence to support the assumptions demanded by these methods.

Reichenbach made another fundamental distinction, namely, between the *context of discovery* and the *context of justification*. Although he often emphasized its importance,[22] he did not, to the best of my knowledge, appeal to it in his arguments concerning the justification of induction. We might begin by asking whether his rule of induction belongs to the former or the latter context. Clearly any decision to examine As to find out whether or how often they are Bs belongs in the context of discovery. What about the inductive posit itself? The use of the rule of induction to *arrive at* a value to posit is also part of the context of discovery; at the same time, it looks like part of the context of justification as well, for the posit is justified by virtue of the rule of induction. Although, as I have argued in detail (1970), it is possible for a given item to belong to both contexts, let us consider it, in the first instance, as a method of discovery only.[23]

Suppose, then, that somehow our curiosity has been aroused regarding the probability that an A is a B. We look at n of the As and find m to be B. We might posit – i.e., guess – that the probability of an A being a B is about m/n. We might further hypothesize – i.e., guess – that the distribution is Bernoullian.[24] On that assumption about the distribution, we can calculate the size n of the sample required to have a given degree of confidence that the actual probability is within *a specified* δ of the observed frequency S_n.[25] Our hypothesis may, of course, be false; the distribution may be far from Bernoullian. This does not vitiate our investigation, however, since the assumption is itself subject to statistical testing by standard means. Thus, the assumption that was introduced as a hypothesis in the context of discovery can be confirmed or disconfirmed in the context of justification.

The Bayesian approach can also employ the distinction between discovery and justification. What the subjective Bayesian takes as a personal probability the objective Bayesian can regard as a guess in the context of discovery. For an objective Bayesian, the guess might take the form of an assumption that the sampling procedure is random. This, too, can be subjected to statistical tests.

It may seem that we run a risk of violating the requirement of total evidence if we decide to ignore the order of items in the sample, but this is not necessarily the case. The requirement of total evidence requires us to take account of all available *relevant* evidence. We normally have a great deal of evidence about our samples that is not taken into account in making various inferences or estimates – e.g., whether the evidence was collected at night or during the day, whether the collector's hair was blond or not, whether the study was conducted in winter or some other season. Notice that such information would clearly be relevant to certain investigations, but for many others it would not. Standard practice, in general, ignores much available information; otherwise statistical studies would be too complicated to be practical. In estimating or inferring values of limiting frequencies, under many

circumstances, the order in the sample can also be ignored as irrelevant.[26]

7. CONCLUSION

Reichenbach sought to resolve Hume's problem of the justification of induction by means of a pragmatic vindication that relies heavily on the convergence properties of his rule of induction. His attempt to rule out all other asymptotic methods by an appeal to descriptive simplicity was unavailing. We found that important progress in that direction could be made by invoking normalizing conditions (consistency) and methodological simplicity (as a basis for invariance), but that they did not do the whole job. I am proposing that, in the end, Reichenbach's own distinction between discovery and justification holds the key to the solution.

8. ACKNOWLEDGMENT

I should like to express my sincere gratitude to Deborah Mayo for critical comments and constructive suggestions on an earlier draft of this paper.

NOTES

See, for example, Salmon (1957), (1965), (1967a), and (1968a).

Or any functional equivalent thereof, such as Keynes's principle of limited independent variety or Russell's postulates of scientific inference. Herbert Feigl (1949) had urged the same point.

It is the counterpart of what Camap called "the straight rule".

Reichenbach gives a very informal sketch of this argument in (1951), chap. 14.

See, for example, Sellars (1964), Sec. XII, pp. 212–14.

This point was first made by Paul Humphreys (1985). I have discussed the propensity interpretation' at length in (1979).

Although many authors, including Aristotle, had hinted at a frequency interpretation of probability, John Venn, in *The Logic of Chance* (1866) was the first to spell it out fully; see Salmon (1980).

See Salmon (1956).

We may as well classify the rule of induction as an asymptotic rule, namely, that for which c_n identically zero for all values of n.

[10] This point was first made by the statistician I. Richard Savage in discussion at the Minnesota Center for the Philosophy of Science.

[11] Using Bayes's theorem, we can calculate that, in the light of the evidence that the ten tosses resulted in heads, the probability that the coin is two-headed is about 0.001.

[12] The key principle seems to be that the frequency composition of any randomly selected sample very probably nearly matches the frequency composition of the population.

[13] By suitably specifying precise degrees of approximation in the three formulas, something akin to transitivity can be salvaged, though it is not actually transitivity, for "≈" does not remain univocal throughout the argument.

[14] The fallacy in arguments of this type was pointed out in my (1961) and (1965).

[15] At that time it had been estimated that 90% of all scientists that ever lived were then alive. I do not know what the current percentage would be – still fairly high I should imagine.

[16] We shall also be easy victims of Kyburg's lottery paradox.

[17] Herbert Feigl made a similar point, and elaborated it somewhat more fully, in (1949).

[18] Actually, I would have preferred to go to Rome, but since, as the saying goes, all roads lead to Rome, that would not have constituted a suitable example.

[19] It has often been noted, as I did above, that any observed frequency is *deductively* compatible with any limit. The issue here is not, however, *deductive relevance* but rather *inductive relevance*.

[20] Linguistic invariance requires invariance under permutations of predicates; statistical invariance requires invariance under permutations of properties.

[21] At any rate, this was my longstanding objection to attempts to justify induction by appealing to standard statistical methods.

[22] See Reichenbach (1938, pp. 6–7), (1947, p. 2), and (1949, pp. 433–34). Robert McLaughlin (1982) has argued persuasively that it would be preferable to call these the *context of invention* and the *context of appraisal*; however, in this historical discussion I shall retain Reichenbach's terminology.

[23] In an unpublished manuscript, Deborah Mayo has pointed out that an excellent method for discovering and justifying a statement about the mean score on a test is to add up all of the scores and divide by the number of students in the class taking the test.

[24] One may, of course, hypothesize a different sort of distribution, but it, too, can be tested statistically.

[25] Reichenbach has often been criticized for failure to provide any way of establishing specific values for δ.

[26] In his (1980) Hacking relinquishes his earlier position on foundations of statistics, and adopts the Neyman–Pearson approach. It seems to me likely that his position regarding the symmetry requirement and the status of information about order in the sample would be revised in consequence.

REFERENCES

Black, Max: 1954, *Problems of Analysis*. Ithaca, NY: Cornell University Press.

Carnap, Rudolf: 1952, *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

Clendinnen, F. John: 1982, 'Rational Expectation and Simplicity', in Robert McLaughlin (ed.), *What? Where? When? Why?* (Dordrecht: D. Reidel Publishing Co.), pp. 1–25.

Feigl, Herbert: 1949, 'The Logical Character of the Principle of Induction', in Herbert Feigl and Wilfrid Sellars (eds.), *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts), pp. 297–304. Originally published in *Philosophy of Science* I (1934).

Feigl, Herbert: 1950, 'De Principiis Non Disputandum. . .?' in Max Black, ed., *Philosophical Analysis* (Ithaca, NY: Cornell University Press), pp. 119–56.

Hacking, Ian: 1968, 'One problem about induction',' in Imre Lakatos, ed., *The Problem of Inductive Logic* (Amsterdam: North-Holland Publishing Co.), pp. 44–59.

Hacking, Ian: 1980, 'The theory of probable inference: Neyman, Peirce and Braithwaite', in D. H. Mellor (ed.), *Science, Belief and Behaviour* (Cambridge: Cambridge University Press), pp. 141–60.

Humphreys, Paul: 1985, 'Why Propensities Cannot Be Probabilities', *Philosophical Review* **XCIV**, 557–70.

McLaughlin, Robert: 1982, 'Invention and Appraisal', in Robert McLaughlin (ed.), *What? Where? When? Why?* (Dordrecht: D. Reidel Publishing Co.), pp. 69–100.

Popper, Karl R.: 1959, *The Logic of Scientific Discovery*. New York: Basic Books.

Reichenbach, Hans: 1938, *Experience and Prediction*. Chicago: University of Chicago Press.

Reichenbach, Hans: 1947, *Elements of Symbolic Logic*. New York: Macmillan.

Reichenbach, Hans: 1949, *The Theory of Probability*, 2nd ed. Berkeley & Los Angeles: University of California Press.

Reichenbach, Hans: 1951, *The Rise of Scientific Philosophy*. Berkeley & Los Angeles: University of California Press.

Reichenbach, Hans: 1954, *Nomological Statements and Admissible Operations*. Amsterdam: North-Holland Publishing Co. Reissued by the University of Califomia Press under the title, *Laws, Modalities, and Counterfactuals* (1976, same pagination).

Salmon, Wesley C.: 1956, 'Regular Rules of Induction', *Philosophical Review* **LXV**, 385–88.

Salmon, Wesley C.: 1957, 'Should We Attempt to Justify Induction?' *Philosophical Studies* **8**, 33–48.

Salmon, Wesley C.: 1957a, 'The Predictive Inference', *Philosophy of Science* **24**, pp. 180–90.

Salmon, Wesley C.: 1963, Review of John Patrick Day, *Inductive Probability*, in *Philosophical Review* **LXXII**, 392–96.

Salmon, Wesley C.: 1963a 'On Vindicating Induction', *Philosophy of Science* **30**, 252–61.

Salmon, Wesley C.: 1965, 'Consistency, Transitivity, and Inductive Support', *Ratio* **VII**, 164–69.

Salmon, Wesley C.: 1967, 'Carnap's Inductive Logic', *Journal of Philosophy* **LXIV**, pp. 725–39.

Salmon, Wesley C.: 1967a, *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Salmon, Wesley C.: 1968, 'Reply', in Imre Lakatos (ed.), *The Problem of Inductive Logic* (Amsterdam: North-Holland Publishing Co.), pp. 84–85.