Peirce and the Nature of Evidence
Author(s): Len O'Neill
Reviewed work(s):
Source: *Transactions of the Charles S. Peirce Society*, Vol. 29, No. 2 (Spring, 1993), pp. 211-224
Published by: Indiana University Press
Stable URL: http://www.jstor.org/stable/40320412
Accessed: 05/09/2012 14:24

Len O'Neill

Peirce and the Nature of Evidence

There has been a long tradition, dating at least from Descartes, that maintains that facts predicted by an hypothesis have peculiarly strong evidential force. This claim is seldom giving an exact formulation and even more rarely is it supported by more than an appeal to intuition. Peirce was a staunch advocate of the predictive testing of hypotheses and one of the few who attempted a justification. The purpose of this paper is to consider this claim which, in deference to Peirce, I shall call the "Predesignation Principle". The implicit, sometimes explicit, contrast is with evidence known prior to forming the hypothesis which is ***explained*** by, or ***accommodated*** by, that hypothesis. How great the implied gap between accommodated and predicted evidence varies and is often unspecified. some would see accommodated 'evidence' as never having evidential value, others that it would ***sometimes*** have no such value, and yet others that it can at most have value as suggesting hypotheses worthy of testing.

Sometimes the contrast is quite stark. In an example that goes back at least to Descartes, the deciphering of a code, two such major thinkers as Whewell and Mill take opposite stands. Whewell[1] had claimed:

> If I copy a long series of letters of which the last half dozen are concealed, and if I guess these as right, as is found to be the case when they are later uncovered, this must be because I have made out the import of the inscription. To say, that because I have copied all that I

Transactions of the Charles S. Peirce Society,
Spring 1993, Vol. XXIX, No. 2.

> could see, it is nothing strange that I should guess those which I cannot see, would be absurd, without supposing such a ground for guessing.

Mill[2] replied:

> If anyone from examining the greater part of a long inscription, can interpret the characters so that the inscription gives a rational meaning in a known language, there is a strong presumption that his interpretation is correct, but I do not think this presumption much increased by his being able to guess the few remaining letters without seeing them: for we should naturally expect. . . that even an erroneous interpretation which accorded with all the visible parts of the inscription would also accord with the small remainder.

More recently Keynes[3] and Carnap[4], amongst others have, explicitly or by implication, dismissed claims of peculiar evidential force to predictions as based on a psychological irrelevancy. Alan Musgrave[5] has dubbed this the "logical" conception of confirmation. On the other hand advocates of both Popperian (Musgrave calls this "historical theory of confirmation") and Bayesian methodologies have supported, or at least given serious attention to, the distinction.

Given the long-standing disagreement on the matter, it is unlikely that we are confronted here by a *single*, clear issue. Probably there is a cluster of problems, discussed at cross-purposes, which will only gradually be sorted out. The purpose of this paper is to see whether the writings of Charles Peirce can shed any light on the matter. The paper is divided into five short sections. The first outlines some of the contemporary discussion of the issue. The second briefly outlines Peirce's "theory of probable inference". The third takes up one interpretation of his defence of his Predesignation Principle in terms of Carnap's "methodological rules". The fourth takes up an interpretation in terms of Bayesian likelihoods. The fifth introduces some considerations about preconditions of inquiry in a final attempt to support Peirce's Predesignation Principle.

I

The problem the Bayesian faces with known evidence was recently brought out by Glymour[6] in his *Theory and Evidence*. It will be useful to have Bayes theorem in two of its forms:

$$Pr(H/E) = \frac{Pr(E/H) \times Pr(H)}{Pr(E)}$$

$$Pr(H/E) = \frac{Pr(E/H) \times Pr(H)}{Pr(E/H) \times Pr(H) + Pr(E/{\sim}H) \times Pr({\sim}H)}$$

relative to background evidence K.

For ease of reference, I shall call Pr(E/H) the likelihood of H and Pr(E/~H) the antilikelihood of H. Where E is part of one's current knowledge, the value of Pr(E/H) (and of Pr(E/~H) will be 1 and hence E will not increase the probability of H.

A defender of the Predesignation Principle might welcome this as confirmation of his view. But Bayesians are not in general advocates of the principle and certainly not in the form where ***no*** evidential credit is ***ever*** given to known facts and they have therefore sought to escape Glymour's argument.

The task is to secure divergent likelihoods and antilikelihoods. Two approaches, I think, have considerable appeal. The first argues that we reconsider what we count as belonging to K, the background evidence, and, I think, in so doing abandons the Predesignation Principle. The second argues that we reconsider the further evidence "E". On the first strategy K is to be identified counterfactually as "that which he would know had he not learnt E". Thus, let K' be what would have been known had Kepler's laws not been known in Newton's time. Relative to this background knowledge, Pr(Kepler's laws/Newton's theory) (likelihood) and Pr (Kepler's laws/Newton's theory false) (anti-likelihood) would dramatically diverge. The crucial issue is the presence of the counterfactual. This strategy finds no difficulty in ascribing evidential force (at least sometimes) to accommodated evidence.

The second strategy focuses on K. While Kepler's laws should remain in K, the fact that Newton's theory ***entails*** Kepler's was not in Newton's time part of K. Given that Kepler's laws are, as it were, in the domain of Newton's theory, it is, presumably, far more likely that Newton's theory will entail Kepler's laws on the assumption that that theory is true than on the assumption that it is false. Hence, we again get the desired divergence of likelihood from anti-likelihood This allows the Bayesian to retain the Predesignation Principle, for it allows no evidential force to known facts per se. But it does involve him in revising a standard Bayesian doctrine, that logical truths (such as the above entailment) be assigned a probability of one—which, of course, would regenerate the problem. That is, the standard Bayesian holds to a doctrine of logical omniscience which would now have to be abandoned and the consequent revision would have extensive ramifications.

The Popperian involvement in the issue has a number of related sources. It ties in most closely with the Bayesian perspective if we attend to Popper's requirement of severity of test. A test is severe only if the predicted result is improbable in the light of background evidence, i.e. we have low anti-likelihood. But Popper's arguments against "justificationism", his rejection of "ad hoc" hypotheses and, especially in Lakatos's hands, the central role of novel prediction all point in the direction of the "Predesignation Principle". The strategies have typically sought to discriminate between contexts in which ***known facts do***, and those in which they ***do not have*** any evidential force. On one view, background evidence consists in whatever is known to the science of the time, and thus would exclude Kepler's laws from constituting any confirmation (survival of severe test) of Newtonian theory. This gives as objective a criterion of severity of test as we have of "known to science" and constitutes a strong version of our principle. Another view, perhaps keeping closer to the ostracism of ad hoc theories, requires that the known fact not be part of the ***problem-situation*** for which the theory was designed. This involves a person-relative notion of background evidence and hence of severity of test. However it gains a more flexible approach to accommodated evidence.

Musgrave[7] adopts a rather different approach, drawing on the work of Lakatos. If we take as our measure of severity of test—our anti-likelihood—the probability of the evidence given the best available ***alternative*** hypothesis and hence look upon evidential value as necessarily involving a ***comparison of hypotheses,*** then known data has, as such, Musgrave argues, no peculiarly inferior status. If 't', the touchstone theory, is incompatible with, or fails to explain E, then whether E is known before, or predicted by, H is of no interest. Lakatos, and I think Musgrave, would be prepared to speak of a severe test in terms of a large difference between likelihood and anti-likelihood (in terms of the touchstone theory). Whether this can be done without importing more background evidence, hence raising the difficulty of what it can contain, is unclear.

I think we should also note George Schlesinger's[8] constructive article, "Accommodation and Prediction". Schlesinger, holds that a preference for predicted evidence is founded on ***an important evidential principle.*** This principle, however, is independent of the prediction/accommodation contrast. When a prediction E from H proves true, then H is such that it entails the further evidence ***without any amendment.*** It is this idea of "explaining without further amendment" that is crucial. Its true force can be brought out by appeal to the notion of simplicity. A theory may be confirmed by the data it is designed to accommodate, but this is not just a matter of its being the simplest theory to accommodate the evidence. It is further required that it be the simplest accommodating hypothesis with regard to at least some lesser portion of the evidence. That is, it is required that it both be the simplest hypothesis to cover some lesser portion of the evidence and *without amendment* be the simplest hypothesis to cover the total evidence. If the simplest adequate theory successfully predicts new evidence, then it fulfills Schlesinger's criterion. But the criterion equally applies in the context of accommodation.

II.

Peirce's 1883 paper "A Theory of Probable Inference" is an extraordinarily fertile and prescient discussion of the nature of empirical reasoning. At the centre of his thinking in this matter is his conviction that all empirical reasoning is essentially reasoning from a sample to a population. Empirical generalisation is just the special case where we infer a 100% frequency of a character in the population. Inference to a scientific theory is an inference from, in effect, a sampling of its logical consequences. Statistical inference can only yield a conclusion that is *merely probably* (never certainly) true and *merely approximately* (never exactly) true. Peirce marks as a fundamental distinction that between what he calls statistical deduction—inference from population to sample—and statistical induction—from sample to population. He holds that while both yield probable and approximate conclusions, the sense of 'probably' is different in each. In the first, it is a matter of long run frequency, whereas in the second it is linked to Peirce's conception of the self-corrective nature of inductive reasoning. In another paper, he develops what is, in effect, Carnap's distinction between state-description and structure-description. Yet he also manages to display similarities to Popper most strikingly in "The Scientific Attitude and Fallibilism".

> . . . it is a great mistake to suppose that the mind of the active scientist is filled with propositions which. . . are at least extremely probable. On the contrary, he entertains hypotheses which are almost wildly incredible. . . The

> best hypothesis. . . is the one which can be most readily refuted if it is false. This far outweighs the trifling merit of being likely. For what is a likely hypothesis? It is one which falls in with our preconceived ideas.[9]

Peirce has much to say on empirical reasoning beside presenting this criterion of validity. Prominent is his "predesignation principle" requiring that a hypothesis be specified prior to evidence being sought for it. As a close reading of the text reveals, Peirce's notion of predesignation is more subtle[10] than this, but, as this subtlety leads into some major difficulties of interpretation not immediately connected with our current concerns, I return now to the central theme of this paper: Peirce's defence of this principle.

III.

Even such staunch defenders as Hempel and Carnap of the "logical" approach, as Musgrave calls it, did not hold that inductive reasoning could be fully understood in terms of formal relations between evidence and hypothesis. Carnap recognised, for example, the need for such principles of methodology as the "requirement of total evidence"[11]. While formal principles determine the degree of support certain evidence K gives to a certain hypothesis, H, this alone does not determine what confidence a person should have in that hypothesis. Should the reasoner be in possession of further evidence E, relevant to H, then his/her confidence should be in accordance with the support his total relevant evidence K&E gives to H. Such a rule bears on the correctness of an inductive conclusion, though not by bearing on the evidential relation itself. It might appear then that we have here a possible resolution of the historical and logical accounts of the role of prediction, at least on the assumption that the notion of a "methodological rule" is a sound one. If the superiority of predicted evidence is akin to the superiority of total evidence, then this would leave untouched the conception of support or confirmation as a logical relation between evidence and hypothesis.

Peirce appears to ***equate***[12] the total evidence requirement and the Predesignation Principle. It seems to me there are grounds for claiming an analogy. A person who undertook to test whether exposure to electromagnetic fields caused leukemia might take numerous large random samples of TV repairmen until he found a sample with a significantly high proportion of leukemia victims. He then publishes only the results for this sample—strikingly supporting his hypothesis. He has plainly behaved improperly. You might say his evidence is ***selective*** or ***tailor-made*** to fit his hypothesis. Now suppose a researcher who gathers a single large random sample of TV repairmen, not with any prior hypothesis in mind, and combs his data until he finds some disease significantly more common than in the general population. He then publishes the result without declaring how he arrived at his "hypothesis". Again we might find impropriety in his proceeding. He has, one might say, "tailor-made" his hypothesis to fit his data. Consider enough samples and you are bound to find one that supports a given hypothesis; consider enough hypotheses and you are bound to find one supported by a given sample. At times, Peirce appears to see these requirements in a similar light to Carnap. They are not principles that determine the strength of the evidential relation between a sample of a given size and nature and a hypothesis about the nature of the population from which it is drawn. Rather, they bear on the proper use of these relations in arguments or reasoning.

Peirce's equation of the Predesignation Principle with the total evidence requirement is present in the general theory of probable inference. The transition is an immediate one—Peirce appears to see no need for defense or explanation. The only interpretation at all adequate is to equate it with Glymour's criticism of Bayesians. That is, where H is not predesignated, the only way we can get an antilikelihood other than 1, is by ignoring some of our total relevant evidence. But this defence of the transition from the uncontroversial total evidence requirement to the controversial Predesignation Principle depends on a particular interpretation of the anti-likelihood ruling, for example, a counterfactual approach out of the question. But the basic difficulty is this: the justification of the total evidence requirement is not something which can be achieved by appeal to Bayes Theorem itself—it is rather a methodological rule about, for example, the application of Bayes Theorem. (Its justification is closely related to the problem of induction.) Failure to observe the total evidence requirement cannot be shown unreasonable *just by appeal to* Bayes Theorem, yet ignoring its alleged equivalent, the Predesignation Principle *is* alleged unreasonable by Bayesian standards (for its failure to secure low anti-likelihood).

In any case Peirce appears to offer a distinct, substantial argument for predesignation elsewhere. This too centres on the failure to secure low anti-likelihood or, as the conclusion of the paragraph has it:

> if the character P were not predesignate, we could not reason that if the Ms did not generally possess the character P, it would not be likely that the S's should all possess this character.[13]

The next section explores the argument to this conclusion.

IV.

Peirce's argument here appears different from the thesis that all likelihoods involving known evidence have a probability of 1. Rather it is an argument which bears a striking resemblance to the work of Nelson Goodman. It is captured in the passage:

> . . .suppose we were to draw our inferences without predesignation of P, then we might in every case find some recondite character in which those would all agree. That, by the exercise of sufficient ingenuity one should be sure to be able to do this, even if not another object of the class M possessed that character, is a matter of demonstration. . .[14]

He illuminates this by finding mathematical characteristics common to all of a sample of poets' ages at death, the projection of which to further cases he declares totally unreasonable. The problem is to determine just how this destroys the evidential force of accommodated data by destroying the appropriate divergence of likelihoods. We might illustrate Peirce's argument in the following way. Suppose we know that both Sam and Mabel have arrived at a certain hypothesis on a certain matter. They have, however, followed different instructions in doing so. Mabel has been given information K and instructed to provide an hypothesis with a true prediction on whether claim E in the domain of the hypothesis is true. It turns out her prediction is correct. Sam is given ***both*** E and K and asked to provide an hypothesis ***accommodating*** this total evidence. Given only the information that both have succeeded, can we outsiders have any rational preference for Mabel's hypothesis over Sam's? Presumably, the advocate of predesignation should favour Mabel's hypothesis. Peirce is offering an argument in favour of that preference. In the light of the passage just quoted, Sam's success is a foregone conclusion—unaffected by whether he selects the true hypothesis or not. In terms of likelihoods,

Pr(Sam accommodates E&k/Sam selects true hypothesis)=1
Pr(Sam accommodates E&k/Sam selects false hypothesis)=1

There is no significant probability of Sam selecting an hypothesis that does not accommodate the data. Mary's success however is very sensitive to whether she selects the true hypothesis—if she selects a false hypothesis she is unlikely to provide a true prediction. In terms of likelihoods,

Pr(Mabel selects an hypothesis that success fully predicts E/ selects true)=1
Pr(Mabel selects an hypothesis that successfully predicts E/ selects false) is low.

Hence, Mabel's achievement constitutes evidence for believing she has selected the true hypothesis, Sam's does not.

This line of thought sometimes yields counterintuitive results. Consider the following game ("Series game") which bears some similarity to the decoding example with which this paper commenced. Suppose Sam and Mabel are both invited to determine a sequence of twenty single-digit numbers. Sam is told the first fifteen (K+E). Mabel is told the first ten (K) and asked to predict the next five (E) and she does so correctly. Ought we to be more confident that Mabel rather than Sam can specify the entire twenty numbers correctly? By the argument of the previous paragraph we should do so. Moreover, that argument is left undisturbed if we add the following crucial datum. The sequence has in fact been generated by a randomising machine. Can we continue to favour Mabel's hypothesis? Of course, Mary's successful prediction makes it more likely that she has selected the true hypothesis, but her success *also* depends on how likely she was to select the true hypothesis in the first place. Here she is *disadvantaged* with respect to Sam, for she knows only the first ten of the sequence whereas Sam knows fifteen.

There is some very large number of single digit twenty member sequences. Suppose that the evidence of the first ten reduces this further to N sequences. We should reason that the probability that from the N sequences that accommodate the first fifteen numbers (there being no question of his choosing any other) he will pick the true one. A probability of 1/N is appropriate as a measure of Sam's having selected the correct sequence given his accommodating success. Mabel starts off in an evidentially worse position, knowing only K and hence assuming she selects only from accommodating hypotheses of which there are M, there is only a 1/M probability that she will select the true sequence as a basis for prediction. But we also learn that she has managed to select an hypothesis which correctly predicts E. Hence we must infer that she has selected one of the hypotheses that entail E&K. Mabel, equally then, has a 1/N probability, in the light of her predictive achievement, of having selected the true hypothesis.

Hence I conclude that in "Series Games" prediction has no evidential advantage over accommodation. My version of Peirce's argument does not discriminate such cases and hence is wrong as it stands. The following and final section considers what marks off Series Game cases and argues for an additional factor needed for the Predesignation Principle to look plausible.

V.

The Series Game is purely a game of skill in accommodating data. That is, in normal circumstances, no player can call into play any other skill in discerning the true hypothesis. Not all situations in which we seek to discern true hypotheses are like this. Whether in scientific inquiry, detective work, judicial judgement or whatever, while most individuals have little trouble providing an hypothesis which accommodates given data, there are some who prove far more successful in discerning, on the same evidence, true hypotheses. Indeed, a person might be, relatively, a good discerner of true hypotheses (say a 50% success rate in some area being far above his competitors) yet, in that area, a relatively poor accommodator of data—through poor calculation getting a mere 50% success rate (compared say with 90% for his/her fellows). Of course, he/she cannot be, percentage wise worse at accommodation than truth-discerning—at least as long as we assume a truth-hypothesis must accommodate the "known" data. But if we allow the correction of such data—he may even be a better truth discerner than he is an accommodator, i.e. his judgements on hypotheses force a revision of evidence!

Peirce, amongst others, has stressed that unless we have such a propensity to discern the true from within the accommodating hypotheses, inquiry could not have progressed. If our selection of hypotheses for testing were an utterly random selection from within all accommodating hypotheses, this being an indefinitely large set, we would have a negligible chance of selecting one anywhere near the truth. Peirce, in "A Theory of Probable Inference"[15] likens this to Martians seeking to generalise from U.S. census statistics.

Suppose it is then granted that Sam and Mabel are undertaking their respective tasks ***in a context*** where the possibility of truth-discernment as distinct from mere accommodation is not precluded (as in Series Games). Should this work in favor of preferring Mabel's hypothesis? The point is this. No amount of accommodating skill on Sam's part can *by itself* constitute any evidence that he has and has used such truth-discernment. However, in the absence of overwhelming evidence to the contrary, repeated predictive success must constitute evidence for a capacity to discern truth more finely than the accommodating skill secures. Hence, a single predictive success constitutes ***some*** evidence for the possession and use of such truth-discernment, and thereby increases the probability that Mabel has selected a true hypothesis.

That is, while
Pr.(Sam accommodates E&K given Sam is a truth-discerner)= Pr (Sam accommodates E&K given Sam is not a truth-discerner)=1.

this contrasts with
Pr(Mabel predicts E given Mabel is a truth-discerner)
is greater than
Pr(Mabel predicts E given Mabel is not a truth-discerner).

That is, Mabel's predictive achievement is better evidence for her hypothesis (***where*** it is) ***indirectly*** by supporting her having such truth-discernment power. Sam's accommodatory achievement can never have such an effect.

University of Melbourne

NOTES

1. W. Whewell, *On the Philosophy of Discovery* (London: John W. Parker, 1860), p. 274.

2. J.S. Mill, *A System of Logic*, Book III, Chapter XIV, Section 6, in J.M. Robson (ed.), *The Collected Works of J.S. Mill*, Vol. VII (Toronto: Univerisity of Toronto Press, 1973).

3. J.M. Keynes, *A Treatise on Probability* (London: Macmillan, 1963), p. 305.

4. R. Carnap, *Logical Foundations of Probability* (Chicago and London: University of Chicago and Routledge and Kegan Paul, 1950), p. 211.

5. A. Musgrave, "Logical Versus Historical Theories of Confirmation", *British Journal for Philosophical of Science* (25), 1974, pp. 1-23.

6. C. Glymour, *Theory and Evidence* (Princeton: Princeton University Press, 1980), pp. 75ff.

7. A. Musgrave, "Logical Versus Historical Theories of Confirmation", *British Journal for Philosophy of Science* (25), 1974, pp. 1-23.

8. G. Schlesinger, "Accommodation and Prediction", *Australasian Journal of Philosophy*, (65) 1987, pp. 33-42.

9. C.S. Peirce, *Collected Papers*, ed. C. Hartshorne & P. Weiss, Vol. I, paragraph 120.

10. Cf. C.S. Peirce, *op. cit.* Vol. II, paragraph 739.

11. R. Carnap, *op. cit.*, paragraph 45.

12. C.S. Peirce, *op. cit.*, Vol. II, pargaraph 735.

13. C.S. Peirce, *op. cit.*, Vol. II, paragraph 737.

14. *Ibid.*

15. *Op. cit.* Vol. II, paragraph 752-753.