PAPERS READ BEFORE THE SOCIETY.

1927–1928.

Meeting of the Aristotelian Society at 21, *Gower Street, London, W.C.*1, *on November* 7*th*, 1927, *at* 8 *p.m.*

I.—THE PRINCIPLES OF PROBLEMATIC INDUCTION.

THE PRESIDENTIAL ADDRESS.

By C. D. BROAD.

(1) DEFINITIONS OF INDUCTIVE ARGUMENTS.

BY *Problematic Induction* I mean any process of reasoning which starts from the premise that all, or a certain proportion of, *observed* S's have had the characteristic P, and professes to assign a probability to the conclusion that all, or a certain proportion of, S's will have this characteristic. It is assumed that no intrinsic or necessary connexion can be seen between the characteristics S and P. Where such a connexion can be seen, the fact that all observed S's have been found to be P can hardly be called a *logical premise;* it is at most a *psychological occasion* which stimulates the observer to intuit the intrinsic connexion between S and P. The latter process is called *Intuitive Induction* by Mr. Johnson, and I do not propose to consider it here. It is generally admitted that, when intuitive induction is ruled out, premises of the kind which we are considering can lead only to conclusions in terms of probability. This has been argued independently by Mr. Keynes and myself, and I am going to assume that it is true.

We can now classify Problematic Inductions as follows: (1) We may divide them up according to the nature of their premises. These may be (1·1) of the form, "All observed S's have been P," or (1·2) of the form, "A certain proportion of the observed S's have been P." Then (2) we may divide them up according to the nature of the proposition whose probability they profess to evaluate. This may be (2·1) of the form, "All S's whatever are P"; or (2·2) "The next S to be observed will be P"; or (2·3) "A certain proportion of the total number of S's are P." The following are the most important types of Problematic Induction: (A) If we combine a premise of the form (1·1) with a conclusion about a proposition of the form (2·1), we have what I will call *Nomic Generalization*, because it professes to assign a probability to a general *law* from an observed *regularity*. (B) Any argument whose conclusion is about a proposition of the form (2·2) we may call an *Eduction*, following Mr. Johnson. There will be two kinds of eduction, according to whether the premise is of the form (1·1) or of the form (1·2). These may be called respectively *Nomic Eductions* and *Statistical Eductions*. Finally (C), if we combine a premise of the form (1·2) with a conclusion about a proposition of the form (2·3), we get what may be called, in a wide sense, a *Statistical Generalization*. (This term would sometimes be confined to the special case in which the proportion in the premise is *the same* as the proportion which is considered in the conclusion.)

(2) The Logical Principles of the Argument.

I shall begin by considering artificially simplified cases. These will be of two kinds, viz., (*a*) the drawing of counters from a bag, and (*b*) the throwing of a counter whose opposite sides are of different colour. I shall try to state clearly both the principles of logic and probability that are presupposed and the assumptions about equiprobability that are made, and to show exactly where each enters into the argument. The only satisfactory way of doing this is to work out the arguments in detail.

(2·1) *Principles of Probability and Formal Logic.*

The following are the only ones that are needed : (1) If p and q be logically equivalent propositions, then p/h (*i.e.* the probability of p on the assumption that h is true) $= q/h$, whatever h may be. This may be called the *Principle of Equivalence.* (2) If p and q be any two propositions, then $(p \,.\, q)/h = (q/h)\ (p/qh) = (p/h)\ (q/ph)$.* Here $p \,.\, q$, $q \,.\, h$, and $p \,.\, h$ are conjunctive propositions—*i.e.* respectively, "p-and-q," "q-and-h," and "p-and-h." This may be called the *Conjunctive Principle.* (3) If p and q be two propositions which cannot both be true, then $(p\text{v}q)/h = p/h + q/h$. Here $p\text{v}q$ means the disjunctive proposition "either-p-or-q." This may be called the *Disjunctive Principle.* (4) If p be any proposition, and $q_1 \,....\, q_n$ be any set of mutually exclusive and collectively exhaustive alternant propositions, then $p \,.\equiv:\, pq_1 \,.\, \text{v} \,.\, pq_2 \,.\, \text{v} \,.\, ... \, pq_n$. This can be called the *Rule of Expansion.*

(2·2) *Bag Problems.*

In these we shall suppose that there is a bag which is known to contain n counters which are qualitatively indistinguishable except in respect of their colours. They are to be drawn out one by one, the colour is to be noted, and the counter is not to be replaced.

* We constantly use what is an immediate consequence of this, viz., $p/qh = \dfrac{(p/h)\ (q/ph)}{(q/h)}$.

(2·21) *Nomic Eduction applied to the Bag.*

Suppose that m counters have been drawn, and that all have been found to have a certain colour, *e.g.*, red. What is the probability that the next counter to be drawn will be red ?

Let us denote the proposition that the sth counter drawn is red by ρ_s. Let us denote our original information about the contents of the bag and the method of drawing by h. For shortness let us denote the conjunctive proposition $\rho_1\rho_2 - \rho_m$ by r_m. Then we are asked to evaluate the probability $\rho_{m+1}/r_m h$.

By the *Conjunctive Principle* $(r_m\rho_{m+1})/h = (r_m/h)\,(\rho_{m+1}/r_m h)$

$$\therefore\ \rho_{m+1}/r_m h = \frac{(r_m\rho_{m+1})/h}{r_m/h} = \frac{r_{m+1}/h}{r_m/h} \quad \ldots\ldots \quad (1)$$

Now there might originally have been in the bag either 0 or 1 or 2 or .... n red counters. Let us denote these $n+1$ mutually exclusive and collectively exhaustive alternant propositions by R_0, R_1, ... R_n respectively. It is evident that r_{m+1} is inconsistent with there having been originally less than $m+1$ red counters in the bag. Therefore combinations such as $r_{m+1}R_1$ vanish.

By the *Rule of Expansion* then we shall get

$$r_{m+1} . \equiv : r_{m+1} R_{m+1} . v. \ldots r_{m+1} R_n$$

$\therefore$ by the *Principle of Equivalence,* $r_{m+1}/h = \sum_{s=m+1}^{s=n} (r_{m+1}R_s)/h.$

By the *Conjunctive Principle* this is equal to

$$\sum_{s=m+1}^{s=n} (R_s/h)\,(r_{m+1}/R_s h).$$

Now a precisely similar argument will obviously lead to the result that

$$r_m/h = \sum_{s=m}^{s=n} (R_s/h)\,(r_m/R_s h).$$

If we now substitute these values in equation (1), we get

$$\rho_{m+1}/r_m h = \frac{\sum\limits_{s=m+1}^{s=n} (R_s/h)\,(r_{m+1}/R_s h)}{\sum\limits_{s=m}^{s=n} (R_s/h)\,(r_m/R_s h)} \quad \ldots\ldots\ldots \quad (2)$$

We must next evaluate terms of the form $r_{m+1}/R_s h$. By definition $r_{m+1}/R_s h = (\rho_1 \rho_2 \ldots \rho_m \rho_{m+1})/R_s h$. By repeated application of the *Conjunctive Principle* it follows that

$$\begin{aligned} r_{m+1}/R_s h &= (\rho_1/R_s h)\,(\rho_2/R_s h \rho_1)\,(\rho_3/R_s h \rho_1 \rho_2) \ldots (\rho_{m+1}/R_s h \rho_1 \ldots \rho_m) \\ &= (\rho_1/R_s h)\,(\rho_2/R_s h r_1)\,(\rho_3/R_s h r_2) \ \ldots\ (\rho_{m+1}/R_s h r_m). \end{aligned}$$

Now if there were originally s reds, and if any one of the n counters was equally likely to be drawn, it is evident that $\rho_1/R_s h = \frac{s}{n}$. At the next drawing there are $n - 1$ counters. If there were originally s reds, and if the one which has been drawn and not replaced was red, there are now $s - 1$ counters. If any one of the $n - 1$ remaining counters is equally likely to be drawn at the second drawing, it is evident that

$$\rho_2/R_s h r_1 = \frac{s-1}{n-1}.$$

So, on these assumptions, $r_{m+1}/R_s h = \frac{s}{n}\,\frac{s-1}{n-1}\,\frac{s-2}{n-2} \ldots \frac{s-m}{n-m}$

We must now explicitly notice that we have been making an assumption at this point about *Equiprobability*. This I will call the *First Premise about Equiprobability*.

It is evident that, on the same assumption,

$$r_m/R_s h = \frac{s}{n}\,\frac{s-1}{n-1} \ldots \frac{s-m+1}{n-m+1}.$$

If we substitute the values just obtained in the numerator and denominator of the right-hand side of equation (2) we get

$$\rho_{m+1}/r_m h = \frac{1}{n-m} \frac{\sum_{s=m+1}^{s=n} (\mathrm{R}_s/h)\, s(s-1)\ldots(s-m)}{\sum_{s=m}^{s=n} (\mathrm{R}_s/h)\, s(s-1)\ldots(s-m+1)} \quad . \; . (3)$$

This is the fundamental formula for Nomic Induction in the case of drawing counters from bags. It is evident that we can get no further unless we can evaluate the terms R_s/h. These are the antecedent probabilities of the various alternative possible original constitutions of the contents of the bag. Now some logicians and mathematicians, notably Laplace, have at this point argued as follows. They have assumed that, when nothing is known about the contents of the bag except that it originally contained n counters qualitatively indistinguishable save in respect of their colours, the $n+1$ possible alternatives—viz., that it contains 0 or 1 or .... n reds (*e.g.*)—are equally probable. And they defend this on the authority of the Principle of Indifference. On this assumption the factors R_s/h cancel out on the right-hand side of equation (3), and we get

$$\rho m + 1/r_m h = \frac{\sum_{m+1}^{n} s(s-1)\ldots(s-m)}{\sum_{m}^{n} s(s-1)\ldots(s-m+1)} \frac{1}{n-m},$$

and it can easily be shown that this is equal to $\frac{m+1}{m+2}$.

This is *Laplace's First Rule of Succession.*

Now this application of the Principle of Indifference has been severely and rightly criticised by Mr. Keynes. A simple way of seeing that it *must* be wrong is to put $m = 0$. We then reach the conclusion that, before any counter is drawn from a bag, the probability that the first to be drawn will be red is $\frac{1}{2}$. But exactly the same reasoning will show that the antecedent probability that the first to be drawn will be blue is $\frac{1}{2}$. Since no counter can be both blue and red it would follow, by the Disjunctive Principle, that the probability that the first to be drawn is *either* red *or* blue must be $\frac{1}{2} + \frac{1}{2}$, *i.e.*, that it is certain to be one or the other, and this is plainly absurd. Nor is it difficult to see why Laplace's application of the Principle of Indifference is wrong. A set of counters $c_1 \ldots c_n$ which are all red can arise only in one way; but a set of counters $c_1 \ldots c_n$ in which one is red can arise in n ways, since it can arise through c_1 being red or c_2 being red or ... c_n being red. Thus the various alternatives R_0, R_1, ... R_n are not all exactly alike in internal complexity, for each is analysable into various numbers of sub-alternatives. It is therefore illegitimate to apply the Principle of Indifference to them. We must therefore reject Laplace's Rule.

This throws us back on the original question. Can we evaluate the probabilities R_s/h? If we cannot, the formula (3), though valid, is useless. Let us try the following way. It is known that any counter has some one colour, including for this purpose black and white as colours. Let us suppose that there are ν distinguishable colours, including black and white. Then any counter taken at random is equally likely to have any one of these ν colours. Let us call this the *Second Premise about Equiprobability*. It must have one and can have only one. Hence the antecedent probability that any counter chosen at random shall have a certain assigned colour—*e.g.*, red—is $\frac{1}{\nu}$. Now a set of n counters containing exactly s red ones can arise in ${}^nC_s$ ways. A typical case might be written

$$\gamma_1 \gamma_2 \ldots \gamma_s \cdot \bar{\gamma}_{s+1} \cdot \bar{\gamma}_{s+1} \ldots \bar{\gamma}_n.$$

Here γ_s is the proposition, "the counter C_s is red"; and $\bar{\gamma}_{s+1}$

is the proposition the counter "C_{s+1} is not red." It is obvious that the probability of any such typical case is $\left(\frac{1}{\nu}\right)^s\left(1-\frac{1}{\nu}\right)^{n-s}$ since the fact that one counter is or is not red is irrelevant to the question whether any other counter is or is not red. It follows that $R_s/h = {}^nC_s\left(\frac{1}{\nu}\right)^s\left(1-\frac{1}{\nu}\right)^{n-s}$.

If we substitute these values in the formula (3) and do a little straightforward algebra, we get

$$\rho_{m+1}/r_m h = \frac{1}{n-m}\,\frac{\sum\limits_{m+1}^{n}\dfrac{(\nu-1)^{n-s}}{(n-s)!\,(s-m-1)!}}{\sum\limits_{m}^{n}\dfrac{(\nu-1)^{n-s}}{(n-s)!\,(s-m)!}}$$

It is easy to prove that the sum in the numerator comes to $\frac{\nu^{n-m-1}}{(n-m-1)!}$, and that the sum in the denominator comes to $\frac{\nu^{n-m}}{(n-m)!}$. We thus reach the extremely unsatisfactory conclusion that

$$\rho_{m+1}/r_m h = \frac{1}{\nu},$$

i.e., that although all the counters that we draw are found to be red, the probability that the next to be drawn will be red remains exactly what it was when no counters had been drawn.

(2·22) *Nomic Generalization applied to the Bag.*

It is now easy to pass from Nomic Eduction to Nomic Generalization about the bag. The question now is: Given that m counters have been drawn, and that all have been red, what is the probability that all the counters in the bag are red? This simply means that we have to evaluate $R_n/r_m h$. By the *Conjunctive Principle* we can write

$$R_n/r_m h = \frac{(R_n/h)\,(r_m/R_n h)}{r_m/h}$$

But $r_m/R_n h = 1$, since the first m to be drawn *must* be red if all in the bag are red. Hence $R_n/r_m h = \dfrac{R_n/h}{r_m/h}$. Making use of the *Principle of Equivalence* and the *Conjunctive* and *Disjunctive Principles*, we get

$$R_n/r_m h = \frac{R_n/h}{\sum\limits_{s=m}^{s=n} (R_s/h)\,(r_m/R_s h)} \quad \ldots\ldots \quad (4)$$

This is the fundamental formula for Nomic Generalization. On the false Laplacean assumption that all terms of the form R_s/h are equal, it is easy to prove that the fraction on the right-hand side of (4) becomes $\dfrac{m+1}{n+1}$. This is *Laplace's Second Rule of Succession*. On the true assumption about equiprobability—viz., that any individual counter is equally likely to have any one of the ν distinguishable colours—it is easy to prove that the fraction on the right-hand side of (4) becomes $\left(\dfrac{1}{\nu}\right)^{n-m}$. Thus, even on the true assumption, the probability of the law that *all* the counters in the bag are red does increase with every counter which is drawn and found to be red, though the probability that the *next* counter to be drawn will be red does not increase.

(2·23) *Statistical Generalization applied to the Bag.*

It remains to consider the most general problem, viz., that of *Statistical Generalization*, for the artificial case of the bag of counters. The problem may be stated as follows. We have drawn m counters and have found that μ of them were red and the rest non-red. What is the probability that there were originally x red counters in the bag? We will denote the proposition that m counters have been drawn and that μ of them are red by $r_{\mu,m}$. Then the probability which we have to evaluate is $R_x/r_{\mu,m}.h$. Applying the same principles as before, we easily find that

$$R_x/r_{\mu,m}h = \frac{(R_x/h)\,(r_{\mu,m}/R_xh)}{\sum\limits_{s=\mu}^{s=n-m+\mu}(R_s/h)\,(r_{\mu,m}/R_sh)}.$$

The limits of the summation in the denominator are determined by the fact that there cannot have been less than μ reds, since μ reds have been drawn, and there cannot have been more than $n-m+\mu$ reds, since $m-\mu$ non-reds have been drawn.

It remains to evaluate $r_{\mu,m}/R_sh$. The μ reds which have been found in the m counters that have been drawn might have been presented in ${}^mC_\mu$ different orders. We are justified in assuming, on the grounds of the Principle of Indifference, that any order of presentment is as likely as any other, with respect to the available data. This constitutes the *Third Premise about Equiprobability*. It was not needed in the two previous problems. Hence, to find the required probability, we may take a single typical order of presentment, *e.g.*, $r_\mu\bar{\rho}_{\mu+1} \ldots \bar{\rho}_m$, and multiply the probability of this by ${}^mC_\mu$. Now it is evident that

$$(r_\mu\bar{\rho}_{\mu+1} \cdots \bar{\rho}_m)/R_sh =$$

$$\frac{s}{n}\frac{s-1}{n-1}\cdots\frac{s-\mu+1}{n-\mu+1}\left(1-\frac{s-\mu}{n-\mu}\right)\cdots\left(1-\frac{s-\mu}{n-m+1}\right).$$

So we finally reach the equation

$$(5) \quad \ldots \quad R_x/r_{\mu,m}\,h = \frac{(R_x/h)\,x\,(x-1)\ldots(x-\mu+1)\,(n-x)\,(n-x-1)\ldots(n-x-m+\mu+1)}{\sum\limits_{s=\mu}^{s=n-m+\mu}(R_s/h)\,s\,(s-1)\ldots(s-\mu+1)\,(n-s)\,(n-s-1)\ldots(n-s-m+\mu+1)}$$

This is the fundamental formula for Statistical Generalization as applied to the case of the bag. On the false Laplacean assumption that all probabilities of the form R_s/h are equal, it is easy to prove that the most probable value of x is such that $\frac{n}{x} = \frac{\mu}{m}$, *i.e.*, that the most probable proportion of reds in the whole contents of the bag is the same as the proportion of reds in the set of counters drawn and observed. On the true assumption that $R_s/h = {}^nC_s \left(\frac{1}{\nu}\right)^s\left(1 - \frac{1}{\nu}\right)^{n-s}$, it is easy to show that the most probable value of x is such that $\frac{x - \mu}{n - m} = \frac{1}{\nu}$, *i.e.*, that the most probable proportion of reds among the remaining $n - m$ counters is $\frac{1}{\nu}$. Now this is exactly what was the most probable proportion of reds in the bag before any counters were drawn, for the antecedently most probable number of reds is that value of s which makes ${}^nC_s \left(\frac{1}{\nu}\right)^s\left(1 - \frac{1}{\nu}\right)^{n-s}$ a maximum, and this is the nearest integer to $\frac{n}{\nu}$. So, no matter how many counters have been drawn, and no matter what may have been the proportion of reds found among them, the proportion of reds which was *antecedently* most probable for the whole contents of the bag will still be the most probable proportion of reds in the remainder. It would be hard to imagine a less satisfactory result.

(2.24) *Summary.*

To sum up. We have seen exactly how the formal principles of probability and logic enter into the inductive arguments about the bag. We have seen that in *every* case *two* different premises about equiprobability are needed, one in the general course of the argument and another in order to evaluate the terms R_s/h.

We have seen that, in Statistical Generalizations, a *third* premise about equiprobability is needed. Finally, we have seen that the Laplacean assumption for evaluating R_s/h is certainly false, and that, when the true assumption is made, the inductive argument fails to establish any high probability.

(2.3) *Problems on Throwing a Counter.*

We here suppose that there is a single counter which is geometrically regular, and has a red face and a white face. I shall suppose that, for all we know at the outset, this counter may be loaded to any extent either in favour of red or in favour of white.

We must first define the notion of "loading." I shall say that "the counter is loaded to a degree s in favour of red," if, and only if, the antecedent probability of its turning up red would be s for anyone who knew in detail how it was constructed. I will denote this proposition by R_s. It is evident from the geometrical fairness of the counter and our complete absence of information as to its loading that $R_s/h = W_s/h$. Again, if the counter be so constructed that the antecedent probability of its turning up red is s for anyone who knows its construction, it is evident that the antecedent probability of its turning up white is $1 - s$ for the same person. Hence $R_s/h = W_{1-s}/h$. We can therefore confine ourselves to loading for red, for we shall cover all the probabilities if we let s range from 0 to 1. That is, $\sum_0^1 R_s/h = 1$.

(2.31) *Statistical Eduction applied to the Counter.*

Let us now suppose that this counter is thrown n times, and that, on m occasions, it is found to turn up red. What is the probability that the next throw will be red? It is easy to prove by exactly the same methods as we used for the bag that

$$\rho_{n+1}/r_{m,n}\,h = \frac{\sum\limits_{0}^{1} (R_s/h)\, s^{m+1} (1-s)^{n-m}}{\sum\limits_{0}^{1} (R_s/h)\, s^{m} (1-s)^{n-m}} \quad \ldots\ldots \quad (6)$$

This is the formula for Statistical Eduction in the case of the counter. More complicated formulæ could be got for a geometrically regular die with ν sides and a different colour on each, but the principles and premises would be of exactly the same kind.

(2.32) *Statistical Generalization applied to the Counter.*

The most general formula would be that for Statistical Generalization. Here we suppose that the counter has been thrown n times, and that red has turned up m times. It is now to be thrown a further n' times, and we ask : " What is the probability that red will turn up m' times in these further n' throws ? " It is easy to prove by the same methods as before that

$$r_{m+m',n+n'}/r_{m,n}h = {}^{n'}C_{m'}\, \frac{\sum\limits_{0}^{1} (R_s/h)\, s^{m+m'} (1-s)^{n+n'-m-m'}}{\sum\limits_{0}^{1} (R_s/h)\, s^{m} (1-s)^{n-m}} \quad \cdot (7)$$

(3) The Causal Pre-suppositions of the Arguments.

Now these formulæ are precisely analogous to those which would be got for the case of drawing counters from a bag, on the supposition that each counter drawn is *replaced* before the next draw. The notion of loading, however, brings out a fundamental pre-supposition of all inductive arguments, which, though really equally present in the case of experiments with bags, is there more likely to be overlooked. The notion of loading is the notion of a constant cause-factor which operates throughout the whole series of throws and co-operates with other and variable cause-factors to determine the result of each throw. Similarly, if each counter is replaced after it has been drawn from the bag and before the next draw is made, the original constitution of the contents of the bag is a constant cause-factor which operates throughout the whole series of drawings and combines with other cause-factors which vary from draw to draw to determine the actual result of each draw. In the case where the counters are not replaced after each draw we have not indeed a *constant* cause-factor; but we know how the original cause-factor, whatever it may have been, has been altered by the results of the previous drawings.

It might perhaps be suggested that there is one fundamental logical difference between the problems on drawing counters from a bag and the problems on throwing a single counter. It might be said that, in the former, we had to use the *First Premise about Equiprobability*, and that, in the latter, it is not used. I think that this is a mistake due to an inadequate analysis of the notion of loading in the latter problems. In the bag-problems the *First Premise about Equiprobability* is needed in order to pass from the datum that there is such and such a proportion of reds in the bag to the probability that the next counter drawn will be red. Now, of course, there is nothing directly analogous to this in the counter problems. But consider the notion of loading. I defined the statement that "a counter is loaded to degree s in favour of red" to mean that it is so constructed that, relatively to a knowledge of the details of its construction, the probability of its falling with the red side upward is s. Now how could one pass from the knowledge of its structure to the probability of its falling with the red side upwards? The relevant point about its structure would be the position of its centre of gravity with respect to its geometrical centre. We should then have to consider all the possible angles which the plane of the counter could make with the table at the moment of contact, and to find in what proportion of these a counter with its centre of gravity in the given position would inevitably fall over with the red side upwards. But this would not enable us to evaluate the probability of such a counter falling with the red side upwards unless we knew the antecedent probabilities of its striking the table at each of the possible angles, and these antecedent probabilities could not be evaluated without some assumption about equiprobability. Thus the alleged distinction between the two types or problem must be rejected.

I think that we are now justified in making the following assertion, which, if true, is very important. Every inductive argument, whether it be a nomic generalisation, an eduction, or a statistical generalisation, equally presupposes the notion of causal determination. It presupposes the following proposition, which I will call the *Fundamental Causal Premise* : "The result of each experiment is completely determined by a total cause composed of cause-factors of two different types. (i) A factor which is known to be constant throughout all the experiments, or whose variations, if it varies, are known at every stage. (ii) A very large number of variable cause-factors, each of which is as likely to vary in one direction as in another of all the directions in which it can vary." The course of every kind of inductive argument is the same. It argues (*a*) backward from the actual results to the present probabilities of the various alternative possible cause-factors of the first kind, and (*b*) forward from these to the probability of a proposed future result. If this be true it is important for several reasons. (i) There are people who profess to reject the notion of causal determination and yet to believe in the validity of some inductive arguments. Some of them think that this position is consistent, provided they content themselves with Eduction and Statistical Generalization and do not attempt Nomic Generalization. If I am right, this is a complete mistake. You must hold either that there is something in causation, or that there is nothing in induction. To reject causation and accept induction is not, as is commonly supposed, hard-headed; it is merely muddle-headed. (ii) On the other hand, there are people who think that, if we could only be sure of the Law of Universal Causation, all the troubles of induction would be over. This is a profound mistake, for the following reasons: (*a*) Even if we knew that the *Fundamental Causal Premise* mentioned above is true, we should still be faced with the question whether any inductive argument can establish a respectable probability for any proposition about as yet unobserved things or events. In our arguments about bags and counters we have assumed this premise, but we have reached only miserably low probabilities. It is thus evident that this premise, though necessary, is not sufficient to justify the claims of induction to make some propositions about unobserved things or events highly probable. (*b*) The Law of Causation is not equivalent to the *Fundamental Causal Premise.* In one respect it is more sweeping. We do not need to assume that *every* event is completely determined by causes. All that we need to assume is that the result of each of our experiments is completely determined by causes. This, however, is not logically important, for knowledge of the general principle would guarantee the particular application; and the general principle might be self-evident, while the particular case, apart from reference to the general principle, might not be self-evident. The really serious objection is that, in another respect, the Law of Causation is not determinate enough. It is not enough to know the general fact that the result of each of our experiments is causally determined. We need to know the more specific fact that it is determined in the particular way mentioned in the *Fundamental Causal Premise.* We need to know that we are in presence of a constant cause-factor, or in presence of a cause-factor whose variation from experiment to experiment is known.

(4) Conditions for a High Final Probability.

Still confining our attention to artificial problems, we can now raise the question: "What further premises would be needed in order that the argument may give a *high* probability to propositions about unobserved things or events?" We see at once that the trouble always arises over the antecedent probabilities of the various permanent cause-factors, *i.e.*, over terms of the form R_s/h. Laplace, by making the preposterous assumption that all these are equal, made them cancel out, and arrived at conclusions which are much too good to be true. We, by making what seems to be the only assumption about equiprobability that is reasonable for the colours of counters of whose origin nothing is known, were able to give certain values to the terms R_s/h; but, by so doing, we arrived at probabilities which are almost beneath contempt. In the case of a die or counter of unknown origin and construction, it is difficult to see that there is any reasonable principle on which the antecedent probabilities of the various possible degrees of loading can be assigned. Here the Laplacean assumption is not so obviously absurd as in the case of counters in a bag. For a given degree of loading is not *prima facie* analysable into a group of a certain number of equiprobable sub-alternatives, as a given proportion of red counters in a bag is. It is not unreasonable to say that, if nothing is known of the construction of a die or counter, any kind and degree of loading is as likely as any other; and, on this supposition, the Laplacean Rules of Succession follow easily from our formulæ (6) and (7); for we may reasonably assume that the probability of any exact degree of loading s is infinitesimal. We may therefore substitute for R_s/h the expression $\phi(s)\,ds$, where ϕ is an unknown function. Formula (6) then becomes

$$\rho_{m+1}/r_{m,n}h = \frac{\int_0^1 \phi(s)\, s^{m+1}(1-s)^{n-m}\, ds}{\int_0^1 \phi(s)\, s^m (1-s)^{n-m}\, ds} \quad . \quad . \quad . \quad (8)$$

The Laplacean assumption amounts to supposing that $\phi(s)$ is a constant. It is then easy to prove, by means of Γ- and B-functions, that the expression on the right $= \frac{m+1}{n+2}$. If we put $m = n$ this becomes $\frac{m+1}{m+2}$, which is Laplace's First Rule of Succession. The other rules follow in the same way from formula (7) on the same assumption. If we suppose that m and n in formula (8) tend to ∞, $\frac{m+1}{n+2}$ will tend to the value $\frac{m}{n}$. The proposition that

$$\underset{\substack{m \to \infty \\ n \to \infty}}{\mathrm{Lt}} \; \rho_{m+1}/r_{m,n}h = \frac{m}{n} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (9)$$

may be called the *Inverted Bernoulli Theorem*, which is thus a consequence of the Laplacean Assumption in the case of dies and counters.

Now, if the supposition that all degrees of loading for a counter, or all original proportions of red counters in a bag, are *equally* likely enables the formulæ of eduction and of nomic generalization to establish reasonably high probabilities, it presumably follows *a fortiori* that any assumption which favours a high degree of loading or a large original proportion of counters of the same colour will act still more strongly in the same direction. Let us call this the *Assumption of Loading*. It is quite distinct from, and independent of, the *Fundamental Causal Premise*. The latter is the assertion that there is a cause-factor of a certain kind operating throughout the whole series of experiments, and it is necessary if *any* inductive argument is to establish *any* probability at all, high or low. The former is an assumption about the relative antecedent probabilities of the various possible cause-factors of the type required by the *Fundamental Causal Premise*. It is required, not to validate inductive arguments as such, but to validate the claims of *some* of them to produce *high* probabilities.

(5) Transition from Artificial to Natural Cases.

We have now completed our analysis of inductive arguments, as applied to artificial cases, and have seen exactly what are the constitutive or ontological conditions which must be fulfilled if such arguments are to be both valid and fruitful. It remains to consider whether there is any reason to believe that these conditions are fulfilled in nature. Let us take the case of investigating swans, finding that all observed swans are white, and arguing to the probability that the next swan, or all further swans, will be white; and let us compare this with the artificial cases which we have so far considered. The analogies are as follows: All swans, past, present and future, may be compared to the total contents of the bag. Drawing a counter, noting its colour, and not replacing it, may be compared to catching a swan, noting its colour, and taking care not to count the same swan again among one's data. So far the analogy is complete. But there are many important differences, and all the more obvious of these are unfavourably relevant to induction as applied to nature, in comparison with induction as applied to the artificial case of the bag. (1) The number of swans past, present and future is unknown; but it is almost certainly very great as compared with the number that have been observed up to any given moment. This would be fatal to the attempt to give a high probability to the proposition that *all* swans are white, even if we accepted the Laplacean assumption, for $\frac{m+1}{n+1}$ would be vanishingly small. It seems to me doubtful whether any assumption of loading which had the faintest plausibility would suffice to give a high probability to a law like "all swans are white," when the evidence is only that all observed swans have been white. On the other hand, the probability that the next swan to be observed will be white might be reasonably large, in spite of the disparity between m and n, if we could accept some assumption about loading much less radical than Laplace's. With Laplace's assumption it is $\frac{m+1}{m+2}$, which is absurdly high. On the assumption that any swan is antecedently as likely to have any one colour as any other, it is $\frac{1}{\nu}$, no matter how great m may be. In either case it is independent of n. It is thus reasonable to suppose that, with some assumption about loading intermediate between these two, the probability that the next swan will be white would be fairly high if m were fairly large, in spite of the fact that n is incomparably larger than m. (2) The same swan might happen to be observed several times, and to be mistaken for different swans. This will cause us to think that m is larger than it really is. This source of weakness is absent when we are dealing with events, and not with relatively permanent substances, like swans; for the same events cannot be observed twice over by the same observer, and we can generally say with fair confidence whether different people are observing the same or different events. We may perhaps sum up this difficulty by saying that the investigation of substances in nature is intermediate between the case where a counter is never put back after being drawn, and the case where a counter is always put back after being drawn. (3) We come now to the difference which is most serious. In the case of the counters in the bag, we assumed that, at any drawing, any counter then in the bag was equally likely to be drawn, and this was an essential premise of the inductive argument. Now, if the bag be not too large, and does not have pockets in it, and the counters be well mixed, this assumption seems to be justified; but it most certainly is not justified in applying induction to nature. It breaks down for two reasons. (i) Spatially, only a very limited range is open to our observation. There may be swans on other planets, and, if there are, none of them could possibly have been included among our data. (ii) Similar remarks apply to time. Obviously the swans that could be observed up to a given date could not include any swans that began to exist after that date, and it is equally certain that our observations (including the reports of our ancestors) do not include swans that existed more than a few thousand years ago. It is as if the bag were so large that the greater part of its contents could not possibly be reached by us. The result is that the *First Premise about Equiprobability* breaks down, and, as we saw, every kind of inductive argument requires this premise. An attempt has been made to evade this criticism by appealing to the principle that mere difference of spatial and temporal position is irrelevant. Even Mr. Keynes seems to attach some importance to this principle; but, whether it be true or false, it is surely altogether beside the mark; for there is no such thing as *mere* difference of spatio-temporal position. If A is in a different place from B, the things that immediately surround A will differ from those which immediately surround B. If A exists at a different time from B, the things and events which are contemporary with or immediately precedent to A will differ from those which are contemporary with or immediately precedent to B; and no one can assert that a difference in a thing's near neighbours in space and time is always irrelevant to its other properties.

It is clear, then, that there are important differences between any subject of inductive enquiry in nature and the artificial cases for which we have worked out the general theory of inductive argument; and all the differences which we have mentioned are unfavourable to induction as applied to nature. The probabilities which can be reached in the artificial examples are the unattainable upper limits of the probabilities that can be reached by the application of induction to nature. Half, and only half, of this fact has been recognized by most writers on Inductive Logic.

They saw the special sources of weakness in the application of induction to nature, and all the various eliminative methods which they have recognized and formulated are simply ways of reducing these sources of weakness to a minimum ; but, having exercised themselves in formulating methods of elimination, they thought that they had done all that was required of them. They failed to notice that they had merely reduced certain obvious sources of weakness, and had given no positive theory of inductive reasoning at all.

We can now see clearly that two tasks must be accomplished if the application of inductive arguments to nature is to be valid, and is to lead to reasonably probable conclusions. (i) We must have some reason to believe that something analogous to "loading" exists in nature, and that certain kinds of "loading" are antecedently much more probable than others. (ii) We must somehow get over the objection that, since future and remote events could not have been included among our observed data, the *First Premise about Equiprobability*, on which the validity of every kind of inductive argument rests, seems to have broken down. It is evident that three general questions can be raised about inductive inference. These questions may be described as the *logical*, the *ontological* and the *epistemic* question. The *logical* question is to determine the formal character of inductive arguments as such ; to state the principles of formal logic and probability which they use, and to see exactly how these enter into the argument ; and to discover what premises about equiprobability they require. It includes the further question as to what further premises are required if the argument is to establish, not merely *some* probability, but a reasonably high probability. This problem has now been completely solved, the first part in detail and the second in outline. The *ontological* question is to determine the minimum assumption about the general structure of nature which will guarantee that the conditions, required in order that an inductive argument applied to natural phenomena may establish a high probability, are fulfilled. If this can be solved there will still remain an *epistemic* question. Do we *know* that nature has this general structure ? And, if so, how do we know it ? Or, if we do not know *it*, do we at least know that it is highly probable ? And, if so, how do we know this ? It is to these questions that we must now address ourselves. We will begin with the ontological question.

(6) The Ontological Question.

The only good treatment of this question with which I am acquainted is contained in Chaps. XXI, XXII and XXXIII of Mr. Keynes's *Treatise on Probability*. Mr. Keynes's theory may be called the *Theory of Generators*. I think that Mr. Keynes's theory is susceptible of improvement in at least two respects. (1) It is stated very briefly, and when one tries to think it out in detail one finds that it is necessary to recognize certain distinctions which Mr. Keynes does not explicitly make, and to deal with certain complications which he does not explicitly consider. (2) I think it is possible to show that generators may be regarded as convenient parameters for stating and working out the theory, but that all that is needed can be accomplished without assuming that they actually exist in nature. I propose, therefore, (1) to begin by assuming the existence of generators, and simply to improve (as I think) the formal exposition of the theory. Then (2) I shall show that the actual existence of generators need not be assumed.

(6.1) *Definitions.*

Suppose there is a certain set of determinable characteristics Γ_1, Γ_2, ... Γ_n, which are logically and causally independent of each other. This means that any of the $2^n - 1$ combinations, which can be got by taking them one at a time, or two at a time, or ... n at a time, is both logically and causally possible. Let C be another determinable characteristic. Suppose that there is a certain sub-set of characteristics of the first kind, *e.g.*, $(\Gamma_1, \Gamma_2, \ldots \Gamma_r)$, and suppose

(i) That anything which had all the characteristics $\Gamma_1 \ldots \Gamma_r$ would have the characteristic C, and

(ii) That anything which had only a selection from the set $\Gamma_1 \ldots \Gamma_r$ might lack C.

Then we say that the set $(\Gamma_1 \ldots \Gamma_r)$ *generates* C. We call $(\Gamma_1 \ldots \Gamma_r)$ a *generating set* for C, and we call each of the characteristics $\Gamma_1, \Gamma_2 \ldots \Gamma_r$ *generating factors* of C.

These definitions obviously leave it possible (1) that a generating set may generate several characteristics, *e.g.*, $(\Gamma_1 \ldots \Gamma_r)$ might generate C and C′. (2) A characteristic may be generated by several different generating sets—*e.g.*, there is nothing in the definition to exclude the possibility that C is generated by (Γ_{r+1}) and $(\Gamma_2 \ldots \Gamma_r, \Gamma_{n+2})$ as well as by $(\Gamma_1 \ldots \Gamma_r)$. All that is excluded is that C should be generated, *e.g.*, by $(\Gamma_1\ \Gamma_2)$ as well as by $(\Gamma_1 \ldots \Gamma_r)$.

I am going to assume, however, until further notice, that the same characteristic C does not in fact have more than one generating set. On that assumption we can talk of *the generator of C*. To say that $(\Gamma_1 \ldots \Gamma_r)$ is *the* generator of C means that

(i) Anything that had $\Gamma_1 \ldots \Gamma_r$ would have C, and

(ii) Anything that had C would have $\Gamma_1 \ldots \Gamma_r$.

It will, of course, still remain possible that some generating sets generate more than one characteristic. The above assumption will be called the *Denial of Plurality of Generators*.

A generating set which contains only one factor—*e.g.*, (Γ_1)—will be called a *set of the first-order*.

A generating set which contains two, and only two, factors—*e.g.*, $(\Gamma_1\ \Gamma_2)$—will be called a *set of the second-order*.

A characteristic which is generated by a generator of the rth order will be called a *characteristic of the r th order*. If we deny plurality of generators, each generated characteristic will be of one and only one order.

The next conception that we need to introduce is that of *fertility*. A generating set is said to be *sterile* if it generates no characteristic. If it generates s characteristics it is said to *have fertility* s. Thus a sterile generating set is one whose fertility is 0.

The fertility of a *generating factor* may be defined as follows: It is the sum of the fertilities of all the generating sets of which it is a factor.

A *generalization* is a universal proposition connecting two mutually exclusive sets of generated characteristics. Thus, the proposition, "Anything that had $C_1C_2C_3$ would have C_4C_5" is a generalization. A generalization whose subject consists of μ characteristics, and whose predicate consists of ν characteristics, is said to be a "generalization of the form $g_{\mu\nu}$."

(6.2) *Assumptions.*

Let us suppose that there are N generated determinable characteristics $C_1C_2 \ldots C_N$, and that they are *logically* independent of each other—*i.e.*, that there is no *a priori* objection to the occurrence of each of the $2^N \ldots 1$ selections that can be made by taking them 1 or 2 or .... N at a time. Let us suppose that there are n generating factors $\Gamma_1 \ldots \Gamma_n$.

We will assume—

(i) That each of the characteristics $C_1 \ldots C_N$ is generated by *some* generating set composed of factors selected from the n generating factors.

(ii) That no generating *factor* is superfluous. This means that every one of the n factors is a factor in *some* generating set which generates *some* characteristic in the set $C_1 \ldots C_N$. (It is, of course, quite possible that some generating *sets* may be sterile and generate no characteristic in the set $C_1 \ldots C_N$.)

(iii) That $N > n$.

(iv) That each of the generated characters is generated by only one generating set. (*Denial of Plurality of Generators.*)

It follows at once that the N generated characteristics fall into $2^n - 1$ mutually exclusive classes (some of which may be null) corresponding to the $2^n - 1$ generating sets. The set of characteristics generated by the rth order generating set $(\Gamma_1 \ldots \Gamma_r)$ may be denoted by $\alpha_{1 \ldots r}$, and similarly for the rest.

(6.3) *Application to Nomic Generalizations.*

Suppose that a certain thing has been found to have the characteristics $C_1 \ldots C_\mu\ C_{\mu+1} \ldots C_{\mu+\nu}$. What is the antecedent probability of the generalization: "Anything that had $C_1 \ldots C_\mu$ would have $C_{\mu+1} \ldots C_{\mu+\nu}$?" This is a generalization of the form $g_{\mu\nu}$.

This generalization will be true if, and only if, the factors which are required to generate the predicate set are contained among the factors which are required to generate the subject set. Suppose, *e.g.*, that $C_1 \ldots C_\mu$ require between them for their generation $\Gamma_1 \ldots \Gamma_r$. Then anything that had $C_1 \ldots C_\mu$ would have $\Gamma_1 \ldots \Gamma_r$. Suppose that $C_{\mu+1} \ldots C_{\mu+\nu}$ between them required a selection from $\Gamma_1 \ldots \Gamma_r$. Then anything that had $\Gamma_1 \ldots \Gamma_r$ would have this selection, and anything that had this selection would have $C_{\mu+1} \ldots C_{\mu+\nu}$. Hence the original generalization must be true.

Now the μ subject properties might between them require 1 or 2 or ... n generating factors. Let us denote the proposition that they require exactly r generating factors by μ_r. Let us denote the proposition that the ν predicate properties require between them exactly s generating factors by ν_s. It is evident that we need not consider cases in which $s > r$, for it would then be impossible that the generating factors of the predicate should be contained in those of the subject. A typical generalization of the form $g_{\mu\nu}$ would be: "Everything that had $C_1 \dots C\mu$ would have $C_{\mu+1} \dots C_{\mu+\nu}$. Let us denote this by $g^{12\dots\mu}_{\mu+1\dots\mu+\nu}$. We want to evaluate the probability $g^{1\dots\mu}_{\mu+1\dots\mu+\nu}/h$, when h includes the assumptions enumerated in (6.2).

By using the *Rule of Expansion* and the *Principle of Equivalence* we find that

$$g^{1\dots\mu}_{\mu+1\dots\mu+\nu}/h = \sum_{r=1}^{r=n}\sum_{s=1}^{s=r}(g^{1\dots\mu}_{\mu+1\dots\mu+\nu}\,\mu_r\,\nu_s)/h, \text{ which}$$

$$= \sum_{r=1}^{r=n}\sum_{s=1}^{s=r}[(\mu_r\,\nu_s)/h]\,[g^{1\dots\mu}_{\mu+1\dots\mu+\nu}/\mu_r\,\nu_s\,h], \text{ by the}$$

Conjunctive Principle.—Now the probability that the s generating factors required by the predicate are wholly contained among the r generating factors required by the subject is obviously the ratio of the number of ways of choosing s things out of r things to the number of ways of choosing s things out of n things, *i.e.*, $\dfrac{r!\,(n-s)!}{n!\,(r-s)!}$. Hence

$$g^{1\dots\mu}_{\mu+1\dots\mu\dots\nu}/h = \sum_{r=1}^{r=n}\sum_{s=1}^{s=r}[(\mu_r\nu_s)/h]\frac{r!\,(n-s)!}{n!\,(r-s)!} \quad .(10)$$

The antecedent probabilities of *all* generalizations of the form $g_{\mu\nu}$ will, of course, be equal.

(6.31) *Effect of the Relative Values of n and* N.

So far we have made no use of the assumption that N, the number of generated characteristics, is greater than n, the number of generating factors. I shall now prove that if $n \overline{\gtrless} N$, *every* generalization of every form *might* be false, and that if $n > N$ *some* generalization *must* be true. (1) Suppose that $N = n - p$. Call the N generated characteristics $C_1 C_2 \ldots C_{n-p}$. Then it is evidently possible that (Γ_1) generates C_1 and it only; that (Γ_2) generates C_2 and it only; and that (Γ_{n-p-1}) generates C_{n-p-1} and it only. Then the remaining characteristic C_{n-p} *cannot* require any of the factors $\Gamma_1 \ldots \Gamma_{n-p-1}$, and it *must* require *all* the factors Γ_{n-p} to Γ_n, for otherwise these factors will be sterile and superfluous. Thus the set $(\Gamma_{n-p} \ldots \Gamma_n)$ generates only a single characteristic C_{n-p}, and all the sub-sets within $(\Gamma_{n-p} \ldots \Gamma_n)$ are sterile. It follows that *no* generalization could be true, on this supposition, and this supposition is clearly possible if $N < n$. Therefore, if $N < n$, all generalizations might be false.

(2) Suppose that $N = n$. Then it is plainly possible that all the generated characteristics should be of the first order. If so, each first-order generating set must generate one and only one characteristic, and all other generating sets must be sterile. Under these conditions it is impossible that any generalization should be true, and these conditions can exist if $N = n$. Therefore, if $N = n$, all generalizations might be false.

(3) Suppose that $N > n$. Two cases will arise, viz., (3.1) that $N > 2^n - 1$, and (3.2) that $N \overline{\gtrless} 2^n - 1$. On the first alternative at least one of the generating sets must generate more than one characteristic. In that case *at least two* independent generalizations must be true. For, to take the weakest case that is compatible with the conditions, suppose that only one generating set generates more than one characteristic, and that this set generates only two characteristics C_1 and C_2. Then both the generalizations $g^1{}_2$ and $g^2{}_1$ must be true.

On the second alternative it is possible that each characteristic is generated by a different generating set, and therefore that each generating set which is fertile generates only one characteristic. In that case there will be no simply convertible generalizations like $g^1{}_2$ and $g^2{}_1$; but, nevertheless, even in this most unfavourable case, there will be some true generalizations. Suppose, *e.g.*, that $N = n + 1$. Either *some* or *none* of the generating sets of these $n + 1$ characteristics have a fertility greater than 1. If any do, then there must be some true generalizations. Suppose, then, that each of the $n + 1$ characteristics is generated by a different generating set. We then have $n + 1$ fertile generating sets, each of unit fertility. The rest of the generating sets are all sterile. It follows that these $n + 1$ generating sets must between them take up all the n generating factors, for otherwise some generating factors would be sterile and superfluous. Let us call the generating sets $\gamma_1, \gamma_2, \dots \gamma_{n+1}$. *Suppose, if possible, that *every* γ contains a factor not contained in *any* of the remaining γ's. From each γ select such a Γ. The Γ's thus associated with the $n + 1$ γ's cannot all be different, for there are only n Γ's in all. So there will be at least one Γ associated in this way with two or more γ's.

But this is self-contradictory. For, if Γ is associated with γ_1 in this way, it will be a member of γ_1 and not of γ_2; whilst if Γ is associated with γ_2 in this way, it will be a member of γ_2 and not of γ_1. Thus the supposition that every γ contains some Γ which is not contained in any of the other γ's leads to a contradiction, and must be rejected. Therefore it is always possible to find *some* selection of γ's from the original $n + 1$ γ's, such that it includes all the factors that are included in some other γ. Consequently *some* generalization of the form $g_{\mu\nu}$ *must* be true. It is obvious that this argument applies *a fortiori* when $N = n + p$, where $p > 1$.

* I have to thank Mr. A. E. Ingham, Fellow of Trinity, for kindly supplying me with the proof which follows.

Now the total number of possible generalizations of the form $g_{N-1,1}$ is N, and if *any* generalizations be true, some of *these* must be true. Thus, the antecedent probability of a generalization of this kind cannot be less than $\frac{1}{N}$. In practice, however, we can be fairly certain that the subject and predicate of our generalization do not together exhaust the total number of generated characteristics, so we have no right to assign so high an antecedent probability to any generalization that we shall actually meet with. I do not see any way of assigning a numerical value to the antecedent probability of a given generalization, even if we know that $N > n$. In order to do so we should have to evaluate the probability $(\mu_r \nu_s)/h$ in equation (10). This is the antecedent probability that the subject-characteristics $C_1 \ldots\ldots C_\mu$ between them require r factors for their generation, and that the predicate factors $C_{\mu+1} \ldots\ldots C_{\mu+\nu}$ between them require s factors for their generation. This could not be evaluated unless we made assumptions either about the antecedent probability that a generated characteristic, chosen at random, shall be of such and such an order, or about the antecedent probability that a generating set, chosen at random, should be of such and such fertility. (These probabilities could not, of course, be independent. Any assumption about the one would obviously affect the other.) I do not see any reasonable principle on which such antecedent probabilities could be assigned. It certainly does not seem reasonable to hold that any of the N generated characteristics is *equally* likely to be of the 1st, 2nd, or nth order; and it certainly does not seem reasonable to hold that any one of the $2^n - 1$ generating sets is *equally* likely to have any degree of fertility from 0 to N inclusive. Common sense would suggest that very high-order and very low-order characteristics would be rare, and that very fertile and very infertile generating sets would be rare. The Principle of Indifference would, I think, allow us to suppose that the antecedent probability of a given fertility would be the same for all generating sets *of the same order;* but it would certainly forbid us to assume that it was the same for generating sets of different orders.

This is a most unsatisfactory result. It is of very little interest to know that the antecedent probability of a generalization is *finite*, for this means only that it is greater than 0. What we want to know is that it is not less than *a certain assignable magnitude*, which would presumably be a function of μ, ν, N, n, and the antecedent probabilities mentioned above. Possibly someone with greater technical ability than I may be able to carry the argument to a more satisfactory conclusion, now that the nature of the problem has been made, as I hope it has, quite clear.

It is important to notice exactly what is the force of the condition that $N > n$. This condition is not needed to prove that any proposed generalization has a finite antecedent probability, as can be seen from equation (10), which makes no use of this condition. The condition $N > n$ simply assures us that *some* generalization *must* be true. This, of course, implies that *any* generalization has a finite antecedent probability. But it can obviously be the case that every generalization has a finite antecedent probability, even though it is not *certain* that any of them is true. The theory of generators, without the assumption that $N > n$, assures us that the antecedent probability of any generalization is finite, in the sense that it is greater than 0. The assumption that $N > n$ assures us that it is finite, in the sense that it is greater than a certain number which is itself greater than 0. The trouble is that we cannot evaluate this number without making assumptions about antecedent probability for which there seems to be very little justification.

(6.32) *Strengthening and Weakening Conditions for a given Generalization.*

Let us now consider what circumstances would tend to strengthen the antecedent probability of a generalization of the from $g_{\mu\nu}$. Let us first consider the subject. It is evidently desirable (a) that the subject-characteristics between them should require as many generating factors as possible. For this will increase the probability that the generating factors required by the predicate are contained among those required by the subject. Mere increase in μ, keeping ν fixed, *may* not secure this, for the added characteristics may between them require no generating factors besides those already required to generate the original μ subject characteristics. Still, an increase in μ does increase the probability that the subject requires a large number of generating factors, and does therefore increase the probability of the generalization. If we assume μ and ν to be fixed, then it is evident that the generalization will have the best chance of being true if (i) the subject contains characteristics of a high order, and (ii) the generating sets of the subject characteristics overlap as little as may be. (b) It is desirable that the generators of the subject characteristics shall be as fertile as possible, for this will increase the probability that the predicate characteristics are contained among those which are generated by the generators of the subject characteristics.

The conditions which the predicate should fulfil are complementary. It is desirable (a) that it shall require as few generating factors as possible, for this will increase the probability that all the generating factors required by the predicate are contained among those which are required by the subject. Mere decrease of ν, keeping μ fixed, *may* not secure this, for the characteristics which remain may require for their generation all the factors required by the original ν. Still, a decrease in ν does increase the probability that the predicate requires only a small number of generating factors, and does therefore increase the probability of the generalization. If we assume μ and ν to be fixed, then it is evident that the generalization will have the best chance of being true if (i) the predicate contains no characteristics of a high order, and (ii) the generating sets of the predicate characteristics overlap as much as possible. (*b*) The fertility of the generators of the predicate characteristics does not seem to be relevant.

(6.321) *Tests for the Fulfilment of these conditions.*

We now see what conditions tend to strengthen the antecedent probability of a generalization. Is there any way of testing whether they are probably fulfilled in a given case? I think there is.

(*a*) If a characteristic be of high order, it requires the presence of a large number of generating factors in any instance in which it occurs. Suppose, *e.g.*, that C is generated by the set $(\Gamma_1 \ldots \Gamma_r)$. Then in anything that has C there will be all the $2^r - 1$ generating sets that can be formed out of these r factors Of course, most of these may be sterile, and none of them need be very fertile. Still, the larger r is the more likely it will be that these $2^r - 1$ sets, which are present whenever C is present, generate between them a good many characteristics. All these characteristics will be present whenever C is present. On the other hand, any selection of them which does not include one of those generated by the complete set $(\Gamma_1 \ldots \Gamma_r)$ can be present without C being present. If, then, we find among the subject characteristics a certain one C, such that, whenever C is present, a certain large group of other characteristics is present, whilst many selections from this group can be present in the absence of C, there is a presumption that C is a characteristic of a fairly high order.

(*b*) If a characteristic C is generated by a highly prolific generating set, we shall find that there is a certain large group of characteristics, such that, whenever C is present, they are all present, and, whenever any of them is present, C and all the others are present. By making supplementary experiments and observations on these lines, we could presumably determine with fairly high probability whether the conditions required for a given generalization to have a high antecedent probability were fulfilled or not; and we see that the question whether μ is large as compared with ν in this generalization will be of relatively small importance in comparison with the other conditions. The relative magnitudes of μ and ν will merely be the test that we shall have to fall back upon if the other tests are inapplicable or lead to no definite results.

(6.4) *Admission of Plurality of Generators.*

Among the assumptions in (6.2) was included the denial of a plurality of generating sets for the same generated characteristic. Let us now consider in outline the result of relaxing this condition. We are to admit now that a generated characteristic C may be generated in some cases by one generating set, *e.g.*, $(\Gamma_1\ \Gamma_2)$; in other cases by another set, *e.g.*, $(\Gamma_2\ \Gamma_3)$; and in other cases by another set, *e.g.*, $(\Gamma_4\ \Gamma_5\ \Gamma_6)$. It is evident that there are three possible kinds of plurality of generators to be considered. The various generating sets which generate C may be either (*a*) all of the same order, *e.g.*, (Γ_1) and (Γ_2); or (*b*) all of different orders, *e.g.*, (Γ_1) and $(\Gamma_2\ \Gamma_3)$; or (*c*) a mixture, *e.g.*, (Γ_1), (Γ_2), and $(\Gamma_3\ \Gamma_4)$. These three kinds of plurality may be described as *Uniordinal*, *Multiordinal*, and *Mixed Plurality*, respectively. It follows from our definition of generation that a pair of generating sets such as (Γ_1) and $(\Gamma_1\ \Gamma_2)$ cannot both be generators of a single characteristic C, for to say that $(\Gamma_1\ \Gamma_2)$ generates C implies that Γ_1 does not do so and that Γ_2 does not do so. We can, of course, no longer speak of *the* set which generates a given characteristic; nor can we speak of *the* order of a characteristic, unless we happen to know that the only plurality of generators possible for this particular characteristic is uniordinal.

There is another very important distinction to be drawn in connexion with plurality of generators. We start, as before, with an observed object, having a certain N generated characteristic $C_1 \ldots\ldots C_N$. We suppose, as before, that each of them is generated by a set selected from a certain n generating factors $\Gamma_1 \ldots \Gamma_n$ which this thing possesses, and that none of these factors is wholly sterile and superfluous. In this particular thing, at this particular time, of course, each characteristic C will be generated by one and only one generating set. But we are now admitting that, in other things, or in this thing at other times, the characteristic C may be generated by other generating sets. Now two cases arise. Are these other sets to be simply *other selections* from the *same* n generating factors $\Gamma_1 \ldots \Gamma_n$? Or are we to admit that other things may have other sets of generating factors, *e.g.*, $\Gamma'_1 \ldots \Gamma'_m$, and that, in them, C may be generated by one or more generating sets selected out of these m different generating factors?

This question is very closely connected with the second difficulty about induction as applied to nature, viz., the fact that it is certain that all the instances that we have observed have fallen within a certain limited region of space and time, whilst we profess to argue to cases which not merely *did* not come, but *could not* have come under our observation. Mr. Keynes's Theory of Generators is not directly addressed to this difficulty, but to the question of something analogous to "loading" in nature. But, if we want to avoid the present difficulty, we shall have to assume that the only kind of plurality possible in nature is of the first kind and not of the second. We must assume that, in every thing, at every time and place, the characteristics $C_1 \ldots C_N$ are generated by the characteristics $\Gamma_1 \ldots \Gamma_n$, and that the only plurality is a *Selective Plurality*, *i.e.*, consists in the fact that the same characteristic C may be generated in one thing by one set selected from $\Gamma_1 \ldots \Gamma_m$, and in another thing by another set selected from $\Gamma_1 \ldots \Gamma_n$.

(6.41) *Application to Nomic Generalizations.*

Let us suppose, to simplify the argument, that we need only consider uniordinal plurality of generators. Suppose that one characteristic C_1, in the subject of a generalization, had a plurality of generators, *e.g.*, suppose that (Γ_1) and (Γ_2) are both generators of C_1. Now it might happen that (Γ_1) also generates a certain characteristic $C_{\mu+1}$ in the predicate of the generalization, whilst (Γ_2) does not. If, now, we were to take another thing with the same subject properties, it is quite possible that in it C_1 should be generated by (Γ_2) whilst Γ_1 was absent altogether. Then if (Γ_1) be the only possible generator of $C_{\mu+1}$, the generalization would necessarily break down for this second thing. This shows the effect of admitting plurality of generators as regards the subject of the generalization.

Let us next consider its effect as regards the predicate of the generalization. So far as I can see, it is *never* a *dis*advantage for the predicate properties to have a plurality of alternative generators, whilst it is *sometimes* a positive advantage. Suppose, *e.g.*, that the generalization was "Everything that had C_1 would have $C_{\mu+1}$." (*a*) Provided that C_1 has only one possible generator, everything that has C_1 must have this generator; and, provided that this generator does generate $C_{\mu+1}$, it cannot matter in the least how many more generating sets are also capable of generating $C_{\mu+1}$. (*b*) Suppose that C_1 is generated both by (Γ_1) and by (Γ_2). If $C_{\mu+1}$ has only one possible generator, *e.g.*, (Γ_1), the generalization will be wrecked, as we saw, by an object in which C_1 is generated by (Γ_2), and in which Γ_1 is not present. But suppose that either (Γ_1) or (Γ_2) will generate $C_{\mu+1}$, then $C_{\mu+1}$ will occur in *any* object that has C_1, and the generalization will be saved.

Thus the correct statement about the effect of a plurality of generating complexes for a single characteristic would seem to be as follows: (*a*) It is *always* unfavourable to the antecedent probability if the *subject*-characteristics have a plurality of alternative generators; and it is *never* unfavourable if the *predicate* characteristics have a plurality of alternative generators. (*b*) If any of the subject-characteristics have a plurality of alternative generators it is favourable to the generalization for the predicate-characteristics also to have a plurality of alternative generators.

If a generalization is to have a finite antecedent probability in the only important sense, *i.e.*, if its probability is to exceed a certain assignable number which is itself greater than 0, the following condition would seem to be necessary. There must be a probability greater than a certain number which is itself greater than 0, *either* (*a*) that none of the subject-characteristics have a plurality of alternative generators, *or* (*b*) that, if some of them do, the predicate-characteristics have at least as great a plurality of alternative generators. For, in the latter case, there will be a probability, which is finite in the non-trivial sense, that any generator which generates the subject-characteristics is also a generator for the predicate-characteristics

(6.42) *Application to Eduction.*

It is plain that the admission of a plurality of alternative generators for a given generated characteristic weakens nomic generalizations, since additional assumptions are now needed to guarantee that the antecedent probability of the generalization is finite. This is not so if we confine ourselves to eductive conclusions. Suppose the eductive conclusion is that the next thing that we meet which has C_1 will also have C_2. Suppose that C_2 has only one generator, *e.g.*, (Γ_2), whilst C_1 has several alternative generators, *e.g.*, (Γ_1) and (Γ_2). Then, so long as there is a finite antecedent probability that the next thing which has C_1 will have (Γ_2), there is a finite probability that it will have C_2. But, since the number of generators is only $2^n - 1$, the number of alternative possible generators for C_1 cannot exceed $2^n - 1$. Hence this condition is automatically fulfilled without making any fresh assumption. As a matter of fact the condition laid down above is needlessly sweeping. It would not matter how many alternative generators C_1 had, provided that the number of alternative generators of C_2 bore a finite *ratio* to it, *i.e.*, a ratio greater than a certain number which was greater than 0.

It is important to notice that, even if we confined our efforts in induction to establishing *eduction*, and gave up all attempts to establish *generalization* inductively, we should still be *presupposing* the existence of universal laws. For the justification of eduction involves the assumption of generators, and the connexion between a generating set and the characteristics which it generates is a universal law. We should thus be in the odd position that the existence of universal laws is presupposed by all induction, though no inductive argument can assign any finite probability to any law connecting observable characteristics.

(6.5) *The Elimination of Generators.*

It is evident that Mr. Keynes thinks of the generated characteristics as qualities like colour, hardness, noise, etc., which we can observe, and that he thinks of the generating factors as their hypothetical physical causes. It seems clear to me that it must be possible to eliminate the hypothetical generating factors, and to state the case wholly in terms of observable characteristics and their relations. I will give a very slight sketch of how this could be done, on the assumption that there is no plurality of alternative generators for a given generated characteristic.

If the theory of generators be true, and the above assumption be made, all the N characteristics which we are concerned with in inductive arguments must fall, as we sa win (6.2), into $2^n - 1$ mutually exclusive classes, such as $\alpha_1, \alpha_{12}, \ldots \alpha_{12 \ldots n}$. Some of these may contain no members. Now, if the theory of generators be true, each of these classes must form what I will call a *Coherent Set*. A *Coherent Set* may be defined as follows: To say that α is a *coherent set* means that it is a set of characteristics such that no member of it can ever occur without all the rest. Any particular coherent set can be defined by means of any characteristic C that falls within it. Thus they coherent set α^c may be defined as follows: It is the set of characteristics consisting of C itself and of every other characteristic which is always present when C is present, and absent when C is absent. Thus $\alpha^c = \hat{X}[X = C \,.\, \mathrm{v}: x \varepsilon X \equiv_x x \varepsilon C]$.

Now we can evidently drop the notion of generators altogether, and take the notion of mutually exclusive coherent sets of observable characteristics as fundamental. The fundamental assumption will now be that each of the N characteristics falls into some one member of a set of mutually exclusive coherent sets, whose total number is not greater than $2^n - 1$, where $n < \mathrm{N}$. It is obvious that every relation between these sets which could be *deduced* from the hypothesis of generators can be stated, without this hypothesis, as part of the original assumption. This I will illustrate very briefly.

(1) A set β is *subordinate* to a set α if the presence of any characteristic from α is always accompanied by that of some (and therefore of all) of the characteristics in β, whilst the converse

does not hold. This is obviously the kind of relation that holds between α_1 and α_{12}, or between α_1 and α_{123}.

(2) A set β is *immediately subordinate* to a set α if β is subordinate to α, and there is no set γ such that γ is subordinate to α and β is subordinate to γ. This is the kind of relation that holds between α_1 and α_{12}.

(3) A set of sets β, γ, δ form an *exhaustive set of subordinates* to α if each is subordinate to α, and whenever a characteristic from each of the sets β, γ, δ is present, one (and therefore all) of the characteristics in α are present. This is the kind of relation that holds between $(\alpha_1, \alpha_{12}, \alpha_3)$ and α_{123}; or between $(\alpha_1, \alpha_2, \alpha_3)$ and α_{123}.

(4) Three sets, α, β, γ, may be so related that none is subordinate to either of the others, but that the presence of a characteristic belonging to any two of them is always accompanied by the presence of one (and therefore of all) the characteristics belonging to the third. This is the kind of relation that holds between α_{12}, α_{23}, and α_{31}.

It is clear that all that is necessary in the assumptions which Mr. Keynes makes in terms of generators and their relations to each other, on the one hand, and to the characteristics which they generate, on the other, could be stated in terms of coherent sets and their relations of subordination, etc., to each other. Since the existence of generators implies the existence of coherent sets having these relations to each other, whilst coherent sets might exist and have these relations to each other even if there were no generators, it is obviously advantageous from a purely logical point of view to state the conditions in terms of coherent sets, and to avoid the assumption of generators. From the practical point of view of expounding the theory and drawing remote consequences from it, it is desirable to continue to employ the notion of generators; but they can now be regarded as no more than convenient parameters. They *may* exist, but it is not necessary to suppose that they *do*.

If we allow plurality of alternate generators for a given characteristic the elimination of generators will be a more complicated business; but it is clear that it must be capable of being carried through even in this case.

(6.6) *The Establishment of Functional Laws.*

There is one other point which it is important to mention before leaving the Ontological Question. I have assumed that the Γ's and the C's are *determinable* characteristics. Mr. Keynes does not explicitly say that they are, but he evidently must intend the Γ's, at any rate, to be determinables and not determinates, for he assumes that the total number of Γ's is finite, and hopes that it may be comparatively small. This assumption would be absurd if the Γ's were supposed to be determinate characteristics, for even a single determinable generating factor might have a large, and even a transfinite, number of different determinate values. But, if the Γ's are supposed to be determinables, so too must the C's be, for it is evident that a set, no matter how complex, of merely determin*able* Γ's would not generate a determin*ate* C. Two consequences follow: (1) It is not nearly so plausible as Mr. Keynes seems to think that the number of C's should largely exceed the number of Γ's. The number of different determin*able* characteristics that we can observe is by no means large; the qualitative variety which we observe in nature is due to the fact that each of the comparatively few observable determinables, such as colour, temperature, etc., has an enormous and perhaps transfinite number of determinates under it. (2) The assumptions that have been made justify only generalizations of the crudest kind, viz., assertions of the form that, whenever certain determinables are present, certain other determinables will be present. Now only the most backward sciences are content with such generalizations. What we want are *Functional Laws*, *i.e.*, laws which will enable us to predict the determinate values of the predicate-characteristics for any given determinate values of the subject-characteristics. To establish such laws, further assumptions have to be made, and something analogous to the *Method of Concomitant Variations* must be used. These assumptions are stated (whether with complete fullness or accuracy I do not here enquire) in Mr. Johnson's treatment of *Demonstrative Induction* in his *Logic*. The only point that I will mention here is that there now arises the possibility of yet another kind of plurality, which Mr. Johnson rules out, quite unjustifiably in my opinion. This is the possibility that *the same* determinate value of the predicate-characteristics may be determined by *several* different determinate values of the subject-characteristics, *i.e.*, that the functional laws of nature may not all be one-valued functions of the variables.

(7) The Epistemic Question.

We have now seen what conditions must be fulfilled in nature if inductive arguments are ever to be able to establish reasonably high probabilities. What evidence, if any, have we for supposing that these conditions are in fact fulfilled ? Let us call the conditions laid down in (6.2) *The Principle of Limited Variety*, and let us denote it by l. What we have shown is that, if g be any generalisation, $g/l \nless \varepsilon$, where ε is a certain number which is greater than 0, but which we have not been able to evaluate.

Now I do not think that anyone would maintain that the *Principle of Limited Variety* has the slightest trace of self-evidence, or that it can be deduced from anything else which is self-evident. Hence it must be admitted that we do not *know* that l is true. So the next question is : Has l a finite probability with respect to anything that we do know to be true ?

Suppose there were certain known facts, f, relative to which l had a finite probability. Suppose further that, if l were true, certain empirical consequences, e, would follow, and that e is found to be true. Now

$$l/fe = \frac{(l/f)\,(e/lf)}{e/f} \text{ by the } \textit{Conjunctive Principle.}$$

But, by hypothesis $e/lf = 1$. Hence $l/fe = \dfrac{(l/f)}{(e/f)}$. Since e/f cannot be greater than 1, $l/fe \nless l/f$; and, if $e/f < 1$, $l/fe > l/f$, which is itself supposed to be greater than 0. So, if these conditions were fulfilled, l/fe would be greater than a certain magnitude which is itself greater than 0. The next question then is: Can we find a set of facts, f, and a set of facts, e, such that $l/f > 0$, $e/lf = 1$, and $e/f < 1$?

It seems to me that there is at least one fact which gives l a faint probability by analogy. We do know that we can actually construct out of simple parts of the same nature complicated structures which behave in very different ways, *e.g.*, watches, motor-cars, gramophones, etc. The differences in observable behaviour are here known to be due simply to differences in arrangement of materials having the same properties; and these materials, and the structures formed of them, are parts of the material world. Relatively to this fact it does seem to me that there is a finite probability that the variety of *material* nature at any rate, should arise in the same way. Hence, if f denotes this fact about artificial machines, I should say that $l/f > 0$.

Next, it is certain that there is a great deal of recurrence and repetition in nature; and that, up to the present, the more we have looked for it the more we have found it, even when at first sight there seemed little trace of it. I have dealt with this point in detail in my second article on *Induction and Probability* in *Mind*, Vol. XXIX, 1920. Now, if the Principle of Limited

Variety were true, there would be recurrence and repetition in nature; whilst if it were not, there is very little reason to expect that there would be. Hence, if e be this empirical fact, it seems evident that $e/lf = 1$ and $e/f < 1$. Consequently, there are facts f and e which fulfil the required conditions, and therefore there are facts f and e such that l/fe is greater than a certain number which is greater than 0.

Finally, we have to apply this result to the question of the antecedent probability of any proposed generalization g.

By the *Rule of Expansion* $g \,.\equiv:gl\,.\,\mathrm{v}\,.\,g\bar{l}$

By the *Principle of Equivalence* $g/fe = (gl)/fe + (g\bar{l})/fe$

By the *Conjunctive Principle* $= (l/fe)\,(g/lfe) + (\bar{l}/fe)\,(g/\bar{l}fe)$

$$\text{Whence } g/fe > (l/fe)\,(g/lfe).$$

Now g/lfe is certainly not less than g/l, for the addition of the two facts f and e certainly does not reduce the probability that g shall be true, given that the *Principle of Limited Variety* is true. Hence $g/fe > (l/fe)\,(g/l)$.

But g/l is greater than a certain number ε, which is itself greater than 0; and l/fe is greater than a certain number γ (viz., l/f) which is itself greater than 0. Hence $g/fe > \eta\varepsilon$, which is greater than 0.

We see then that any generalization about the material world has a finite initial probability, relative to the known facts that we can construct a variety of differently acting machines from similar materials and that there is a great deal of repetition and regularity in the material world; and this initial probability will increase as we find more regularity and repetition.

Thus a more or less satisfactory answer can be made to the *Epistemic Question*, so long as we confine ourselves to inductive arguments about the material world. But, so far as I can see, we have no ground whatever to trust inductive generalizations about *mental* phenomena; for here there are no known facts analogous to f, the fact that we can construct machines of the same materials to act in different ways.

(8) Summary of Conclusions.

Every inductive argument presupposes, beside the general principles of formal logic and of probability, certain assumptions about equiprobability, and what I have called in (3) the *Fundamental Causal Premise.* If it is to establish a high probability, it requires in addition the assumption that "loading" in favour of a certain one alternative is antecedently highly probable. In the case of induction applied to things and events in nature, these conditions will not be fulfilled unless nature has a certain particular kind of structure, which may be expressed by saying that it answers to the *Principle of Limited Variety.* We stated this principle in terms of the notion of generating factors, and deduced its consequences, first on the assumption that plurality of generators is excluded, and then on the assumption that it is admitted. We also stated the conditions which tend to strengthen or weaken the antecedent probability of a generalization on the assumption that nature is subject to the *Principle of Limited Variety*, and we gave certain tests for judging whether a given generalization does or does not fulfil these conditions. Then we showed that the notion of generators, though highly convenient, is not essential to the statement of the *Principle of Limited Variety.* The actual existence of generators may be left an open question, and the fundamental notion may be taken to be that of coherent sets of characteristics related to each other in certain ways. We pointed out that, even on the assumption of the *Principle of Limited Variety,* only crude generalizations connecting determinables can be established by induction. To establish functional laws further assumptions about nature are needed. Finally, we said that the *Principle of Limited Variety* is neither intuitively nor demonstratively certain. But there are two known facts about the material world which are so related to it that the antecedent probability of any proposed generalization about material phenomena with respect to these two facts is greater than a certain number which is greater than 0. Lastly we saw that the same argument does not apply to inductive generalizations about mental phenomena. So that, with our present knowledge, we have no good reason to attach any great weight to the conclusions of inductive argument on these subjects.