RICHARD C. JEFFREY

Carnap's Empiricism

"Our knowledge is of matters of fact and of relations of ideas. Logic — inductive and deductive — concerns relations of ideas. As to our factual knowledge, some of it we have direct, through the senses, and the rest we have from the direct part, through logic." This is a rough sketch of Carnap's empiricism. Let me now try to fill it in, to get a fair likeness.

1. The Confirmational Net

For at least his last thirty years, this much of Carnap's view was constant: Inductive logic ought to be representable by a network, the nodes of which are the sentences (or, indifferently, the propositions expressed by the sentences) of a formalized, interpreted language. The nodes of the net are connected by arrows, each of which has a number associated with it, viz., the degree of confirmation of the sentence at the arrow's head, given the sentence at the arrow's tail. Thus, the configuration $e \xrightarrow[x]{} h$ means that $c(h,e) = x$, where c is the confirmation function we shall eventually adopt, and e and h are sentences in the explicitly structured language we shall eventually adopt, either in actual practice or as a "rational reconstruction" of our wayward linguistic practice.

One's beliefs at time t ought to be representable by a credence function, cr_t, which assigns to each node h a numerical value $cr_t(h)$ in the interval from 0 to 1. Carnap's paradigm was the case in which there is a sentence e_t which expresses the part of our knowledge which we obtain at time t directly, through the senses: the experiential component. If we define

$$e(t) = e_1 \cdot e_2 \cdot \ . \ . \ . \ \cdot e_t$$

AUTHOR'S NOTE: I wish to express gratitude to the National Science Foundation for support of my research on probability and scientific method. This paper forms a pair with Jeffrey (1973), which it partly overlaps.

then one's entire credence function at time t ought to be determined by the formula

$$(1)\qquad cr_t(h) = c(h,e(t)).$$

In other words, one's degree of belief in any sentence h ought to be the number associated with the arrow to h from the sentence $e(t)$. One might imagine one's degrees of credence in the various nodes of the net at time t to be represented by luminosities: Where $cr_t(h) = 1$, node h shines with a certain maximum brightness; where $cr_t(h) = 0$, node h is quite dark; and credences between 0 and 1 are represented by intermediate luminosities. In these terms, $e(t)$ might be regarded as the *origin* of all the light in the net at time t. The number associated with the arrow from $e(t)$ to h indicates the portion of the full luminosity of $e(t)$ which the net transmits to h. As t increases, some parts of the net increase in brightness while others dim.

There is a point to the image of the twinkling, luminous net. One can see that different nodes have different brightnesses at time t, even though the luminosities represent someone else's beliefs. Dropping the metaphor, this "seeing" is a matter of knowing the fellow's *preference ranking* at time t: If he would rather have h be true than have h' be true, then his preference at time t is for h over h', and we write hP_th'. If the relation P_t has certain properties (Bolker, 1967) there will exist an expected utility function eu_t where

$$(2)\qquad hP_th' \textit{ iff } eu_t(h) > eu_t(h')$$

from which one can derive the credence function cr_t by the relation

$$(3)\qquad cr_t(h) = \frac{eu_t(-h) - eu_t(h \vee -h)}{eu_t(-h) - eu_t(h)} \quad \textit{if } hP_t{-h} \text{ or } {-h}P_th.$$

(In case the condition of applicability of (3) fails, use the method of section 7.3 in Jeffrey, 1965.) It is (3) which determines the *actual* credence function at time t of the man whose preferences at that time are given by the relation P_t. It is (1) which gives the *rational* credence function at time t of the man whose actual observational evidence up to time t is $e(t)$. In using the symbol 'cr_t' for both functions, I am assuming that the actual credence function is as it ought to be, given what the man has observed.

Now let us make a slight alteration in the image of the net, prompted by the fact that logically equivalent sentences are confirmationally indistinguishable. We write '$[h]$' for the logical equivalence class to which h belongs, i.e., for the set of sentences which are logically equivalent to h. Imagine that all nodes (= sentences) in $[h]$ are gathered together into a supernode. All arrows between nodes in the same supernode have the number 1 associated with them, and all arrows from one supernode to another will have the same number associated with them. Now the nodes of *the reduced net* will be the supernodes of the original net, and we shall have the configuration $[e] \underset{x}{\rightarrow} [h]$ in the reduced net if and only if we had $e \underset{x}{\rightarrow} h$ in the original net. From now on, "net" will mean *reduced net*.

The structure of the (reduced) net is static. (By "structure" I mean not only the configuration of arrows, but also the numbers associated with the arrows.) In the cases which Carnap has discussed in his publications, the play of light from the nodes is to be explained by (a) the trajectory of the points $[e(1)]$, $[e(2)]$, . . . , at which experience pumps light into the net, and (b) the static structure of conductive arrows through which the light of experience is propagated throughout the net. (The number associated with an arrow indicates its conductivity.) In this way, the experiential element in knowledge, (a), is neatly separated from the inductive element, (b), which Carnap viewed as purely logical.

2. In What Sense Is c "Logical"?

What did Carnap mean by his repeated assertions that his c-functions are purely logical?

I think he had two things in mind, the first of which is unassailably true, but perhaps a disappointment: If the net is to be static, we ought to identify our chosen c-function in such a way that, once e and h are identified either ostensively (as in quotation-mark names) or by structural descriptions, it is a matter of calculation to determine the value which c assigns to the pair (h,e) — no empirical research should be needed. In Carnap's terms, the values which c assigns to such pairs ought to be determined by the semantical rules of the metalanguage in which such sentences as 'c^*('Pa_1', 'Pa_2') $= 2/3$' appear. If 'M' is the name of that metalanguage, all such sentences should be M-determinate, e.g., the one just given is M-true.

Thus the first thing Carnap had in mind, in speaking of 'c' as a *logical functor*, was a requirement about the form of definition of c, e.g., it would be all right to define it by using a computing machine program which, given descriptions of h and e, eventually prints out the value of $c(h,e)$; but it would not do to define c by some stipulation which, when we seek the value of 'c('P_1a', 'P_2a')', instructs us to discover the limiting relative frequency of P_1's among P_2's, even though (as luck would have it) the computing machine program mentioned above always gives the very answers to which the unacceptable sort of stipulation would lead us to by way of empirical research. The *function* c would be the same in both of these cases, being describable by a certain set of ordered triples. But in the first case, the functor which has that function as its extension is logical, while in the second case the functor is factual.

Carnap thought this point important, and he argued for it by inviting us to imagine the trajectory $[e(t)]$, $[e(t-1)]$, . . . , traced back to its source, in a primal state of innocence in which our experience is null: $[e_0] = [p \vee -p]$. The structure of the net is constant, throughout the trajectory, so that the numbers associated with the arrows are the same at time 0 as at all successive times. But at time 0 one had no experience, and therefore the values of cr_0 and, with them, the values of c, must be independent of experience: they are logical, not factual. Of course, this argument will have no weight with someone who rejects the picture of the static net. Its point is rather to show that once you accept that picture you are committed to a definition of c which makes it a matter of computation, not experience, to determine the numbers $c(h,e)$.

This aspect of Carnap's claim that 'c' must be a logical functor is then a rather trivial truth about Carnap's program of representing inductive logic by means of a static confirmational net. Read in this way, the claim is disappointing because it leaves us with the question, "Granted that the form of definition ought to be 'logical,' how are we to identify the *right* c-function, among the infinity of functions you have already considered, and the infinity that you have not yet considered?"

Here, Carnap's answer is available in his paper on inductive intuition (Carnap, 1968) and in parts of his paper entitled "A Basic System of Inductive Logic" (Carnap and Jeffrey, 1971). I would simply add that if the program is carried on and succeeds in the sense that we eventually find a language as well as a confirmational net on that language, which conforms to the clear cases of our inductive practice and resolves the unclear ones in a way that convinces people, then in the finding we shall have sharpened and developed and tuned and fixed our inductive intuitions far beyond anything we have now, and the resulting system will of course be describable by the net we shall have found. The corresponding c-function (or small family of c-functions, all of which seem to do the job about equally well) will then be uniquely "logical" in the sense of according with our inductive logical intuitions, as they will then exist.

I conclude that in its first aspect Carnap's claim that 'c' is logical is a clear, analytic truth about the character of his program for developing inductive logic; and in its second aspect his claim has the same truth-value as the statement that in time Carnap's program will be successfully carried out. (If and only if the truth-value is *t*, we shall eventually have highly developed inductive intuitions which are describable by way of a c-function.)

3. Experience and Confirmation: Correspondence with Carnap

In 1957 I finished my doctoral dissertation and sent a summary to Carnap, who expressed interest in my treatment of probability kinematics, viz., an alternative to the use of the trajectory $[e_t]$ $(t = 1, 2, \ldots)$. (See Jeffrey, 1965, chapter 11.) His own, very different ideas on the subject can be seen in the following correspondence.

From Carnap, July 17, 1957

To the first point: Your *rejection of conditionalization.* I believe that this idea of yours has a certain relationship to an old idea of mine; I never worked it out because I was never able to find a satisfactory solution. . . .

My idea was as follows. The c-method (i.e., the rule: if e represents all observational results of A then it is reasonable for A, to believe in the hypothesis h to the degree c(h,e)) is an oversimplification because it proceeds as if A knew certain observational results with certainty, while in fact we can never have absolute certainty with respect to a factual sentence. This simplification is useful to a certain extent, like many other schematizations in science. But I thought that it would be worthwhile to search for a method which takes into consideration the uncertainty of the evidence. . . . Thus I considered the following problem: The person A should represent his evidence (at a given time point) not simply as a list or conjunction of observational sentences, but rather as a list of such sentences together with a number attached to each sentence, where the number is to indicate the subjective certainty of the sentence on the basis of the observational experience. Thus the new form of evidence E may [be as follows]:

$$E = \{e_1, b_1; \ldots ; e_i, b_i; \ldots \}, \text{ where } 0 < b_i < 1.$$

Let e_1 be "the ball just drawn is red." If A could look at the ball for some time with good illumination and under otherwise normal circumstances, then b_1 is close to 1. If on the other hand A could see the ball only during a short moment and the illumination was not good or the air was misty or otherwise the circumstances were not favorable for an observation, then b_1 may be rather low. The important point is that b_1 represents the certainty of e_1 merely on the basis of the observational experience, without regard to any inductive relation to earlier observations. My problem now was to find a belief function cr(h,E) which is to determine the rational degree of belief (credibility) for any sentence h of a given simple language L, for which we have a c-function. My guess was that cr(h,E) somehow was to be defined on the basis of the numbers b_i and the values of c(*h*,—) with respect to the e_i or combinations of such. Many years ago I explained this problem to Helmer and Hempel (in connection with their papers with Oppenheim ca. 1945) and later, when I was in Princeton, to Kemeny and Putnam. As far as I know, none of us has made any serious efforts to find a solution. I was discouraged from continuing the search for a solution by the fact that very soon I found some difficulties and did not know how to overcome them. One of them arises even before sentences outside of E are considered. The difficulty lies in the question how to determine the value of $cr(e_1,E)$? The value cannot simply be taken as equal to b_1 because the value must also take into account the inductive influence on e_1 of the other sentences in E, and this influence is not taken into account in the number b_1. Suppose that $b_1 = 0.8$ (for example, because the illumination was not quite good). Even if we assume for the other sentences of E the simplest case, namely that every b_i (for $i \neq 1$) is so close to 1 that we can take it as approximately = 1, the difficulty still remains. Let E′ represent these other sentences of E. Let $c(e_1,E') = 0.9$ (E′ may for example describe a great number of balls drawn from the urn of which almost all are red). Then $cr(e_1,E)$ is neither simply $= b_1 = 0.8$, nor simply = 0.9. I would guess that it should be > 0.9, since E contains in comparison with E′ an additional observation about e_1 and this observation, although it has only $b_1 = 0.8$ nevertheless represents an additional favorable item of evidence for the truth of e_1.

When I read your letter I had first the impression that your a's were about the same as my b's, and that therefore your method was perhaps a solution of my problem. It is true that you say only: "Now suppose A changes his mind about e . . . : A's degree of belief in e changes from m(e) to a" without specifying the reasons for A's change of mind. I presumed that you had in mind a change motivated by an observation, or at least that this was one possible case of a rational change. But now I doubt that your a's are the same as my b's. For, if I understand you correctly, a_1 is the degree of belief in e_1 after the change, in other words, after the observation, hence, in my terminology $cr(e_1,E)$ which is different from b_1.

My main question now is this: does your theory still somehow supply the solution to my problem . . . ? If not, what is the rule of application which your theory would give to A for the determination of the new value a_1 for the sentence e_1? It seems to me that any practicable system of inductive logic must give to A the following: (1) rules for choosing a function m [viz, cr_0] as the initial belief function . . . and (2) rules for A to determine the rational change of his belief function step for step on the basis of new observations which he makes.

To Carnap, September 4, 1957

Let $c_i = cr(e_i, E)$. The difficulty you describe . . . is: to determine c_i as a function of b_1, b_2, ... As you remark . . . it is your *c*'s rather than your *b*'s which correspond to my *a*'s, i.e., to the person's actual degrees of belief in the *e*'s after his observations. Actually, the *c*'s will not be identical with the *a*'s unless (i) *E* reports all the observational experience the person has ever had which is relevant to the *e*'s, (ii) the person in question is rational, and (iii) there really is a function cr which uniquely describes *rational degree of credibility*.

Now the *a*'s are measurable by observing the believer's behavior, e.g. by offering to bet with him on the *e*'s at various stakes, immediately after his observations, and noting which bets he accepts. But the *b*'s are high-level theoretical terms, relative to the *a*'s. I am inclined to turn your question around, and ask, not "How are the *c*'s determined by the *b*'s?" but rather "How are the *b*'s determined by the *c*'s (or, actually, by the a's)?"

This question is not factual; it asks for an explication, and might be rephrased: "How shall we define the *b*'s in such a way that they will adequately explicate that component of the rational observer's belief in the *e*'s which corresponds to what he has seen with his own eyes?"

From Carnap, December 26, 1957

I am still thinking about the question which I raised in . . . my letter of July 17: which rule does your inductive logic give to the observer in order to determine when and how he should make a change according to your method? You are indeed right that the customary c-method has the disadvantage that it refers only to certain situations which can never be completely realized, namely cases where the observer knows the result of an observation with certainty. But the method gives at least a clear rule for these idealized situations:

(3) When you experience an observation o_i which is formulated by the sentence e_i, then add e_i to the prior evidence and take as rational degree of belief in any sentence h its c-value with respect to the total evidence.

You have so far not given any rule. You emphasize correctly that your a_i is behavioristically determinable. But this concerns only the *factual* question of the *actual* belief of A in e_i. But A desires to have a rule which tells him what is the *rational* degree of belief. One part of the rule should presumably be as follows:

(4) Make a change with respect to e_i if and only if you have made an observation whose result is formulated by e_i.

But then a rule of the following form must still be added:

(5) If the prior degree of belief in e_i is m_i, then take as its posterior value after the observation the value $m'_i =$ . . .

The new value m'_i might depend upon the clarity of the observation(or the feeling of certainty connected with it or something similar) and further on m_i. . . .

P.S. Now some comments on your letter of November 14. . . . You are right with your interpretation of Neurath's paradoxical denial of the comparison of sentences and facts; I criticized this already in "Wahrheit und Bewährung" (1935); partly translated in Feigl and Sellars, *Readings*, 1949. Your idea of clarifying the situation of observing and recording by imagining a machine doing this, seems to me very helpful. But I still think it might be useful to regard the deposits of the observational results in the memory organ of the machine as sentences (in the interior language of the machine). My coefficients b_i might then perhaps enter in this way: the machine deposits in its memory after each observation o_i not only the corresponding observation sentences e_i but also a number b_i which depends only on the circumstances of the observation (e.g., in the case of a visual observation, on the intensity of the illumination, the sharpness of the picture on the screen, etc.) but not on the results of earlier observations.

To Carnap, January 16, 1958

I think this special case [viz., what I called "conditionalization" and Carnap called "the customary c-method"] is an important one. In physics, for example, there are very high standards which an event must meet in order to count as an observation, and there is perhaps not much variation in degree of reliability among such events. At least, this is true if by *observation* one means observations of the positions of needles on meters or such events. But if one thinks of the observer as consisting of the human being together with a certain amount of apparatus, e.g. together with an oscilloscope, and thereupon speaks of "observing" the presence of harmonics superimposed on a sine wave of voltage (as engineers, at any rate, are prone to do), the situation is different. I am intrigued with this broader sense of "observation." It seems that in physics, one can dispense with it, and speak instead of observations in the narrower sense, to which we apply portions of physical theory in order to derive or induce the "observation" in the wider sense. But where the science is not so highly developed as physics is, it is difficult to dispense with "observations" in which a great deal of interpretation is mixed up; e.g. in psychology we might say we observe that someone is angry, or we might say, more cautiously, that we observe that the person flushes, trembles, and curses, and add that there are laws which say that when these symptoms are exhibited, the subject is usually angry. Empiricists have usually tried to isolate the empirical element, i.e., the observational element, by analyzing "observations" in the broad sense into observations in the narrow sense plus theory. But I'm not sure this is the best way. When we learn, as children, to use words like "angry," we do so in much the same way in which we learn to use words like "red" and "flush." So I am inclined to think that the difference between "angry" and "red" — as far as their relation to experience is concerned — is not that one is theoretical while the other is observational, but rather that one term can more reliably be attributed to a subject, on the basis of observation, than the other. But they differ in degree, not in kind. It is desirable, where possible, to be able to reduce observation reports to terms like "red," which are highly reliable; but I doubt if this is always possible, and I expect that if we study only those situations in which such reduction is possible, we are apt to get a distorted notion of meaning.

From Carnap, March 6, 1958

I am still inclined to take as primitive simple properties like "red" rather than complex ones like "angry." With respect to complex properties, great difficulties would arise for the question of dependencies; such dependencies would then have to be expressed by complicated meaning postulates. And for the inductive treatment of two families [where a family is a set of properties P^i_1, P^i_2, . . . , which are mutually exclusive and collectively exhaustive on logical grounds] it is quite essential that, e.g., P^1_1 be simpler than $(P^1_1 \cdot P^2_1) \vee (P^1_1 \cdot P^2_2)$, although both properties have the same logical width. If this condition of simplicity were not fulfilled, there would not be a good reason for the axiom of analogy A16 (Notes, p. 29) [viz., U.C.L.A. "Lecture Notes on Probability and Induction"].

4. Prospects for Carnap's Program

To some extent, I have exaggerated the futuristic aspects of Carnap's program for developing an account of induction based on a static confirmational net. *He* was quite prepared to consider the question, "What c-function is appropriate for a language in which the primitive predicates are organized into a finite number of finite families?" Such a language would be very far from the very expressive languages for which he hoped that suitable c-functions would eventually be discovered. He would consider such simple languages rather abstractly, seeking to define families of c-functions for them. The family of c-functions for such a language would be parametrized by a number of adjustable constants, whose values would be determined by the interpretations chosen for the primitive predicates. Examples of such parameters: the *logical widths* of the primitive predicates; the *logical distances* between them.

Carnap, of course, sided with Humpty Dumpty: "When *I* use a word, it means exactly what I choose it to mean — neither more nor less." Others (e.g., Quine and Alice) attacked this linguistic Jacobinism from the right: "The question is," said Alice, "whether you *can* make words mean so many different things." To this, Carnap would reply, with Humpty Dumpty: "The question is, which is to be master — that's all."

Consider the case in which a small child thinks there may be an elephant in the next room. He stands in the doorway and looks about, sees nothing out of the ordinary, but still thinks there may be an elephant there, e.g., behind the easy chair. Quine points out that there is no telling whether the child's ignorance of the sizes of elephants reveals lack of factual knowledge or lack of semantical understanding: it is only by an arbitrary stipulation that we can see it as a matter of definition (or, alternatively, as a matter of fact) that elephants are rather large, compared with easy chairs. But that did not trouble Carnap, who was quite prepared to make such stipulations, after comparing the merits of making them one way or another; he was a linguistic revisionist, and quite unabashed about it. He was not given pause by the fact that the distinction between primitive and defined terms is vacuous in connection with natural languages. He was prepared to impose and police such distinctions as part of the business of reconstructing language closer to the mind's desire. The instrument which he sought to forge was a unit consisting of a language and a c-function, fused. It was no part of his claim that either element now lies ready to hand, fully formed.

For my part, I subscribe to a form of right-wing deviationism called "subjectivism." Subjectivists, like Carnap, use relations like (3) to determine the actual credence function, cr_t, at time t. Thus, subjectivists can measure the luminosities of the various nodes. Using the relationship

$$c_t(h,e) = \frac{cr_t(h \cdot e)}{cr_t(e)} \text{ if } cr_t\ (e) > 0$$

they then determine the number which is to be associated with the arrow from [e] to [h] *at time t*: subjectivists take a dynamic view of the net. So may Carnap, if he wishes, for $c_t(h,e)$ is simply $c(h,e \cdot e(t))$ if the subject is being completely rational. But subjectivists do not assume that there is an underlying, static net corresponding to the function c. Nor need they assume that the sentences in the nodes of the net belong to a rationally reconstructed language. One can indeed follow the sequence c_t, c_{t+1}, . . . , some distance forward under favorable circumstances: perhaps the change from c_t to c_{t+1} is clearly to be explained by an observation which convinced the subject that e_{t+1} is true, so that we have

$$c_{t+1}(h,e) = c_t(h,e \cdot e_{t+1})$$

Perhaps, too, one can follow the path backward some distance — but not, on the subjectivistic view, back to an a priori c-function c_0 which determines an underlying static net of the Carnapian sort.

Over his last ten or fifteen years, Carnap kept reporting encouraging signs of movement: subjectivists like L. J. Savage were getting closer to his position. For his part, Savage explained the relative motion as originating in Carnap. Indeed, from a great enough distance, e.g., from the point of view of K. R. Popper, Savage and Carnap were Tweedledum and Tweedledee. Perhaps, indeed, the difference is largely temperamental: you are a subjectivist if, like Carnap, you take credence to be determined by (3) but, unlike Carnap, boggle at the idealization (or falsification) which is involved in the attempt to trace the sequence c_t, c_{t-1}, . . . , all the way back to an underlying static net, corresponding to the a priori function $c_0 = c$. It is not that subjectivists, rejecting the a priori c-function, are better empiricists than Carnap. On the contrary, they lack the means to formulate empiricism as I did (for Carnap) at the beginning of this paper.

That formulation was designedly archaic. Carnap would have no truck with such psychologism ("relations of ideas") and was scrupulous in speaking rather of relations between sentences (or, later, between propositions or attributes, etc.). But the point of the archaism is to suggest that even after such Carnapian refinement, empiricism seems unworkably simplistic. Here I take the distinguishing mark of empiricism to be the insistence on isolating the experiential element in knowledge from the logical element — at least in principle — so that we can conceive of testing our beliefs by tracing their provenance in impressions or in *protokollsätze* or in sets of ordered pairs (e_i, b_i) where the number b_i indicates the degree of belief in e_i which should be warranted by the observation itself, apart from the confirmation or infirmation which e_i derives from its relationships to the rest of the observer's beliefs.

I have no proof that Carnap's program for fleshing out empiricism cannot work. I take the difficulties which Carnap discussed in our correspondence, fifteen years ago, to be symptoms of unworkability, but that was never Carnap's view. He was sure that a solution could be found, and had projected a section on the topic for *Studies in Inductive Logic and Probability* (1971) before he found time getting short. It may be that, had he lived longer, he would have solved the problem, and it may be that someone else will solve it yet. But my own uncertain sense of plausibility leads me off on a different tack.

REFERENCES

Bolker, E. (1967). "A Simultaneous Axiomatization of Utility and Subjective Probability," *Philosophy of Science*, vol. 34, pp. 333–40.

Carnap, R. (1968). "Inductive Logic and Inductive Intuition," in I. Lakatos, ed. *The Problem of Inductive Logic*. Amsterdam: North Holland.

Carnap, R., and R. Jeffrey, eds. (1971). *Studies in Inductive Logic and Probability*, vol. 1. Berkeley, Los Angeles, and London: University of California Press.

Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw-Hill.

Jeffrey, R. (1973). "Carnap's Inductive Logic," *Synthese*, vol. 25, pp. 299–306.