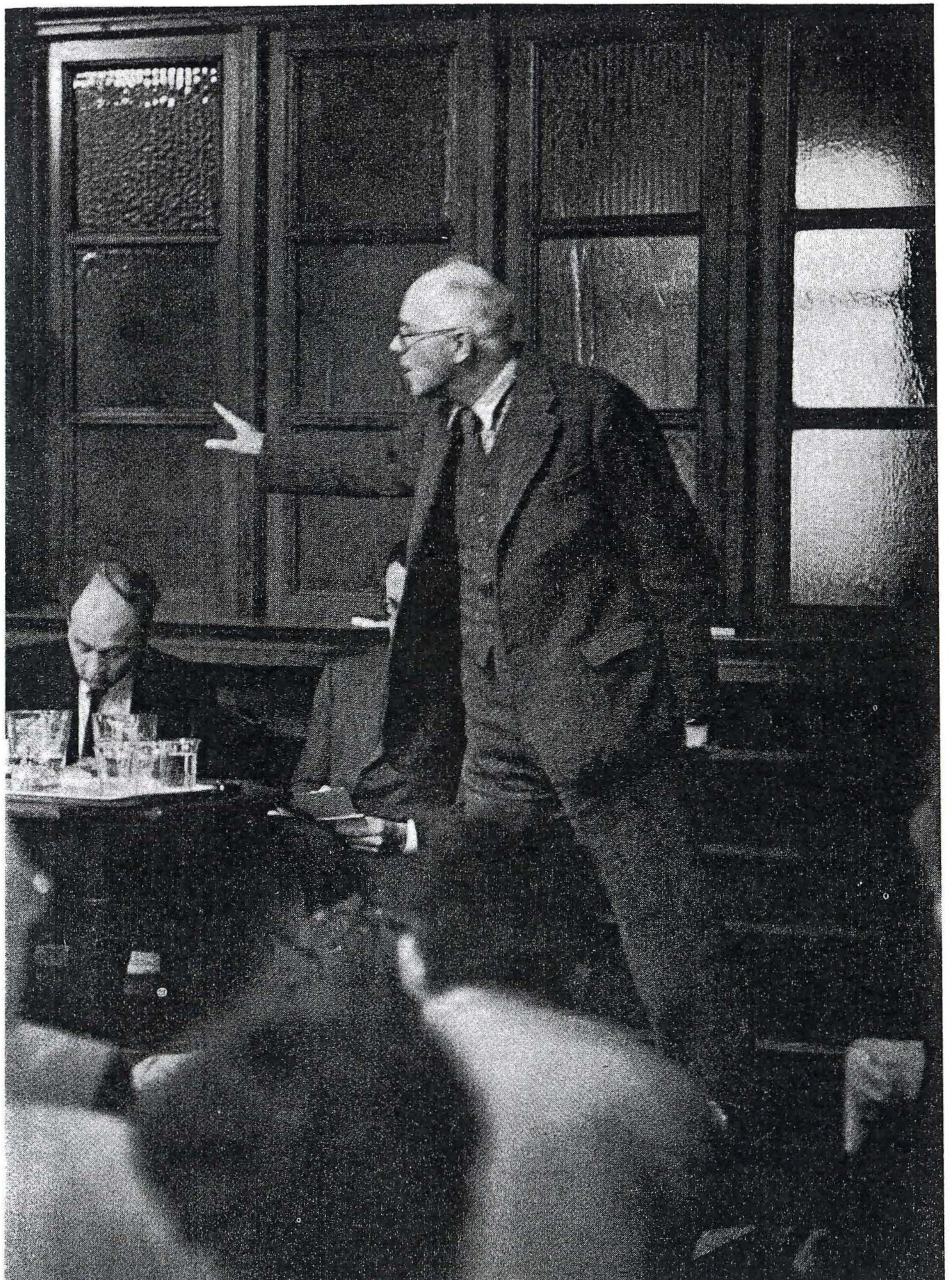

R. B. Braithwaite lecturing at the 1952 Joint Session of the Aristotelian Society and the Mind Association (*Radio Times Hulton Picture Library*)

Science, belief and behaviour

ESSAYS IN HONOUR OF R. B. BRAITHWAITE

Edited by D. H. Mellor

1980

CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

REFERENCE

Clopper, C. J. and Pearson, E. S. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–13.

8 *The theory of probable inference: Neyman, Peirce and Braithwaite*

IAN HACKING

This paper will show that the Neyman–Pearson theories of testing hypotheses and of confidence interval estimation are sound theories of probable inference. That is worth arguing because Neyman insisted that there is no such thing as inductive inference; there is only inductive behaviour. Moreover his chief critics, whether they be Bayesian or follow in the footsteps of R. A. Fisher, contend that scientific practice needs a theory of inference, and seriously fault Neyman on that score. I contend that this *contretemps* is provoked by a misunderstanding about inference, and that Neyman has, indeed, provided just the fulfilment of the theory of probable inference that was sketched a century ago by Charles Sanders Peirce.

This volume is a good place to argue the point, for R. B. Braithwaite was the first philosopher to incorporate modern statistical practice into his discussions about science, and his book *Scientific Explanation* makes ample use of the Neyman–Pearson approach. Braithwaite was also an enthusiast for Peirce's ideas about probability and induction, although he did not connect that part of his exposition as closely to Neyman as I now do. It is high time to revive Braithwaite's enthusiasm for Neyman–Pearson statistics. A recent careful article in *The Journal of Philosophy* dismisses the confidence interval approach: 'I believe adequate criticism of this theory is already available in the literature' (Seidenfeld 1978: 710). That is a mistake for which I must share responsibility. My *Logic of Statistical Inference* took vigorous issue with Neyman. This essay is a retraction. I now believe that Neyman, Peirce and Braithwaite were on the right lines to follow in the analysis of inductive arguments.

The scandal of philosophy. A little local history will serve to fix ideas. In 1926 Francis Bacon's third centenary was celebrated in the Senate House of Cambridge University. C. D. Broad concluded a notable lecture by hoping 'that when Bacon's next centenary is celebrated the great work which he set going will be completed; and that Inductive Reasoning, which has long been the glory of Science, will have ceased to be the scandal of Philosophy' (Broad 1952: 142). Whether or not induction is the glory of science, induction, in at least one standard sense of that term, ceased to be the *scandal* of philosophy within a decade, although it will remain problematic for a long time. Each of the three probabilistic routes to the problem of induction had been well worked out within the following ten years, with a good deal of the work done a few hundred yards from the Senate House.

The first way to apply probability to induction is, loosely speaking, "logicist". Keynes had published his version in the 1921 *Treatise on Probability*, which had already been in circulation for some years before appearing in print. The ideas of Harold Jeffreys' *Theory of Probability* were evolving. But the most important work was that of R. A. Fisher. In order to apply logicist conceptions one needs new concepts and *a priori* principles to connect evidence and hypotheses. Sufficiency, ancillarity, likelihood, measures of information and the like were all produced or understood around this time. Indeed had one fully grasped the potentiality of Fisher (1922), that great paper on the foundations of statistics, one might not, even in 1926, have used Broad's phrase, 'the scandal of philosophy'.

Broad lectured on 5 October. The published version of a paper read to the Moral Sciences Club on 26 November repeats the phrase with, perhaps, a touch of derision. We now know that paper well; it is 'Truth and probability' (Ramsey 1931: 197 or 1978: 99). Although Ramsey is less of a subjectivist than is commonly made out, he did begin the second great modern application of probability ideas to induction, the personalist approach rediscovered by Bruno de Finetti and elaborated by L. J. Savage.

Finally there are the ideas associated with Jerzy Neyman and E. S. Pearson: confidence intervals and hypothesis testing using size and power. The two men began their collaboration early in 1926 at University College, London. The confidence interval approach was in fact anticipated by the Harvard mathematician Edwin B. Wilson, a life long devotee of C. S. Peirce.[1] Peirce had long before formulated the chief philosophical idea of the Neyman–Pearson work, that probable inference does not assign a probability to individual hypotheses, but rather draws inferences according to a system with some known rate of success. Despite such earlier formulations, the contributions of Neyman and Pearson were of course independent of Wilson and Peirce. Neyman came to make as trenchant a statement about induction as can be found in the work of any philosopher. There is no such thing as inductive inference – but it does not matter for we can still act wisely by choosing the best patterns of inductive behaviour (Neyman 1957). Hume was right to say that induction is a matter of habit rather than of reason for individual conclusions, but wrong not to see that we can assess our habits. In their classic joint paper of 1933 Neyman and Pearson put the claim more cautiously: 'As far as a particular hypothesis is concerned no test based upon the theory of probability can by itself provide valuable evidence of the truth or falsehood of that hypothesis' (Neyman & Pearson 1933: 290).

It may seem remarkable that Braithwaite made so much use of the ideas of Neyman and Pearson. He was taught about probability and induction by Keynes, and he was Ramsey's contemporary and posthumous editor. Why should he have become so attached to the third great idea about probability and induction, rather than the other two? Perhaps it was partly because there was at the time much enthusiasm, in Cambridge, for some strands in pragmatist thought.

[1] For Wilson's biography and bibliography, see (Hunsaker & Maclane 1973). One of Wilson's teachers was B. O. Peirce and he was fascinated by the entire family, even to the extent of using their genealogical table as a basis for statistical analysis (Wilson & Doering 1926). He recomputed Peirce's experimental studies on the normal distribution (Wilson & Hilferty 1929); the paper has interesting consequences. He described Peirce as 'an expert in making refined physical observations and in reducing them, and a great logician and philosopher' (Wilson 1926a); his admiration for Peirce's writings on probable inference is stated in (Wilson 1926b); the 'confidence interval' paper (Wilson 1927) echoes Peirce in the very title, 'probable inference'. When Neyman graciously attributed the confidence theory to Wilson (Neyman 1952: 222) Wilson said he could not claim that priority. He had merely tried to correct the "logic" of reasoning that employs standard deviation (Wilson 1964: 293). Wilson was rather conservative in demeanour, and dubious of all generalizations not only about statistics but about scientific methodology in general. He did not think of confidence intervals as a universal tool but as a device for a particular job. He expressed at least one reservation about the philosophical basis of Neyman and Pearson (Wilson 1942: 90).

Peirce understood the foundation of the confidence idea. The purpose of this extended footnote is not to claim priority for Wilson but to show that Peirce's ideas are connected, by direct lines of filiation, with a more modern exponent of the technique: his work did not lie entirely fallow. Of course the general confidence idea occurs, in some form, in numerous nineteenth-century writers; A. A. Cournot is often cited in this connection.

'What follows', Ramsey put in a footnote to the concluding section of his paper, 'is almost entirely based on the writings of C. S. Peirce.' Ramsey called his subjective theory 'the logic of consistency'; the subsequent "logic of truth" based on Peircean ideas is less well known today. Much of Braithwaite's exposition runs parallel to Ramsey and Peirce. Braithwaite draws on Peirce most evidently in his "predictionist justification of induction", but I shall show how pertinent it is to the theories of Neyman and Pearson too.

Logic. What I call a 'logicist' approach to induction holds (a) that correct induction is to be analysed in terms of relations which exist between a body of evidence and an individual hypothesis that it supports or disconfirms, and (b) that these relations are logical in character, so that propositions asserting that these relations hold are *a priori* true or false. Keynes thought that the canonical forms of such propositions are:

(1) The probability of H on A is p,

and

(2) H is more probable on A than H^* on A^*.

He apparently believed that the probability relation in (1) is like a partial (logical) implication, so that (1) bears an analogy with,

(3) A implies H.

Just as the propositional calculus and *Principia Mathematica* provide constraints on and interconnections among statements of the form (3), so a probability calculus would do the same for (1) and (2). Quantitative probabilities obey the normal rules for probability, while Keynes sketched out a calculus for comparative probabilities, completed in the axioms for "intuitive probability" of Koopman (1940). Koopman used the term 'intuitive' because Keynes held that many judgements of forms (1) and (2) are in the end answerable only to intuitions, on analogy with what G. E. Moore had written about moral judgements in *Principia Ethica*.

Ramsey had a great many criticisms of Keynes' logicist approach, but the most immediately telling "is the obvious one that *there really do not seem to be any such things as the probability relations he describes*".

It may seem surprising to call R. A. Fisher a logicist, but in at least one respect he is close to Keynes. He differs in that he never succumbed to what I call 'the hegemony of probability'. Unlike Keynes he does not believe that all relations between hypotheses and evidence are to be couched in terms of a quantitative, comparative or qualitative relation to probability. Instead he devised his likelihood, significance levels, analysis of contingency tables and so forth. Only when a body of data has a peculiar (and usually deliberately contrived) structure can we engage in that form of "exact inference" that involves overt probability statements. But although Fisher never attached that exaggerated importance to probability statements (1) and (2) that is so common among statisticians and philosophers, he does satisfy my definition of 'logicist' with which I began this section. There are, he held, a variety of *a priori* relations that hold between bodies of data and individual hypotheses, in virtue of which we draw inductive conclusions. My own *Logic of Statistical Inference* unquestioningly made the same assumption. I can now say, at least for the purposes of this paper, what Ramsey said about Keynes: there do not seem to be any such relations. I do not mean that there is no role for likelihood or significance levels *etc.*, but only that these fundamental concepts should not be understood in a logicist way. Nor does my present opinion impugn Fisher's achievements, for it only calls in question a particular way of understanding what he was doing.

It may be tempting to sum up this opinion in the words, 'There is no such thing as a logic of statistical inference'. But to say that is to grant too much to the logicist, for it is to suppose that (1), (2) and the like are the province of inductive logic. On the contrary they are founded on a false analogy with deductive logic. In the next two sections I shall try to destroy the analogy. Unfortunately this involves writing the sort of prose that I do not care to read – with lots of numbered sentences and fine distinctions. For the reader with a similar lack of taste for this sort of work, the results are abstracted in the third section to follow, headed *summary*.

Statement and argument. Peirce took for granted that the topic of logic is argument, that is, the transition from some sentences to others as in the simplest form,

(4) S, so H.

Here S is a set of sentences (whose conjunction, if finite, we may represent by A). Keynes thought that the topic of inductive logic would focus on probability relations as exemplified by (1) or the comparative (2) or the qualitative (5):

(5) Given A, H is probable.

I believe that this inclination of Keynes and others was partly fostered by a number of linguistic slides. First, it may be supposed that if (4) is the topic of deductive logic, then inductive logic must concern itself with arguments of the form,

(6) *S*, so probably *H*.

In the following section I shall argue that inductive and deductive logic are both concerned with arguments of the form (4), and that (6) is not a new form of argument but a way of stating the argument of form (4). Postponing that discussion, suppose that as inductive logicians we are concerned with (6). Moreover, we propose to do inductive logic on analogy with deductive logic. We should first consider, then, what the deductive logician does with (4).

Peirce and the older generations of deductive logicians may have thought that (4) should be studied directly, but by the time of *Principia Mathematica* (*PM*) a false analogy with mathematics had made it standard to regard logic as an axiomatic science of truths, not as a science of rules of inferences, of transitions between sentences. Indeed *PM* somewhat surprisingly tries to smuggle in its rules of inference among the axioms as one more "primitive proposition". Only in the 1930s did the logician Gerhardt Gentzen restore the older tradition, showing how to get first-order logic from rules of inference alone. Certainly when Keynes wrote he could accept that logic was to be studied not directly as in arguments (4) but by statements of implication such as (3):

(3) *A* implies *H*.

Moreover, as Quine has observed, use and mention were confused. Conditionals of the form 'If *A* then *H*' are treated in the propositional calculus by the truth-functional conditional,

(7) $A \supset H$.

The horseshoe was read, '*A* materially implies *H*.' But a statement such as (3) asserts that a relation holds between two sentences, namely the relation of implication. It mentions (to use Quine's jargon) these sentences *A* and *H*, and says something about them. But the conditional (7) uses these sentences to form a complex sentence. '*A* implies *H*' and 'If *A*, then *H*' are *not* synonymous. Now the argument '*A*, so *H*' is truth-functionally valid if and only if (7) is a tautology, or a theorem of the propositional calculus. So it was thought that the argument '*A*, so probably *H*' must be inductively valid if and only if a statement such as (5), 'Given *A*, *H* is probable' is a theorem of a probability calculus – or if not (5) then some quantitative statement such as (1). Moreover, it was thought that (1) or (5) would be parts of a probability calculus in the way that (7) is, *and, at the same time*, that (1) and (5) express relations between the propositions *A* and *H*. That is the result of the confusion of use and mention that conflates (3) and (7). On the relational view of probability, (1) is like (3), expressing a relation between propositions, but then it is not like (7), *i.e.* not like an ingredient in the propositional calculus. But, like most confusions of use and mention, this error would have been unlikely to do any harm in itself, had it not been compounded with a more serious mistake.

Probable inference. An argument of the form,

(4) *S*, so *H*,

will be valid if and only if the statement of the form,

(3) *A* implies *H*,

is true. So can we not say that an argument of the form,

(6) *S*, so probably *H*,

is inductively valid if and only if the relational statement of the form,

(5) Given *A*, *H* is probable,

is true? It may only be a matter of taste to prefer Gentzen's regarding logic as a matter of argument (such as (4)) as opposed to *Principia Mathematica*'s treating it as a matter of axiomatizing sentences (such as (3)). May we not then as a matter of taste go along with Keynes and make (1) or (5) the topic of inductive logic? I think not, because I believe that the 'probably' in (6) is like the 'necessarily' in one reading of,

(7) *S*, so necessarily *H*.

Now one can say 'necessarily *H*' to mean that *H* is a logically necessary proposition. But in the context of (7) the common use of 'necessarily' is to indicate that *H* follows necessarily from *S*. That is, 'necessarily' modifies the argument, the way in which *H* is reached. It is not an adjective modifying *H simpliciter*.

Thus (7) is not a different argument from (4), '*S*, so *H*.' It is a different way of expressing the same argument, at the same time indicating the force with which one may draw the conclusion from the premisses. Likewise (6) is not a different argument from (4); the 'probably' in (6) modifies the argument, not the conclusion *H*.

Yet it will be protested that we can say simply,

(8) Probably *H*.

Why cannot (8) be the conclusion of the argument (6)? In answer I shall recall two observations from the time when Oxford linguistic philosophy was at its apogee. One is due to D. G. Brown, and concerns the word 'inference'; the other is by Stephen Toulmin, about the word 'probably'.

I retain the logician's use of the word 'argument': a string of sentences with a concluding sentence. Brown (1955) says that 'inference' does not mean 'argument' in this sense. 'S, so H' represents an argument, not an inference. The *inference* is H. An inference is a concluding statement made in an argumentative context. Incidentally, Peirce sometimes appears to be using the word in this way, as do most writers of his time who treat of logic. But unlike Brown I do not claim that this is the only or even the favoured way of using the word 'inference', only that it is a correct and for us instructive way to use it.

Next examine what Toulmin (1965) says about 'probably' and combine it with Brown's idea about inference. He says that 'probably H' is not a statement about H. It is a way of stating H. 'Probably' is as it were in parentheses, employed to make the statement H in a certain way. To state H in normal conversation is to imply that H is warranted, or that one will stick by H, backing it up with reasons, and so forth. To say 'probably H' is to state H in a guarded way. That is, to imply that one's back-up for H is less than conclusive, and that one grants that H might turn out to be false, although one doesn't expect that.

Now we can combine Brown and Toulmin to comment on the form of words,

(6) S, so probably H.

This is just the argument, 'S, so H'. In either case the inference is H. This inference is a statement which may be guarded, like any other, by prefixing 'probably'. But in the context of the argument one knows the source of the caution that prompts this guarded statement: the argument is not a conclusive one.

It will be noticed that an argument may be inconclusive for more than one reason. The premisses may not entail the conclusion, so the argument may be offered as a merely probable argument. But, also, the premisses may be stated in a guarded way, as when one says, 'probably S, so probably H'. If S entailed H I would not call that a probable argument. But the function of 'probably' in the conclusion has the same core role as in probable argument, namely

to express caution about the statement H. Naturally none of this takes us any positive distance towards understanding probable argument. These have been merely negative remarks, intended to dispel the illusion that inductive logic should concentrate on statements of the form,

(1) The probability of H on A is p.

Perhaps I should add that Ramsey seems to have subscribed to the view that "probable inference" would be a theory of drawing inferences to conclusions (1). 'The suggestion of Mr Keynes that [Hume] can be got round by regarding induction as a form of probable inference cannot in my view be maintained. But to suppose that the situation which results from this is a scandal to philosophy is, I think, a mistake.' Ramsey wrote these words in the course of elaborating on Peirce. I differ from Ramsey chiefly because I believe that what results from Peirce's ideas is exactly what should be called a theory of probable inference, and that only logicism prevented one from seeing that. Peirce himself always called it either probable inference or probable argument. E. B. Wilson, one suspects with Peirce in mind, and writing at exactly the same time as Ramsey, titled his version of confidence theory just that: 'probable inference'.

Summary of the preceding two sections. Logic is concerned with the transition between sentences, *i.e.* with arguments 'S, so H'. In such argument the inference is the conclusion H. An inference is not a transition but a statement. A statement may be expressed in a guarded way, *e.g.* by prefixing the word 'probably'. Thus the argument 'S, so probably H' is not an argument to a conclusion 'probably H' but an argument to the conclusion H, which inference is expressed in a guarded way. Inductive logic is not the logic of statements of the form, 'the probability of H on A is p'. These have been taken to be the province of inductive logic for a mixture of poor reasons. Such statements were wrongly taken to be the conclusion of any quantitative inductive argument: they are not. Moreover certain false analogies with implication, and confusions of use and mention, abet this tendency among earlier inductive logicians.

A genus of arguments. A valid demonstrative argument, wrote Peirce, 'is a member of a genus of arguments all constructed in the same way, and such that, when their premisses are real facts, their conclusions are so also' (Peirce 2.649).[2] Today's logician would give a semantic account of this in terms of logical consequence. That requires some apparatus. A model structure for a class of sentences closed under suitable syntactic operations consists of a class of models or interpretations of the sentences, together with functions that have the effect of assigning truth values to sentences in the models. *H* is a logical consequence of *S* when *H* is true in all models in which every member of *S* is true.

Evidently there are formal devices that will define a notion of partial logical consequence, even one which is amenable to a probability measure. That might seem the right way to explicate Peirce's frequent pronouncements to the effect that 'We may, therefore, define the probability of a mode of arguments as the proportion of cases in which it carries truth with it' (2.650). That, however, would not do. He had more than proportion in mind, for in the same paragraph he writes: 'In the long run there is a real fact which corresponds to the idea of probability, and it is that a given mode of inference sometimes proves successful and sometimes not, and that in a ratio ultimately fixed.' A model theoretic approach needs to be augmented by this "real fact" of stable long run frequency.

Frequency. When Peirce was young he was a nominalist who had no truck with contrary-to-fact conditionals. The probability of getting an ace in tossing a die is a relative frequency of aces in a sequence actually performed. A diamond at the bottom of the sea is hard only if actually tested. As Peirce matured he adopted a scholastic realism and came to see that what would be the case is an integral part of the way the world is. The diamond is hard if it would have proved hard; the probability of getting an ace is the "would-be" of the die, as Peirce says in what Braithwaite rightly calls a 'striking description' (Peirce 2.664; Braithwaite 1953: 187).

Braithwaite's account of objective probability has not been significantly bettered. Probability in this sense does not mean "relative frequency", but probabilities are typically manifested by stable frequencies. Probabilities conform to the usual probability axioms which have among their consequences the essential connection between individual and repeated trials, the weak law of large numbers proved by Bernoulli. Probabilities are to be thought of as theoretical properties, with a certain looseness of fit to the observed world. Part of this fit is judged by rules for testing statistical hypotheses along the lines described by Neyman and Pearson. It is a "frequency view of probability" in which probability is a dispositional property, subject "to the whole gamut of logical and epistemological considerations developed in" Chapter VI of *Scientific Explanation*. Naturally various niceties will be added to this account, *e.g.* Patrick Suppes' programme of proving representation theorems for propensities, and there will be debates about what the propensities are propensities *of*, whether truth about probability requires indeterminism, and the relation between single cases and the long run (as if Bernoulli had not settled that). But Braithwaite's account is a good careful philosophical explication of what statisticians mean, as when Fisher talks about relative frequencies in hypothetical infinite populations. Neyman thanked Braithwaite for this quite explicitly and it is entirely in the spirit of Peirce. How are we to apply it to Peirce's conception of probable inference?

[2] References to Peirce are in the customary form of volume and paragraph numbers of his collected papers (1932).

The truth-producing virtue. Peirce defined *validity* as 'the possession by an argumentation or inference of that sort of efficiency in leading to the truth, which it professes to have' (2.779). He went on to call this the 'truth-producing virtue'. In the case of inductive argument, this virtue consists in the fact that the argument is of a form which would lead to the truth, "for the most part".

> It may be conceived, and often is conceived, that induction lends a probability to its conclusion. Now that is not the way in which induction leads to the truth. It lends no definite probability to its conclusion. It is nonsense to talk of the probability of a law, as if we could pick universes out of a grab-bag and find in what proportion of them the law held good (2.780).

This is equally the opinion of Wilson, Neyman and Pearson. Peirce did not adequately develop the mathematical implications of his insight, so we pass at once to more modern procedures.

To take the simplest but sufficiently general case, suppose we require an estimate of a quantity such as the chance of getting an ace with a die or the mean of a normally distributed population. We incorporate idealized statistical assumptions about the chance set-up into a statistical model of a family of distributions with unknown parameter θ (*e.g.* the chance of an ace or the mean height of the population). A class of possible observations is represented in a sample space.

An interval estimator for θ is a function f with the following properties. First, it is a function from points in the same space to intervals on the parameter space. Secondly, for any θ in the parameter space the probability on θ of getting a sample x that covers θ is (say) 95 per cent.

Hence, assuming that the statistical model is correct, the probability – regardless of the value of θ – of making a correct estimate using f is 95 per cent. Let us then make an observation, say x_0. Using f we estimate that the unknown true value of θ, call it θ_0, is in $f(x_0)$.

This estimate is reached by a procedure that gives correct results 95 per cent of the time. All authors in this tradition rightly insist that we cannot therefore attach 95 per cent probability to the statement 'θ_0 is in $f(x_0)$.' Yet, contrary to this tradition, we may still talk about inference. There are two premisses, one stating the assumptions built into the statistical model, and the other reporting the observation. This leads to the following argument:

(i) (A premiss setting forth the statistical model.)
(ii) The sample x_0 was observed.
So, (iii) θ_0 is in $f(x_0)$.

The statement (iii) is the inference. One may express the inference in a guarded way: 'Probably θ_0 is in $f(x_0)$.' That does not mean that (iii) has some definite probability, only that one is making a cautious statement. The most informative way to express reservations about (iii) is to say that it is reached by a method that has a 95 per cent probability of being correct. An exactly parallel account may be given for inferences made using a fixed significance level or size of test.

Critics of Neyman and Pearson sometimes say that confidence intervals perpetrate a confidence trick on the innocent research worker. The routine technician conducts an experiment and obtains a 95 per cent confidence interval. But even when this person has been taught somewhere along the line that you cannot attach a 95 per cent probability to the statement, 'θ_0 is in $f(x_0)$', what the interval *means* to the researcher is just, 'the probability that θ_0 is in $f(x_0)$ is 95 per cent'. That is what the confidence interval "feels like" to the research worker, and that is how it is often used, or so the critics say. I think it is only the logicist instincts of the critics, and their false view of language, that leads them to impugn the research

worker. The person in the laboratory does draw an inference (Neyman was wrong about that) but the inference is just, 'θ_0 is in $f(x_0)$.' That is the inference one wants, together with an indication of the reliability of the inference. Indeed the older version of the same way of inferring is still taught to physics students; one concludes an experimental measurement with an indication of the probable error. The old term 'probable error' is a horror, but the idea is sound enough.

Efficiency. Peirce, as quoted above, said that the validity of an argument is a kind of efficiency, but only the precise analysis of Neyman and Pearson makes clear the aptness of this metaphor. Their 1933 paper employs the word in its title – 'the most efficient tests of statistical hypotheses'. In our terms the problem of efficiency arises because if there are any 95 per cent estimators there are typically too many. How to choose the best 95 per cent estimator, or the best 95 per cent test?

Peirce took a thoroughly practical attitude towards reasoning. Good reasoning serves certain ends. He idealized the end as the pursuit of truth. But as Isaac Levi says in a related connection, this pursuit involves two goals, a desire to be correct, and "relief from agnosticism" (Levi 1967: 62). One can minimize mistakes by never engaging in ampliative inference; it is the desire for more information that leads to risk. 'Believe truth! Shun error! – These we see are two materially different laws', as William James puts a not unrelated point with characteristic flair.[3] In terms of estimators one can serve both these masters, truth and informativeness, if one can find a 95 per cent estimator which, for all θ, gives uniformly narrowest estimates. That is, we would prefer an estimator which, at given security level, produces narrower estimates than other estimators, regardless of the sample observed and the true known value of the parameter. Unfortunately this tidy solution is mathematically available only in the simplest cases.

E. S. Pearson has described how he and Neyman tried to formulate this kind of problem in general terms and arrived at some of the solutions that have since become standard (Pearson 1966). It has been increasingly realized that unique solutions become available

[3] (James 1897: 18). John Etchemendy drew my attention to the parallel between William James' injunction and the Neyman–Pearson motto, 'control the size of a test; maximize its power'.

only when one specifies more precisely what the estimate or test is for. Mathematical work in this branch of statistics has been transformed into decision theory. I do not take it to be a criticism of the Neyman–Pearson theory that there is no mindless procedure for applying it to any statistical problem. It is surely a virtue sometimes to be speechless or to invite more data or to ask exactly what is the point of the inquiry. A great many real-life problems, even after they have been forced into parametric models, simply do not have uniquely best solutions without further analysis, abstraction or experimentation.

My remarks are not intended to endorse every Neyman–Pearson style solution on the books. The most serious criticism of the method is that it does not always use the information which may, perhaps by ill luck, be cast on the experimenter's plate. A policy which in general is a good one might on a particular occasion reveal itself as inept, as in the distinction between before-trial and after-trial betting (Hacking 1965: 95–102). But it is important to recall that Neyman–Pearson techniques, understood as rules of inference, are not rules that tell you what you must do on every occasion. To see this, some further discussion of the use of such rules is in order.

Acceptance. The theory of testing hypotheses was first presented as a decision procedure with two outcomes: either accept or reject the hypothesis under test. Braithwaite observed that 'don't reject' does not imply 'accept'. He himself had something of a falsificationist methodology, according to which only falsifiable statements have real content; since he was examining the meaning of statistical hypotheses, he emphasized rejection. The terminology of 'accept' and 'reject' had won some favour among statisticians because of applications to quality control where one is literally accepting a consignment of light bulbs. It was soon realized that testing readily generalized to three or more options, *e.g.* 'accept', 'reject', 'suspend judgement until examining a further batch'. The idea of accepting hypotheses did no harm in statistics. The same may not be said of philosophy.

The philosopher's exemplar of a rule of inference is *modus ponens*, 'From A and $A \supset B$ to infer B.' It seems absurd to assert the premisses while denying the conclusion, so such rules have been compared to obligations: they *must* be followed, on pain of unintelligibility. Traditionally one said that if the premisses are true, the conclusion

must be true, or follows necessarily. There is little harm in comparing rules of inference in deductive logic to commands.

Since in an ampliative argument the conclusion does not necessarily follow, we should not think of inductive rules as commands; perhaps they are more like licences, granting permission. Although such deontic comparisons are largely futile, they can prevent some obfuscation. Consider the lottery paradox. I learn that a fair lottery has 10 000 tickets, of which only one will receive a prize. I am offered ticket 401. Reflecting on the information to hand, I infer, *i.e.* state, with hardly any reservations, that ticket 401 will not win. I have used a form of argument whose probability is 99.99 per cent; only one time in 10 000 would I err if I regularly reasoned in this way. Then it is protested: in parity I must draw the same inference about tickets 402, 403, 9999, 1, 2, . . ., 399. So I must infer that no ticket will win!

I "must" infer nothing of the sort. I need infer nothing whatsoever. To infer is to make a statement on the basis of reasoning. When offered 401 I infer with almost no hesitation that it will not win. If I am offered my pick of any of the tickets, I infer with equal confidence that whatever ticket I pick, it will not win. But I do not propose to make 10 000 inferences; I should get hoarse. But if I make several inferences taken together, I may have more reservations about the joint inference than about any one inference made individually. If I am offered all the tickets 401–500, I shall say that *probably* none will win. That inference is drawn according to a rule with 99 per cent probability, and perhaps that is when I start expressing caution.

Henry Kyburg, who has placed much store by the lottery paradox, agrees one can infer that ticket 401 won't win. But he goes on to propose that I should "admit into my corpus of belief" – which is how he understands inferring – all the propositions of the form, 'ticket n won't win'. According to him, I infer each of 'n won't win', from $n = 1$ to $n = 9999$. But he denies that if I infer A and infer B, then I must infer A & B (Kyburg 1970). On my account, to infer is to perform a certain act, to wit, making a statement. To state A and to state B, on the same occasion that one declines to state A&B, can only invite an inquiry as to what on earth one means. But to state A guardedly on the same occasion as one states B guardedly is consistent with expressing even greater caution about the joint truth of A and B. Perhaps I am taking a position, on the lottery paradox, that

[illegible] ~~reject~~ accept [illegible] resembles Kyburg's. But it seems much less tendentious when couched in terms of guarded inference. The school of Neyman and Pearson sometimes said of course it dealt with inference, for it provided a theory of acts, and saying that θ_0 is in $f(x_0)$ is just as much an act as buying θ_0 board feet of lumber. Although this is sometimes dismissed as idle sophism, it is not. The theory of inference must be part of what J. L. Austin came to call the theory of speech acts.

The lottery paradox is analogous to the situation with statistical inference because Peirce, Neyman and Braithwaite transformed inverse inference – in which one infers a statistical hypothesis from experimental observations – into direct inference, in which one infers a particular case from knowledge of a statistical distribution. Knowing the distribution of the estimator f, and having observed x_0, one infers that θ_0 is in $f(x_0)$. But since there may be any number of 95 per cent estimators, one could, if one drew all the inferences that they license, infer any number of incompatible statements. The lottery paradox is exacerbated. But not only is one not obliged to infer anything that follows by any 95 per cent rule, but also there is, on Peirce's methodology, a fundamental reason for using just one rule. This arises from an aspect of his theory that I have thus far left aside: habit.

Habit. Hume notoriously concluded that inductive conclusions cannot be justified by reason, and are drawn only as a matter of habit. Peirce made a few direct allusions to Hume, but he did adopt many doctrines from the Scottish school of philosophers, Reid, Steward, Bain and so forth. The Scots dominated mid nineteenth-century philosophy and it was natural for him to take up their idea that belief is to be understood in terms of habits of action. Reasoning too was a matter of forming habits, the habits of drawing conclusions. But whereas Hume had thought habit and custom are not susceptible of criticism, "habit", for Peirce, had no connotations of irrationality. One seeks habits that will tend to lead to the truth. Use of a 95 per cent estimator is one such habit. But use of an arbitrary 95 per cent estimator, or of a lot of conflicting estimators, is not such a habit. The rational agent will choose an estimator with some optimal characteristics. But it is not required for the theory to say, in general, what is optimal. That will depend, perhaps, on the purposes to hand. Note, moreover, that the present analysis does not commit one to each of the particular solutions which have been

urged by the Neyman–Pearson school. I am contending only that the theory provides an account of inference.

Yet there remains a question. What is the merit of adopting a habit with the truth-producing virtue, if one is in fact to deploy the habit only once? If I were to reason many times by the use of a habit, then I might be rewarded by being right most of the time, but what about those cases when only one inference is to be drawn? How can long run virtues justify short run policies? My *Logic of Statistical Inference* will not be accused of understating this difficulty, yet in effect it merely repeats Peirce (2.652). How can Peirce have persisted in his theory of inference when he was so aware of this objection? One answer is, in his words, akin to Faith, Hope, and Charity (2.655). The agent that does not identify his interests with those of all mankind is irrational. I may deploy my habit only once, but my act of reasoning is only one among a host of human acts. Now this is a "nominalist" answer, attempting to justify a habit in terms of a long run frequency among actual choices, albeit not my choices. But it is defective. We believe that an argument with the truth producing virtue would be a good argument even if it were the last argument ever to be propounded; even if our race were about to become extinct. Perhaps 'Hope' is supposed to rule that out. But there is another difficulty to which Hope does not cater. We may devise a particular Neyman–Pearson solution to a problem of a sort which we have no reason to hope will ever occur again, simply because the statistical model applies to a bizarre and fortuitous concatenation of circumstances. We may have every reason to hope that our inference is unique. How then do the long run merits of a habit justify the inference?

The nominalist answer of Faith, Hope and Charity is unsatisfactory. But as Peirce insisted, a passage from nominalism to what he called scholastic realism is part of the maturing of the philosopher. We have to admit what "would be" into our account of what the universe is. What would happen in this world is an irreducible fact about this world. More surprisingly, benefits which "would accrue if" must be included among our roster of what counts as reason.

Induction. According to Peirce there are two kinds of ampliative reasoning. One he called 'probable argument', and the other he variously called 'hypothesis' or 'abduction'. He kept on changing his opinion about the relative domain of each. He long hoped to get some kind of probability for abductive inferences, 'but when I finally succeeded in clearing the matter up, the fact shone out that probability proper had nothing to do with the validity of Abduction' (2.102). Now the method of hypothesis or abduction is that of finding explanations for otherwise inexplicable collections of facts. A great deal of what Peirce says about hypothesis fits Braithwaite's classic account of the hypothetico-deductive method. Peirce thought that induction is for testing hypotheses that have been conjectured as explanations, and he foresaw the enormous role that statistical hypotheses were to have in twentieth-century science. That branch of logic which shows how to test them is to be called 'quantitative induction' (2.758). But it relates 'to but a small part of the Logic of Scientific Investigation' (2.751). I began by quoting Broad: 'induction, which is the glory of Science'. As Peirce and I understand the term, induction is *not* the glory of Science. Induction is a matter of good habits. In different epochs we attend to different glories of Science. Ancient Greece taught that the glory lies in the power to produce demonstrative conclusions, and took geometry to be the model for Science. In our day we much emphasize the power to break the old habits of thought and create new, albeit fallible, versions of the Universe. It is not a defect in a theory of induction that it should fail to account for all or even the most important parts of the logic of scientific investigation.

It is of course possible to use the word 'induction' to mean whatever it is that science does, and to cover every species of ampliative reasoning. Perhaps that is a Baconian use of the word, one found in Whewell's histories and philosophy of 'The Inductive Sciences'. It is wiser to follow Peirce, and allow of at least two kinds of ampliative reasoning. Then induction becomes a matter of humdrum habit. A theory of induction should indeed be relevant to Hume, who notoriously cared little about most of the things we now call the glories of Science. Instead he worried about everyday reasoning – not reasoning that would shift our concepts and redeploy our imaginations, but the plain, the humdrum, the habitual.

In answering Hume we need not tell him the basis for every belief. On the contrary, we need only tell him how to infer, on the basis of some propositions to which he already subscribes, some new statements about the future or about the world in general. That is precisely what is done by the Neyman–Pearson theory of probable inference. Common people no more use the categories of that theory than they employ first-order deductive logic. But just as the latter is the idealization that furnishes our best present understanding of demonstrative reasoning, so the former is one way to abstract and analyse the inchoate structure of inductive argument. But there may be different abstractions of the same thing, both instructive, yet not isomorphic. That is the present state of our understanding of inductive argument. The confidence approach furnishes one sort of model of learning from experience. The neo-Bayesian subjective tack furnishes another. I have never seen any incompatibility between the two. It is seldom noticed that F. P. Ramsey's 'Truth and probability' (Ramsey 1978) ends by trying to combine his subjective theory – 'the logic of consistency' – with Peirce's approach – 'the logic of truth'. He goes so far as to suggest that only such an embedding will provide the real explanation of why degrees of belief satisfy the probability axioms.

bk

Stanford University

REFERENCES

Braithwaite, R. B. 1953. *Scientific Explanation*. Cambridge.

Broad, C. D. 1952. *Ethics and the History of Philosophy*. London.

Brown, D. G. 1955. The nature of inference, *The Philosophical Review* **64**, 351–69.

Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London A* **222**, 309–68.

Hacking, Ian. 1965. *Logic of Statistical Inference*. Cambridge.

Hunsaker, Jerome and Maclane, Saunders. 1973. Edwin Bidwell Wilson, *Biographical Memoirs of the National Academy of Sciences*. **43**, 285–320.

James, William. 1897. *The Will to Believe*. London.

Koopman, B. O. 1940. The axioms and algebra of intuitive probability, *Annals of Mathematics* **41**, 269–92.

Kyburg, Henry E., Jr 1970. Conjunctivitis. *Induction, Acceptance and Rational Belief*, ed. Marshall Swain, pp. 55–82. Dordrecht.

Levi, Isaac. 1967. *Gambling with Truth*. London.

Neyman, Jerzy. 1952. *Lectures and Conferences on Mathematical Statistics and Probability*, Second edition. Washington.

Neyman, Jerzy. 1957. 'Inductive behavior' as a basic concept in the philosophy of science, *Revue de l'Institute Internationale de Statistique* **25**, 7–22.

Neyman, Jerzy, and Pearson, E. S. 1933. On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society of London A* **231**, 289–337.

Pearson, E. S. 1966. The Neyman–Pearson story 1926–34. In *Research Papers in Statistics*, ed. F. N. David, pp. 1–24. New York.
Peirce, C. S. 1932. *The Collected Papers of Charles Sanders Peirce*, eds. C. Hartshorne and P. Weiss. Cambridge, Mass.
Ramsey, F. P. 1931. *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite. London.
Ramsey, F. P. 1978. *Foundations*, ed. D. H. Mellor. London.
Seidenfeld, Teddy. 1978. Direct inference and inverse inference. *The Journal of Philosophy* **75**, 709–30.
Toulmin, S. E. 1950. 'Probability'. I, *Aristotelian Society Supplementary Volume*, **24**, 27–62.
Wilson, Edwin Bidwell. 1926a. Statistical inference. *Science* **43**, 289–96.
Wilson, Edwin Bidwell. 1926b. Empiricism and rationalism. *Science* **44**, 47–57.
Wilson, Edwin Bidwell. 1927. Probable inference, the law of succession and statistical inference. *The Journal of the American Statistical Association* **22**, 209–12.
Wilson, Edwin Bidwell. 1942. On confidence intervals. *Proceedings of the National Academy of Sciences* **28**, 88–93.
Wilson, Edwin Bidwell. 1964. Comparative experiment and observed association. *Proceedings of the National Academy of Sciences* **51**, 288–93.
Wilson, Edwin Bidwell and Doering, Carl R. 1926. The elder Peirces. *Proceedings of the National Academy of Sciences* **12**, 424–32.
Wilson, Edwin Bidwell and Hilferty, Margaret M. 1929. A note on C. S. Peirce's experimental discussion of the law of errors. *Proceedings of the National Academy of Sciences* **15**, 120–5.

9 *Statistical statements: their meaning, acceptance, and use*

HENRY E. KYBURG, JR

1 In *Scientific Explanation* (1953) Braithwaite offers a novel and picturesque interpretation of statistical statements. It is offered as a discussion of the meaning of 'probability' as it occurs in scientific hypotheses. In the discussion that follows, I shall use the more neutral term 'measure'. Since "statistical hypotheses" in science more often concern the distribution of a certain random quantity in a class, than merely the measure of some subclass in a reference class, 'measure' seems more appropriate than 'probability'; and this choice of terminology allows us to reserve 'probability' for purposes which will become apparent in due course.

Let us consider the meaning of 'the measure of *A*s among *B*s is *p*'. When *B* is non-empty and finite, this may be construed simply as the *proportion* of *A*s among *B*s (Braithwaite 1953: 122). According to Braithwaite, however, we cannot construe the reference class *B* as finite in the case of a *scientific* hypothesis. 'For if it were assumed that the class of reference had only a finite number of members, the generalization would be restricted to apply only to a limited number of its instances, and would thus not have the generality which we require of a generalization for it to be ranked as a scientific hypothesis' (Braithwaite 1953: 123).

On the face of the matter, this is not persuasive. A cosmological theory according to which there were only a finite number of massive bodies in the Universe would, on this ground, deprive a theory of gravitation of its status as a "scientific" hypothesis; statements in biology concerning the distribution of colours or weights in a certain species would not be allowed as "scientific", since we have good grounds for supposing that the number of instances of any species is finite; similar remarks would apply to hypotheses in