WILLIAM H. HANSON

# Names, Random Samples, and Carnap

An important feature of Carnap's approach to the problems of probability and induction is his sharp distinction between inductive logic proper and the methodology of induction.[1] His formula for the singular predictive inference, for example, belongs to the former, and his requirement of total evidence to the latter. Under the methodology of induction he subsumes, among other things, all matters concerning random samples. His view seems to be that a methodological rule requires samples to be chosen by random methods, and hence the theorems of inductive logic itself need not concern themselves with considerations of randomness.[2]

A less-noticed and seemingly unrelated feature of Carnap's inductive logic is its heavy reliance on individual constants (i.e., names). For example, his formula for the singular predictive inference (which can be used to give a complete characterization of an inductive method) is stated in such a way that names of all the individuals involved must appear in the evidence and hypothesis statements.[3]

The purpose of this paper is to show that these two features of Carnap's approach — the relegation of considerations of randomness to methodology and the prominence of individual names — are related in an important way. More particularly, it will be demonstrated that at least in certain cases the use of names in probability problems gives exactly the same result as would be obtained in an intuitively more natural approach if evidence statements were allowed to contain the information that the samples they describe had been selected by random

[1] *Logical Foundations of Probability*, 2nd ed. (Chicago: University of Chicago Press, 1962), sec. 44. Hereafter cited as *Foundations.*

[2] *Ibid.*, p. 494.

[3] See *The Continuum of Inductive Methods* (Chicago: University of Chicago Press, 1952), secs. 4, 9, and 10. Hereafter cited as *Continuum*. See also *Foundations*, pp. 567–68.

methods. It will also be shown that use of this more natural approach would avoid certain little-noticed[4] but serious difficulties that arise at a very elementary level in Carnap's inductive logic.

Section I presents the basic difficulty in question through the use of a simple example. Section II then discusses and rejects two seemingly plausible suggestions for resolving this difficulty. In section III it is shown that the difficulty extends far beyond the simple example discussed in I and II and hence is really a serious problem for Carnap's inductive logic. Finally section IV makes some tentative suggestions for resolving the difficulty.

## I

The problem I wish to consider can be illustrated by a very simple example. Let us suppose that we have a set of boxes, each of which contains two buttons, and that each button is either amber or black. Suppose also that the set contains an equal number of boxes with no amber buttons, with one amber button, and with two amber buttons. Suppose finally that in each box one of the buttons has the letter 'a' engraved on it and the other the letter 'b', and that these identifying letters are independent of the colors of the buttons (i.e., among those boxes with one button of each color there are just as many boxes in which the button marked 'a' is amber and the one marked 'b' is black as there are boxes with the reverse combination of colors and letters). Let a box (call it '$\alpha$') be drawn at random from the set. There are now two probability problems concerning $\alpha$ that I wish to consider.

*Problem 1.* The box $\alpha$ is given to an honest and competent person who examines its entire contents without letting us see it. We then ask this person whether or not $\alpha$ contains at least one amber button and he answers, "Yes." What is the probability that both of the buttons in $\alpha$ are amber?

[4] Some of the matters considered in this paper have been discussed in the literature by Popper, Jeffrey, and Bar-Hillel. But the discussion is not widely known, and none of the discussants seems to have noticed the crucial point — namely, the relation between the use of names and random sampling. The papers in question are the following: Sir Karl Popper, "On Carnap's Version of Laplace's Rule of Succession," *Mind*, 71 (1962), 69–73, and "The Mysteries of Udolpho: A Reply to Professors Jeffrey and Bar-Hillel," *Mind*, 76 (1967), 103–110; Richard C. Jeffrey, "Popper on the Rule of Succession," *Mind*, 73 (1964), 129, and "The Whole Truth," *Synthese*, 18 (1968), 24–27; Yehosua Bar-Hillel, "On an Alleged Contradiction in Carnap's Theory of Inductive Logic," *Mind*, 73 (1964), 265–67.

*Problem* 2. We randomly select a button from $\alpha$ for observation, and we discover that it is amber. What is the probability that the unexamined button is amber and hence that both of the buttons in $\alpha$ are amber?

I think that everyone who knows a little about the mathematical theory of probability will agree that the answers to these problems are ½ and ⅔, respectively. For in the first problem our observer's report merely tells us that the range of equally probable alternatives has been narrowed from three to two — hence the answer ½. In the second problem, however, our seeing an amber button gives us all the information contained in the observer's report of the first problem and more besides. For if $\alpha$ contains just one amber button the likelihood that it will be selected by a random draw is only ½. Hence for the second problem it seems intuitively that the answer should be higher than ½, and a simple calculation shows that it is ⅔.

In spite of their different answers, the two problems are quite similar. For the hypothesis ('Both buttons are amber') is the same in each, and what we know of the composition of $\alpha$ is also the same ('At least one button is amber'). The only difference between the two is in the way in which we learn what we know of the composition of $\alpha$. In Problem 1 we are merely told that at least one button is amber by a person who has examined the entire contents of $\alpha$ and who is therefore omniscient for purposes of the problem. In Problem 2, however, we find this out by random sampling of $\alpha$. Clearly it is this difference in the way we obtain our incomplete knowledge of the composition of $\alpha$ that makes the problems different. I shall call problems of the first sort *omniscient-observer problems* (hereafter OO problems) and those of the second sort *random-sampling problems* (hereafter RS problems). Notice, incidentally, that the identifying marks that appear on the buttons play no role in either problem.

Let us now try to express and solve in a Carnapian system the probability problems that we have been considering. We can use the language $L_2^1$, which contains just one primitive predicate, A, and two individual constants, a and *b*. This language has just two Q-predicates, A and —A, which I take to be abbreviations for 'amber' and 'black', respectively. Each individual constant is taken to be the name of the correspondingly engraved button in box $\alpha$. The confirmation function c* (which at one time, at least, was Carnap's preferred confirmation function)[5] will be used to determine probabilities in $L_2^1$ since it conforms exactly to the conditions of our problems.[6]

It is easy to obtain the correct answer to Problem 1 in $L_2^1$. Clearly the hypothesis is *Aa* & *Ab* and the evidence is *Aa* v *Ab*. Values are assigned by m* to the state-descriptions of $L_2^1$ in accordance with Table 1. Hence it is easy to show that c* (*Aa* & *Ab*, *Aa* v *Ab*) = ½.

| State-descriptions of $L_2^1$ | m* | Structure-Descriptions of $L_2^1$ Are the Following Sets of State-Descriptions |
|---|---|---|
| 1. *Aa* & *Ab* | 1/3 | [1] |
| 2. *Aa* & —*Ab* | 1/6 | [2, 3] |
| 3. —*Aa* & *Ab* | 1/6 | [4] |
| 4. —*Aa* & —*Ab* | 1/3 | |

Table 1

The situation with Problem 2, however, is quite different. It will be remembered that the crucial difference between the two problems was that in Problem 2 our knowledge that at least one button is amber was obtained by random sampling. Hence the evidence statement for Problem 2 should include this information. But how do we express this information in $L_2^1$? There seems to be no way to do so, since $L_2^1$ contains only the single color predicate A. And besides, Carnap apparently thought that this sort of information need never appear in the evidence statement of a probability problem.[7] Yet if our evidence statement fails to specify the way in which we obtained our incomplete knowledge of the composition of α and states only what we know about that composition — namely, *Aa* v *Ab* — then c* will give us the same answer for Problem 2 as it did for Problem 1. Hence it looks as though Carnap's c* is unable to accommodate correctly Problem 2.

There is, however, a way to make c* yield the correct answer to Problem 2. All we need to do is look closely at the amber button our random draw produced and note its identifying mark. Suppose that it is 'a'. Then we can express the evidence as *Aa*, and, since c* (*Aa* & *Ab*,

[5] *Foundations*, pp. 562–67.

[6] The assumption that α is selected at random from a set of boxes in which each statistically different distribution occurs as often as any other entails that all the possible structures of α are equally probable. Analogously, c* assigns equal probabilities to each of the structure-descriptions of $L_2^1$. It also assigns equal probabilities to each state-description within a given structure-description, and this corresponds to our assumption that there are just as many boxes in which the button marked 'a' is amber and the one marked 'b' is black as there are boxes with the reverse combination of letters and colors.

[7] See *Foundations*, p. 494. Cf. fn. 20.

*Aa*) = ⅔, Problem 2 receives its correct answer. Of course we get the same answer if the identifying mark is 'b', since c*(*Aa* & *Ab*, *Ab*) = ⅔. So Carnap's c* is able to solve Problem 2 after all. But there is something very strange about the solution. When we first considered Problems 1 and 2 we noted that the identifying marks played no role at all in determining the correct answers. Hence for purposes of these problems the buttons need have no identifying marks. Yet without them the solution to Problem 2 for c* that I have just given is impossible.

The bizarreness of this situation is perhaps more apparent if we suppose that the marks exist but are so small that they can be read only with a magnifying glass. Then if we randomly select a button from $\alpha$, find it to be amber, but don't use the glass, the only evidence we have that can be stated in $L_2^1$ is *Aa* v *Ab*. Hence c* will give ½ as the answer to Problem 2. But if we use the glass on the selected button, then no matter what we see, 'a' or 'b', c* will give the correct answer to Problem 2. But surely the odds on a fair bet based on Problem 2 do not change because of the use of a magnifying glass.

Summarizing, it seems that Carnap's c* gives the correct answer to an RS problem like Problem 2 only if the evidence statement contains names of the individuals in the sample. If the evidence does not contain names, then c* treats an RS problem as if it were an OO problem (like Problem 1) and may give an incorrect answer. In section III, I shall prove rigorously that this result holds not only for c* but for all the c-functions that satisfy Carnap's conditions C1–C11.[8] But first I shall discuss informally, in terms of the simple examples given above, two proposals for resolving the difficulty.

## II

As was pointed out earlier (in footnote 4), the difficulty under consideration has been discussed in the literature by several authors. In "The Whole Truth" Jeffrey proposes a solution to the problem that at first seems simple and natural. He suggests that we treat names (i.e., the individual constants of Carnap's languages) like disguised descriptions rather than tags.[9] Thus in Problem 2 instead of using *a* and *b* for the

[8] See *Continuum*, pp. 12–29.

[9] I am not here suggesting that Carnap thought of individual names as mere tags,

correspondingly engraved buttons, we might let *a* stand for 'the button I will see' and *b* for 'the other button'. As long as we look at one and only one button it is clear that such an approach avoids the difficulty that arises in Problem 2 when names are treated like tags. For if the uniqueness conditions of the descriptions abbreviated by *a* and *b* are satisfied, then we will be able to state the evidence without using a disjunction. The difficulty with this approach is, of course, that the uniqueness conditions of the descriptions may not always be satisfied. Suppose that we are dealing with an *N*-membered population and that we are using the individual constants $a_1, a_2, \ldots, a_N$. We might then think of letting $a_i$ abbreviate 'the $i^{th}$ thing I see'. But suppose that in making my third observation I inadvertently see two things at once. Then $a_3$ does not denote anything, and familiar problems arise for sentences that contain it. Similar problems arise for $a_j (j \geqq 5)$ if I never see more than four things. Or suppose that on my sixth observation I inadvertently see the same thing I saw on the second observation. Then $a_6$ and $a_2$ denote the same thing, contrary to one of the basic assumptions of Carnap's semantics, and more than *N* individual constants will be needed to provide a name for each object.

Of course all these difficulties can be avoided if just the right descriptions are used. But there is no assurance that we will always be able to construct just the right descriptions, and this for two reasons. First we simply may lack the ingenuity or the vocabulary, and second (as the previous paragraph shows) the problem may change in ways that we do not anticipate. In addition to these difficulties, Jeffrey's approach has the drawback that the semantics of the language has to be modified constantly to accommodate different probability problems, even different problems dealing with the same set of objects. Suppose I intend to make several estimates of the composition of the same population of *N* objects based on several series of random draws. Then $a_i$ can no longer abbreviate 'the $i^{th}$ thing I see', but will have to be relativized to the set of draws and become something like 'the $i^{th}$ thing I see in the first set of draws'. But then when the second set of draws begins it will have to be modified in an appropriate way. For all these reasons it seems to me

that is, as signs having denotation but no sense or meaning. It is clear from his talk about individual concepts in *Meaning and Necessity*, for example, that he did not. The point is rather that he did not think of them as disguised descriptions either, and, furthermore, for purposes of his inductive logic they could just as well be thought of as tags. On this latter point see *Foundations*, p. 59.

that Jeffrey's approach is a desperate expedient and should be avoided if at all possible. Surely Carnap himself would not have endorsed it since he did not think of names as disguised descriptions.

Another solution to the problem under consideration has been suggested by several statisticians with whom I have discussed it. They typically find my distinction between RS and OO problems a bit arbitrary since, they say, the former can be transformed into the latter by merely enlarging the sample space and building the assumption of randomness into the distribution. With Problem 2 this can be done by adding a predicate O for 'is observed', yielding the eight possibilities of Table 2.

| | |
|---|---|
| 1. *Aa* & *Ab* & *Oa* | 1/6 |
| 2. *Aa* & *Ab* & *Ob* | 1/6 |
| 3. *Aa* & —*Ab* & *Oa* | 1/12 |
| 4. *Aa* & —*Ab* & *Ob* | 1/12 |
| 5. —*Aa* & *Ab* & *Oa* | 1/12 |
| 6. —*Aa* & *Ab* & *Ob* | 1/12 |
| 7. —*Aa* & —*Ab* & *Oa* | 1/6 |
| 8. —*Aa* & —*Ab* & *Ob* | 1/6 |

Table 2

Following the constraints on Problem 2 we have that each of the first two and the last two sentences in Table 2 has a probability of 1/6, while each of the remaining four has a probability of 1/12. We can now consider the evidence in Problem 2 to be merely that an amber button is observed, and express this as (*Aa* & *Oa*) v (*Ab* & *Ob*). Problem 2 can now be handled correctly without having to notice the mark engraved on the observed button since

$$P(Aa \,\&\, Ab/(Aa \,\&\, Oa) \vee (Ab \,\&\, Ob)) = 2/3.$$

Since the statement of evidence in the equation above contains no explicit mention of random sampling, we are here treating an RS problem in the way we previously treated OO problems. The reason this approach works is that we have divided each of the four state-descriptions of Table 1 into two parts by adding either *Oa* or *Ob*, and then divided the value previously assigned to the state-description *equally* between its two parts. For example, state-description 1, which previously was assigned the value 1/3, has been replaced by the first two sentences of Table 2, each of which is assigned the value of 1/6. It is this equal division of the old state-description values between their new parts that amounts to asserting that the observations were made at random. This is what my statistician friends mean by building the assumption of randomness into the distribution. It should also be noticed that we continue to get the correct answer to Problem 1 on this approach since

$$P(Aa \,\&\, Ab / Aa \vee Ab) = 1/2.$$

The success of the method sketched in the preceding paragraph may lead one to believe that the difficulty involved in handling RS problems in Carnap's inductive logic has been solved. That such a belief is premature can be seen by noticing that the sentences in Table 2, unlike those in Table 1, are not state-descriptions. For in adding the new predicate O we have transformed the language $L_2^1$ into $L_2^2$, and the state-descriptions of $L_2^2$ all contain four conjuncts. Hence there are sixteen state-descriptions of $L_2^2$; they are given in Table 3. In order to handle Problem 2 in $L_2^2$ without giving favored status to either *a* or *b*, we have to express the evidence, e, as $(Aa \,\&\, Oa \,\&\, -Ob) \vee (Ab \,\&\, Ob \,\&\, -Oa)$. The hypothesis, *h*, continues to be *Aa* & *Ab*. It now turns out, unfortunately, that

$$c^*(h, e) = 1/2.$$

We also get the wrong answer for Problem 1, since

$$c^*(h, Aa \vee Ab) = 3/7.$$

The reason for the discrepancy between the c*-values in $L_2^2$ and the values based on Table 2 is that $L_2^2$ has *ten* structure-descriptions, each of which is weighted equally by m*, while Table 2 is just a refinement of Table 1 and hence preserves the effect of that table in which m* assigns equal values to the *three* structure-descriptions of $L_2^1$. In order to get the right answers to Problems 1 and 2 in $L_2^2$ we would have to construct a special measure function that preserves the effect of Table 2. The function m** given in Table 3 is the obvious choice. It preserves all the values assigned by Table 2 and also makes explicit what was implicit in that table, namely, that cases in which neither or both of the objects are observed simply are not being considered. Unfortunately there are several difficulties with m**. One is that it violates both conditions C9 and C10 of *Continuum* and hence is unavailable within a strictly Carnapian approach.[10] This is not so serious, since if there were

[10] $c^{**}(Aa \,\&\, Oa, Ab \,\&\, -Ob) = 2/3$ *while* $c^{**}(Aa \,\&\, Oa, -(Ab \,\&\, Ob)) = 1/3$. According to C9 these two values should be the same, and according to C10 the latter should fall within the interval 0–1/4.

| State-Descriptions of $L_2^2$ | m* | m** | Structure-Descriptions of $L_2^2$ Are the Following Sets of State-Descriptions |
|---|---|---|---|
| 1. *Aa* & *Ab* & *Oa* & *Ob* | 1/10 | 0 | [1] |
| 2. *Aa* & *Ab* & *Oa* & —*Ob* | 1/20 | 1/6 | [2, 3] |
| 3. *Aa* & *Ab* & —*Oa* & *Ob* | 1/20 | 1/6 | [4] |
| 4. *Aa* & *Ab* & —*Oa* & —*Ob* | 1/10 | 0 | [5, 9] |
| 5. *Aa* & —*Ab* & *Oa* & *Ob* | 1/20 | 0 | [6, 11] |
| 6. *Aa* & —*Ab* & *Oa* & —*Ob* | 1/20 | 1/12 | [7, 10] |
| 7. *Aa* & —*Ab* & —*Oa* & *Ob* | 1/20 | 1/12 | [8, 12] |
| 8. *Aa* & —*Ab* & —*Oa* & —*Ob* | 1/20 | 0 | [13] |
| 9. —*Aa* & *Ab* & *Oa* & *Ob* | 1/20 | 0 | [14, 15] |
| 10. —*Aa* & *Ab* & *Oa* & —*Ob* | 1/20 | 1/12 | [16] |
| 11. —*Aa* & *Ab* & —*Oa* & *Ob* | 1/20 | 1/12 | |
| 12. —*Aa* & *Ab* & —*Oa* & —*Ob* | 1/20 | 0 | |
| 13. —*Aa* & —*Ab* & *Oa* & *Ob* | 1/10 | 0 | |
| 14. —*Aa* & —*Ab* & *Oa* & —*Ob* | 1/20 | 1/6 | |
| 15. —*Aa* & —*Ab* & —*Oa* & *Ob* | 1/20 | 1/6 | |
| 16. —*Aa* & —*Ab* & —*Oa* & —*Ob* | 1/10 | 0 | |

Table 3

good reasons to adopt m** Carnap's conditions could be modified to accommodate it. A more serious difficulty with m** (and indeed with the entire approach being discussed here) is that use of a special predicate for 'is observed' forces us to revise constantly the semantics of the language. For *Oa* does not mean simply that *a* is observed at some time by someone. If this were the case then the probability of *Oa* would be much higher than ½, and the probability of *Oa* & *Ob* would not be zero. Clearly *Oa* is to be interpreted as saying that *a* is observed within a certain span of time or within the context of a certain specified set of procedures. But then if the drawing of an object from the set containing just *a* and *b* is to be repeated at a later time, the interpretation of O must be changed. Surely it would be better if this sort of frequent modification could be avoided.

The most serious difficulty with m**, and with similar functions that might be defined to accommodate the predicate O, is that for a given domain of individuals we will need many such probability functions, one for each possible sample size. To illustrate this, suppose that we have a set of ten objects each of which is either amber or black. Suppose also that we intend to draw five of them at random, observe their color, and on the basis of our observation assess the probability of predictions about the frequency of the two colors in the entire set. Then we will need a probability function that assigns zero to each state-description in which some number of individuals other than five is observed. But if we next decide to make similar assessments on the basis of a six-membered sample, the probability function must change. Now we will need one that assigns zero to each state-description in which some number of individuals other than six is observed. It hardly needs to be said that changing the probability function as the problem changes, even though the number of properties and individuals remains the same, is not the sort of thing Carnap had in mind. And independently of Carnap's own preferences I think it is desirable to avoid constant modification of the probability function, just as in the previous case it seemed desirable to avoid constant modification of the semantics of the language.

The upshot of all this is that neither Jeffrey's approach nor the approach suggested by my statistician friends is very satisfactory for handling the difficulty that I have been considering. In section IV I will sketch the beginnings of a solution that avoids the kinds of problems encountered by these two approaches. Unfortunately, this solution will require the formal languages and the probability functions to be more complicated than any considered by Carnap.

## III

The conclusions of section I regarding difficulties and limitations inherent in Carnap's inductive logic were based on the single example that gave rise to Problems 1 and 2. Hence it is legitimate to ask if similar examples give rise to similar difficulties. If not, then Carnap's inductive logic will merely have been shown to contain a single anomaly, and this is probably not of much interest. But if similar examples do give rise to similar difficulties, then the criticisms of the previous sections will have been shown to be broadly based and serious. In the present section I will show that the latter is indeed the case. Specifically, the difficulties involved in handling Problem 2 in Carnap's inductive logic will be shown to be independent of the number of individual constants, the number of predicates, and the c-function.

I shall begin by giving Carnap's formula for the singular predictive inference as it appears in *Continuum*. This will serve as a point of departure for the results to be presented in this section, since it is Carnap's general formula for a large class of RS problems. But first some notation must be introduced. In *Continuum* Carnap restricts his attention to languages with a finite number of primitive predicates ($\pi$) and a finite number of individual constants ($N$). These languages are designated $L_N^\pi$. The strongest noncontradictory predicates that can be formed in $L_N^\pi$ are called $Q$-predicates. $L_N^\pi$ contains $\kappa = 2^\pi$ $Q$-predicates, and they are designated as $Q_1, Q_2, \ldots, Q_\kappa$. A $Q$-sentence is a sentence that results from applying a $Q$-predicate to an individual constant. The function $\lambda(\kappa)$ takes as values the non-negative real numbers and $\infty$ and gives a complete characterization of the confirmation function $c_\lambda$. In the formula for the singular predictive inference to be given below the evidence statement, $e_Q$, says that $s_1$ individuals have property $Q_1$, $s_2$ have $Q_2, \ldots, s_\kappa$ have $Q_\kappa$. $s_1 + s_2 + \ldots + s_\kappa = s$. The hypothesis, $h_i$, says that some individual not mentioned in the evidence has $Q_i$. More precisely, $e_Q$ is a conjunction of $s$ distinct $Q$-sentences each of which contains a distinct individual constant, where $s_1$ of these constants occur in $Q$-sentences with $Q_1$, $s_2$ with $Q_2, \ldots, s_\kappa$ with $Q_\kappa$. For each $i$, $1 \leq i \leq \kappa$, $h_i$ is a $Q$-sentence made up of $Q_i$ and an individual constant that does not appear in $e_Q$. Carnap's formula for the singular predictive inference involving $h_i$ and $e_Q$ is as follows:[11]

$$(1) \qquad c_\lambda(h_i, e_Q) = \frac{s_i + (\lambda(\kappa)/\kappa)}{s + \lambda(\kappa)}$$

To see how (1) applies to an RS problem, recall that Problem 2 used the language $L_2^1$ and the c-function $c^*$. So we have $\pi = 1$, $N = 2$, $\kappa = 2^\pi = 2$, $Q_1 = A$, $Q_2 = -A$, and $\lambda(\kappa) = \kappa = 2$.[12] In Problem 2, $s_1 = s = 1$ and $s_2 = 0$, so $e_Q = Aa$ and $h_1 = Ab$. Hence (1) gives

$$c^*(Ab, Aa) = (1 + 1)/(1 + 2) = 2/3.$$

And this is the same result that we got in the previous section after we had noticed the identifying mark on the observed button.[13]

[11] Equation (1) follows from (9-6) and (4-4) of *Continuum*. The evidence statement $e_Q$ is what Carnap calls an individual distribution. Equation (1) also apparently holds if the evidence is expressed as a statistical distribution. (See *Foundations*, pp. 111, 567–68.) Replacing individual distributions with statistical distributions does not, however, obviate the criticisms of Carnap's inductive logic made in this paper since a statistical distribution is always relative to a fixed set of individual constants.

[12] Function $c^*$ is that c-function for which $\lambda(\kappa) = \kappa$. See *Continuum*, sec. 15.

[13] Actually in section I our hypothesis was $Aa$ & $Ab$ rather than $Ab$. But as long as the evidence is $Aa$ in both cases the probabilities must be the same. Similarly, in (1) the hypothesis could be $h_i$ & $e_Q$ rather than $h_i$ without affecting the probability. This fact should be kept in mind when comparing (1) with (2) below.

In order to generalize the difficulty that was encountered in section I in connection with Problem 2, I shall have to produce and prove an equation that is similar to (1), but that contains instead a general formula for OO problems.[14] If it turns out that the general formulas for RS and OO problems give different results in a large number of cases, then I will have shown that the difficulty of section I is widespread. For the difficulty was just that sometimes one is forced to treat an RS problem like an OO problem (i.e., when the individual constants that have been assigned to the individuals in the sample are unknown). In giving the general result about OO problems it will be convenient yet sufficient to restrict attention to those cases in which $s = N - 1$. With this restriction in mind I introduce some additional notation.

For any non-negative integers $s_1, s_2, \ldots, s_\kappa$ such that $s_1 + s_2 + \ldots + s_\kappa = s = N - 1$, let $h_i'$ $(1 \leq i \leq \kappa)$ be the structure-description with Q-numbers $s_i + 1$ and $s_j$ (for all $j \neq i$).[15] Then let $e_Q'$ be the disjunction $h_1' \vee h_2' \vee \ldots \vee h_\kappa'$. This has the effect of making $e_Q'$ a sentence which says that at least $s_1$ individuals have $Q_1$, at least $s_2$ individuals have $Q_2, \ldots$, at least $s_\kappa$ individuals have $Q_\kappa$. Yet $e_Q'$ doesn't say which individuals have which Q-properties. And each $h_i'$ gives a complete yet purely structural description of the domain with which we are concerned. Hence $e_Q'$ and $h_i'$ are, in a certain obvious sense, structural analogues of $e_Q$ and $h_i$, respectively, for the case $s = N - 1$.[16] As an example consider how this notation applies to Problem 1. The language is again $L_2^1$. Since $s_1 = s = N - 1 = 1$, $h_1'$ is $Aa \,\&\, Ab$, $h_2'$ is $(Aa \,\&\, -Ab) \vee (-Aa \,\&\, Ab)$, and $e_Q'$ is $h_1' \vee h_2'$ which is logically equivalent to $Aa \vee Ab$. Hence $h_1'$ and $e_Q'$ correspond precisely to the hypothesis and evidence statements of Problem 1.

In view of the interpretation of $h_i'$ and $e_Q'$, it should be obvious that a formula for the value of $c_\lambda(h_i', e_Q')$ will give us the answers to a large

[14] In most of this section I simply assume that (1) is a correct formula for the singular predictive inference in those cases where the initial probability distribution is given by $m_\lambda$ and the sample is selected at random. Hence I am assuming that (1) is a correct formula for a large class of RS problems. This assumption seems unproblematic. Nevertheless equation (3), which appears at the end of this section, supports (1) since it gives the same result for the case in which $s = N - 1$.

[15] That is, $h_i'$ is a structure-description in which $s_j$ individuals have $Q_j$, for all $j \neq i$, and $s_i + 1$ individuals have $Q_i$.

[16] Actually, $h_i'$ is more closely analogous to $h_i \,\&\, e_Q$ than to $h_i$, since the first sentence is a structure-description and the second is a state-description, while the third is just a single Q-sentence. But this is an inessential difference for my purposes since $c_\lambda(h_i, e_Q) = c_\lambda(h_i \,\&\, e_Q, e_Q)$. Cf. fn. 13.

class of OO problems, just as (1) gives us the answers to a large class of RS problems. The formula is contained in equation (2) below.

For any $h_i'$ $(1 \leqq i \leqq \kappa)$,[17]

$$(2) \qquad c_\lambda(h_i', e_Q') = \frac{(s_i + \lambda/\kappa)/(s_i + 1)}{[(s_1 + \lambda/\kappa)/(s_1 + 1)] + [(s_2 + \lambda/\kappa)/(s_2 + 1)] + \ldots + [(s_\kappa + \lambda/\kappa)/(s_\kappa + 1)]}$$

Equation (2) holds for any c-function that satisfies conditions C1–C11 of *Continuum* and any choice of non-negative integers $s_1, s_2, \ldots, s_\kappa$ such that their sum $s = N - 1$. The proof of (2) is given in the appendix.

As a formula for OO problems (2) is correct and unobjectionable. But if one were to attempt to use (2) rather than (1) for RS problems the results would be disastrous. To illustrate, suppose that we have a domain of $N$ objects for which we distinguish $\pi$ primitive predicates, and that we have applied $L_N^\pi$ by way of the usual semantical rules to this domain. Part of the semantics of $L_N^\pi$ will involve specifying a one–one mapping from the $N$ individual constants of $L_N^\pi$ to the objects of the domain. Suppose also that we randomly select $N - 1$ objects and determine that $s_1$ of them have the property $Q_1$, $s_2$ have $Q_2, \ldots, s_\kappa$ have $Q_\kappa$. We now want to know what the probability is that the remaining object has the property $Q_i$ (when $1 \leqq i \leqq \kappa$) according to our favorite c-function, $c_\lambda$. If we know which individual constants have been assigned to the objects we have observed, then we can apply (1) and get an answer that is correct, given the initial assignment of probabilities made by $m_\lambda$. If, however, we do not know which individual constants have been assigned to the observed objects, then we may be tempted to state the evidence and hypothesis as strongly as we can (i.e., with $e_Q'$ and $h_i'$) and apply (2). But an analysis of the consequences of (2) will show that this approach is unthinkable for an inductivist like Carnap. For Carnap thought that any c-function worth serious consideration should, when applied to problems of the kind we are considering, satisfy at least two conditions: (a) The value of the c-function should vary directly with $s_i/s$; and (b) for a sufficiently large sample it should be close to $s_i/s$.[18] These conditions are often summarized by saying that a

[17] In (2) I use 'λ', as Carnap often does, as an abbreviation of the functional notation 'λ(κ)'.

[18] *Continuum*, p. 24.

satisfactory c-function should allow us to learn from experience. It is now easy to show that if (2) were applied to RS problems, learning from experience would be sadly deficient.

If $0 \leqq \lambda < \kappa$, then each of the terms $\frac{s_j + \lambda/\kappa}{s_j + 1} (1 \leqq j \leqq \kappa)$ of (2) varies directly with $s_j$. Hence (2) allows some learning from experience, but usually very little. For even when $\lambda = 0$ (2) becomes

$$c_0(h_i', e_Q') = \frac{(s_i)/(s_i + 1)}{[(s_1)/(s_1 + 1)] + [(s_2)/(s_2 + 1)] + \ldots + [(s_\kappa)/(s_\kappa + 1)]}$$

which (although it gives the same results as the "straight rule" for $s_i = 0$ and $s_i = s$) gives results quite close to $1/\kappa$ as long as each of the numbers $s_1, s_2, \ldots, s_\kappa$ is fairly large, even though these numbers may differ widely among themselves. For example, let $\kappa = 2$, $N = 500$, $s_1 = 100$, and $s_2 = 399$. Then

$$c_0(h_1', e_Q') = .498$$

and

$$c_0(h_2', e_Q') = .502.$$

It should be remembered that when $\lambda = 0$ learning from experience is supposed to be at its maximum in Carnap's systems.

As $\lambda$ moves from 0 toward $\kappa$ learning from experience declines, and when $\lambda = \kappa$, which gives us the function $c^*$,[19] we have

$$c^* (h_i', e_Q') = 1/\kappa.$$

For $\lambda > \kappa$ each of the terms $(s_j + \lambda/\kappa)/(s_j + 1)$ $(1 \leqq j \leqq \kappa)$ of (2) varies inversely with $s_j$. Hence if (2) is applied to an RS problem it here yields counterinductive methods. The worst of these is $c_\infty (h_i', e_Q')$ which according to Carnap's conventions is defined as

$$\lim_{\lambda \to \infty} [c_\lambda (h_i', e_Q')].$$

It is not difficult to show that

$$c_\infty (h_i', e_Q') = \frac{(1)/(s_i + 1)}{[(1)/(s_1 + 1)] + [(1)/(s_2 + 1)] + \ldots + [(1)/(s_\kappa + 1)]}$$

[19] See fn. 12.

As an example of the strongly counterinductive nature of (2) when $\lambda = \infty$ let $\kappa = 2$, $N = 100$, $s_1 = 10$, and $s_2 = 89$. Then

$$c_\infty(h_1', e_Q') = .891$$

and

$$c_\infty(h_2', e_Q') = .109.$$

It should now be clear that the discrepancy between equations (1) and (2) is widespread, and hence that (2) cannot reasonably be used for RS problems. So the criticisms of Carnap's inductive logic made in section I turn out to be broadly based.

Up until now I have simply assumed that (1) gives a correct formula for a large class of RS problems. I still see no reason to doubt this assumption, but it will be instructive to show that for the case in which $s = N - 1$ it is possible to state the problems in such a way that no individual constant is given favored status and yet to obtain the same formula. In order to obtain the desired result I retain all the notation introduced in connection with (1) and (2), and in addition I introduce $e'$ as an abbreviation of the sentence 'A sample selected at random contained $s_1$ objects with $Q_1$, $s_2$ objects with $Q_2$, . . . , $s_\kappa$ objects with $Q_\kappa$'. It is fairly clear that Carnap did not intend any of the predicates of his languages $L_N{}^\pi$ to be interpreted as 'is (was) selected by a random method',[20] so $e'$ should not be thought of as a sentence of $L_N{}^\pi$. But suppose that we have probability functions $P_\lambda$ that assign values to sentences like $e'$ as well as to the sentences of $L_N{}^\pi$, and suppose further that if $A$ and $B$ are any sentences of $L_N{}^\pi$

$$P_\lambda(A,B) = c_\lambda(A,B).$$

If we now take $s_1, s_2, \ldots, s_\kappa$ to be any non-negative integers whose sum $s = N - 1$, we will find that equation (3) gives the result we seek. The proof of (3) is given in the appendix.

For any $h_i'$ $(1 \leqq i \leqq \kappa)$,

[20] In *Foundations*, p. 494, Carnap argues that no theorem of inductive logic proper need contain the requirement that a sample be random (or even that it not be known not to be random) since this will be taken care of by the requirement of total evidence — a methodological principle. Thus he avoids bringing considerations of randomness into inductive logic proper by saying, in effect, that whenever we seriously *apply* inductive logic to a practical problem we should, as a matter of correct methodology, make sure that we have no evidence indicating that our samples are not random.

$$(3) \qquad P_\lambda(h_i',e') = \frac{s_i + (\lambda(\kappa)/\kappa)}{s + \lambda(\kappa)}.$$

Notice that the right side of (3) is identical with that of (1), Carnap's equation for the singular predictive inference.

Equation (3) shows that it is possible (at least when $s = N - 1$) to understand the singular predictive inference in such a way that individual constants are entirely irrelevant to it, and yet to agree with Carnap's formula for it as given in (1). More precisely, we can take the only essential evidence to be that $s = N - 1$ individuals have been randomly selected and have been found to be distributed among $\kappa$ properties in a certain specified way. This is what $e'$ expresses. We can then construe any prediction about the remaining individual as a statement about the complete structure of the population. The sentences $h_i'$ $(1 \leq i \leq \kappa)$ express all such hypotheses that are compatible with $e'$. Finally (as the proof of (3) in the appendix shows), if we take Carnap's $m_\lambda$ to assign the correct probability value to each $h_i'$, and if we assign values to each of the likelihoods $P_\lambda(e', h_i')$ in the only reasonable way, then the value for the singular predictive inference is the same as that given by (1).

With (3) we have come full circle, for (3) is the formal embodiment of the informal approach that was initially used to solve Problem 2 in section I. And the fact that (3) and (1) give the same result can be thought of as vindication of Carnap's formula for the singular predictive inference, at least for $s = N - 1$. It also shows that given a probability distribution $m_\lambda$, there is a trade-off between framing the evidence and hypothesis statements in terms of the preassigned names of the objects involved and stating the evidence and hypothesis in purely structural terms but including in the evidence an explicit statement that the sample was obtained by random methods. Thus the use of names can be thought of as a device for coding the information that the sample was chosen at random. Whether or not Carnap would have been better advised to make use of sentences like $e'$ instead of the coding device provided by names will be discussed in the next section.

## IV

I think the foregoing results show that giving an adequate formulation of the logical approach to probability is much more difficult than Carnap and his sympathizers thought. Serious problems arise at an entirely elementary level, and they do not seem to have any simple solutions. All this, I fear, bodes ill for any approach that attempts to make comprehensive assignments of probabilities to the sentences of a formalized language. Nevertheless, in closing I will attempt to sketch the first step or two of an adequate solution to the problem presented in sections I and III. It is not surprising that this solution will be much more complicated than either of those considered in section II.

As I pointed out at the end of the previous section, the difficulty that gave rise to this paper can be traced to Carnap's requirement that the evidence and hypothesis statements of RS problems be framed in terms of preassigned names of the objects involved. But as equation (3) showed, this requirement could be dropped (at least where $s = N - 1$) if we had (a) some way of including in the evidence statement of RS problems the information that the sample was selected by a random method, and (b) some procedure for assigning probabilities to statements that contain talk of random selection. I shall here concern myself only with a sketch of a solution to (a).

It should be fairly easy to see that (a) cannot be solved simply by adding to Carnap's languages a special one-place predicate that is always interpreted as 'was selected by a random method'. For all the difficulties that arose for the special predicate O ('is observed') discussed in section II would also arise for this predicate. But something like this can be done. Suppose that we introduce a special set of individual constants $t_1$, $t_2$, . . . that designate moments of time. We could then introduce a special two-place predicate $R$ such that $R\alpha\beta$ is always interpreted 'Object $\alpha$ is selected by a random method at time $\beta$'. The fact that $R$ takes a temporal argument eliminates some of the problems that arose for O in section II. And it is possible to use $R$ and the $t_i$ to give plausible statements of the evidence and hypotheses in RS problems. These statements will not contain preassigned names of particular objects, but will express what is important in such problems — namely, the frequency of the observed properties and the fact that the sample was selected by a random method.[21] As an example, consider a

[21] It might be objected at this point that the only important difference between R and O is that the former is a two-place predicate that takes a temporal argument whereas the latter is a one-place predicate. Hence, the objection might go as follows:

typical case of the singular predictive inference in which the evidence is that two objects have been selected at random from some population and found to be amber ($A$), and the hypothesis is that the next object to be selected at random will be amber. I suggest that this problem can be understood in terms of e, *i*, and *h*, below, as the probability of *h* given e & *i*. (Here e and *h* are the evidence and hypothesis, respectively, and *i* is background information that specifies the nature of the singular predictive inference. $E\alpha\beta$ is to be interpreted as '$\alpha$ is earlier than $\beta$'.)

$$e = (\exists x)(\exists y)(x \neq y \,\&\, Rxt_1 \,\&\, Ryt_2 \,\&\, (z)(Rzt_1 \supset z = x) \,\&\, (z)(Rzt_2 \supset z = y) \,\&\, Ax \,\&\, Ay)$$

$$i = Et_1t_3 \,\&\, Et_2t_3 \,\&\, (\exists x)(Rxt_3 \,\&\, (y)(Ryt_3 \supset y = x)) \,\&\, (x)((Rxt_1 \vee Rxt_2) \supset -Rxt_3)$$

$$h = (\exists x)(Rxt_3 \,\&\, (y)(Ryt_3 \supset y = x) \,\&\, Ax)$$

Sentence e should require no explanation. Sentence *i* assures that the thing mentioned in the hypothesis is distinct from and is selected later than the things mentioned in the evidence. It also says that exactly one thing is selected at random at $t_3$. This clause is included because without it P(*h* given e & *i*) would be the probability that a third thing is amber and is selected at $t_3$, given e. But in the singular predictive inference it is assumed as part of the problem that this additional thing will be selected. What we wish to know is the probability that this thing will be amber given e *and* that it will be selected.[22]

if in section II I had allowed O to be a two-place predicate with a temporal argument, then the problems that this paper raises for Carnap's inductive logic could have been solved along the lines suggested there without eliminating preassigned names. But this objection overlooks three important points. The first is that the intended interpretation of *R* involves random selection while that of O involves merely selection or observation, not necessarily by a random method. Hence O might be interpreted in ways that do not reflect a crucial feature of the problems under consideration. The second point overlooked by this objection is that since the use of preassigned names is at best a misleading way of coding other information (namely, that the sample was selected at random) their use should be avoided if at all possible. The third point is that whichever approach one chooses for *expressing* the evidence and hypothesis in an RS problem (i.e., whether one chooses to use R and eliminate names as in the present section or to use O and retain names as in section II) the problem of determining how to *assign probabilities to* those statements remains. This is the problem referred to as (b) in the present section. We encountered this same problem in section II when we noticed that functions like m** were relative to the size of the sample referred to in the evidence statement.

[22] It might be objected that the uniqueness claims of e and *h* are too strong since

Notice that no individual constants designating objects appear in $e$, $i$, or $h$. Hence the sort of difficulty that faces Carnap's inductive logic in handling RS problems cannot arise. Individual constants for times do appear, but I do not think they create any special difficulties. So I suggest that anyone interested in pursuing Carnap's program of assigning probability measures to formal languages begin by using something like the techniques sketched in the example above for expressing RS problems. I leave to him the much more difficult task of determining how the measure functions are to be specified in order to yield the desired values. The only thing about this latter task that is clear to me is that $R$ will have to be treated differently from other predicates, and this will probably mean substantial revision of Carnap's approach.

## V

The lesson to be learned from this paper, if there is just one such lesson, is that Carnap's move in banishing considerations of random sampling from inductive logic proper to the methodology of induction[23] was a mistake. Undoubtedly this move greatly simplified matters and facilitated Carnap's impressive technical achievements. But it ultimately led to confusion and absurdity, and if my general line of argument is correct it must be abandoned by the serious logical probability theorist. I have tried to indicate above how such a theorist might begin making the necessary repairs. Their completion will very likely be a long, difficult, and thankless task which I gladly leave to him.[24]

## Appendix

*Proof of* (2). Since $e_Q'$ is defined as $h_1' \vee h_2' \vee \ldots \vee h_K'$ and the $h_j'$ are pairwise mutually exclusive, we have by Bayes's theorem that

it is highly implausible, for example, that just one thing in the entire world is selected at random at $t_3$. This could be remedied by making $R$ a three or more place predicate that takes arguments for places and/or observers as well at times.

[23] See Foundations, pp. 493–94.

[24] This paper was written before I became aware of the work on inductive logic done by Hintikka and his followers. (See, for example, several of the articles in J. Hintikka and P. Suppes, eds., *Aspects of Inductive Logic*, Amsterdam: North Holland, 1966.) As far as I am able to tell, however, these authors have neither discovered nor attempted to resolve the difficulties raised in this paper.

$$(4)\ c_\lambda(h_i', e_Q') = \frac{m_\lambda(h_i') \cdot c_\lambda(e_Q', h_i')}{m_\lambda(h_1') \cdot c_\lambda(e_Q', h_1') + \ldots + m_\lambda(h_\kappa') \cdot c_\lambda(e_Q', h_\kappa')}.$$

But clearly each $h_j'$ entails $e_Q'$, so (4) reduces to

$$(5) \qquad c_\lambda(h_i', e_Q') = \frac{m_\lambda(h_i')}{m_\lambda(h_1') + \ldots + m_\lambda(h_\kappa')}.$$

Hence the problem reduces to finding the $m_\lambda$-values of structure-descriptions. Since all the state-descriptions in a structure-description have the same Q-numbers, and since the $m_\lambda$-value of a state-description is a function only of $\lambda$, $\kappa$, and its Q-numbers, we will be able to calculate the $m_\lambda$-value of any structure-description if we know the Q-numbers associated with it and the number of state-descriptions it contains.[25] In view of the fact that $s = N - 1$, the Q-numbers associated with the structure-description $h_j'$ are $s_j + 1$ and $s_i$, for all $i \neq j$. Hence by (10-7) of *Continuum* each state-description in $h_j'$ has as its $m_\lambda$-value,[26]

$$(6) \qquad \frac{(\lambda/\kappa + s_j) \prod_i [\lambda/\kappa(\lambda/\kappa + 1)(\lambda/\kappa + 2) \ldots (\lambda/\kappa + s_i - 1)]}{\lambda(\lambda + 1)(\lambda + 2) \ldots (\lambda + N - 1)}.$$

The number of state-descriptions in $h_j'$ is

$$(7) \qquad \frac{N!}{(s_j + 1)! \prod_{i \neq j} s_i!}$$

since this is just the number of those statistically equal distributions of $N$ individuals among $\kappa$ properties which assign $s_i$ individuals to the $i^{th}$ property, for all $i \neq j$, and $s_j + 1$ individuals to the $j^{th}$ property, where

$$s_j + 1 + \sum_{i \neq j} s_i = N.$$

Hence the $m_\lambda$-value of $h_j'$ is the product of (6) and (7). By substitut-

[25] The Q-numbers of a state-description are the numbers of individuals assigned to each Q-property by that state-description. Equation (10-7) of *Continuum* gives the $m_\lambda$-value of a state-description as a function only of $\lambda$, $\kappa$, and its Q-numbers.

[26] It is important to notice that in the iterated product in the numerator of (6) $i$ takes all values from 1 to $\kappa$, including $j$.

ing an appropriate instance of this product for the numerator and each addend of the denominator of (5) and then removing from each of these terms the common factor

$$(8) \qquad \frac{\prod_{i}[\lambda/\kappa(\lambda/\kappa+1)(\lambda/\kappa+2)\ldots(\lambda/\kappa+s_i-1)]N!}{\lambda(\lambda+1)(\lambda+2)\ldots(\lambda+N-1)\, s_1!\, s_2!\ldots s_\kappa!}$$

we arrive at (2).

*Proof of (3).* Since the $h_j$ $(1 \leq j \leq \kappa)$ are the only structure-descriptions compatible with e′ and are pairwise mutually exclusive, we have by Bayes's theorem that

$$(9) \qquad P_\lambda(h_i', e') = \frac{P_\lambda(h_i') \cdot P_\lambda(e', h_i')}{P_\lambda(h_1') \cdot P_\lambda(e', h_1') + \ldots + P_\lambda(h_\kappa') \cdot P(e', h_\kappa')}.$$

Each $h_j'$ is a sentence of $L_N^\pi$, so $P_\lambda(h_j') = m_\lambda(h_j')$. Hence $P_\lambda(h_j')$ is the product of (6) and (7) as in the proof of (2). The value of $P_\lambda(e', h_j')$ is $(s_j+1)/(N)$, since in randomly selecting the $(N-1)$-membered sample described by e′ from the N-membered population described by $h_j'$ there are just $s_j+1$ different individuals that might remain unselected. We now have a value to substitute for each term of (9). Under this substitution an appropriate instance of the product of (6) and (7) appears as a factor of the numerator and each addend of the denominator, so we can remove from each of these terms the common factor (8). This substitution and simplification reduce (9) to

$$(10) \qquad P_\lambda(h_i', e') = \frac{[(s_i+\lambda/\kappa)/(s_i+1)] \cdot [(s_i+1)/(N)]}{[(s_1+\lambda/\kappa)/(s_1+1)] \cdot [(s_1+1)/(N)] + \ldots + [s_\kappa+\lambda/\kappa)/(s_\kappa+1)] \cdot [(s_\kappa+1)/(N)]}$$

(Comparison of (10) and (2) shows that the only difference between them is the likelihoods. In (2) the likelihoods have dropped out, but not so in (10).) Further obvious simplifications of (10) reduce it to (3).