VOL. LIII. No. 211.] [July, 1944.

MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—HR. VON WRIGHT ON THE LOGIC OF INDUCTION (III.).

BY C. D. BROAD.

THEOREMS CONNECTING PROBABILITY WITH INDUCTION (*Continued*).

Theorem 4. *The Universal Generalisation Theorem.*—We have seen that it is necessary, in the case of infinite classes, to draw a distinction between the Extreme Statistical Generalisation, ' The proportion of R's in an indefinitely extended sequence of Q's is 100 % ', and the Universal Generalisation, ' All the Q's in a certain indefinitely extended sequence will be R's '. In our notation the former proposition is symbolised by $f_\infty(R\,;\ Q) = 1$. The latter might be symbolised by U(R ; Q).

We want to evaluate the probability

$$U(R\,;\,Q)/:k\,.\,\&\,.f_N(R\,;\,Q) = 1,$$

and to see what happens to it when N is indefinitely increased. For shortness I shall write U for U(R ; Q), and, as before, $P_N(1)$ for $f_N(R\,;\,Q) = 1$.

There are two theorems to be proved, which we will call (4·1) and (4·2). The first states that, under certain conditions, $U/k\ \&\ P_N(1)$ increases with every increase of N. The second states that, under certain conditions, $U/k\ \&\ P_N(1)$ approaches to 1 as its limit if N increased indefinitely.

(4·1) By using the Lemmas it is very easy to show, as Keynes does in his *Treatise on Probability*, that

$$U/k\ \&\ P_{N+1}(1) = .\,\frac{U/k\ \&\ P_N(1)}{R(x_{N+1})/k\ \&\ P_N(1)}.$$

Hence it is plain that U/k & $P_{N+1}(1)$ is greater than U/k & $P_N(1)$ provided that (i) U/k & $P_N(1)$ is not zero, and (ii) $R(x_{N+1})/k\&P_N(1)$ is less than 1.

The following is the meaning of the second condition. It must not be the case that the fact that the first N trials of Q's have all been R's makes it *certain* that the next Q tried will be an R. This condition can be granted without hesitation. Keynes, in a very obscure argument which uses the Identity of Indiscernibles as one of its premisses, draws from this condition the conclusion that an increase in the number of instances favourable to a universal generalisation strengthens the latter *only* in so far as it increases the negative analogy. Hr. von Wright attempts to state this argument of Keynes's clearly and to refute it. I have never been able to see any force in the argument myself, and I do not propose to linger over the refutation of it.

(4·2) It is also easy to show, as Keynes does in his *Treatise*, by means of the Lemmas that

$$U/k \,\&\, P_N(1) = \frac{U/k}{U/k + \overline{U}/k \times P_N(1)/k \,\&\, \overline{U}}.$$

The necessary and sufficient conditions for this expression to approach 1 as its limit when N is indefinitely increased are the following.

(i) U/k is not zero.

(ii) As N tends to infinity $P_N(1)/k \,\&\, \overline{U}$ tends to 0.

Hr. von Wright has no difficulty in showing that the second of these conditions cannot be granted. The supposal that U is false would be satisfied if even a single instance of Q turned out not to be R. But, as we have seen, the occurrence of any finite number of such counter-instances in an indefinitely long series of trials would not suffice to reduce to 0 the probability that the proportion of Q's which are R tends to the limit of 100 % as the sequence of trials is indefinitely prolonged.

The point can be made perfectly plain by the following examples of drawing counters from a bag. We have to compare the following two cases. (i) The bag used for the experiment may contain a large or a small number of counters, but in either case there is literally *no* non-white counter among them. (ii) The bag contains an enormously great number of counters and among them are a very few non-whites. In the first case the universal generalisation, 'All drawings will be white', is true from the nature of the bag. In the second case there is always a possibility at each drawing that the counter drawn will not be white, no matter how small the probability may be. The probability that there will be *at least one* non-white drawing in the first N drawings is in fact $1 - \left(1 - \frac{m}{n}\right)^{N}$, if n be the number of counters in the bag and m be the number of these which are not white. In this case, then, the universal generalisation U is false. Now the probability that all the first N drawings will be white is in this case $\left(1 - \frac{m}{n}\right)^{N}$. Now, if $\frac{m}{n}$ is finite, this does tend to 0 as N tends to infinity. But suppose that n, the number of counters in the bag, tends to infinity as well as N, the number of drawings, whilst m is finite. Then this probability does not necessarily tend to 0 as N tends to infinity, for it assumes the indeterminate form 1^{∞}. Suppose we put $N = kn$ and allow n to increase without limit. Then the expression $\left(1 - \frac{m}{n}\right)^{N}$ becomes e^{-km}, *i.e.*, $e^{-\frac{mN}{n}}$. Now this might have any value between 0 and 1 according to whether the ratio between the two infinitely great quantities N and n was great or small.

Theorem 5.—*Laplace's Rule of Succession.*—As usual, this theorem divides into two. They may be called respectively (5·1) *The Non-numerical Rule*, and (5·2) *The Numerical Rule*, of Succession. I have dealt with them fairly fully in my Presidential Address to the *Aristotelian Society*, which will be found by anyone whom it may interest to look for it in Vol. XXVIII of their *Proceedings*.

(5·1) *The Non-numerical Rule.*—If N trials of Q's have been made under Bernoullian conditions and all of them have been R, then the probability that the $(N + 1)$th Q to be tried will be an R approaches the limiting value 1 as N is indefinitely increased. The formal statement in our notation is

$$(\varepsilon) :: . (\exists \nu) :: N > \nu \supset_{N} : . \mathrm{R}(x_{N+1}) / : f_{N}(\mathrm{R} ; \mathrm{Q}) = 1 \,.\, \& \,.\, k : \geqslant 1 - \varepsilon.$$

The condition under which this proposition holds is that $\mathrm{R}(x)/\mathrm{Q}(x) \,\&\, h = 1 \,.\, /k$ shall be greater than 0.

The following example of drawing counters from a bag will make the general line of reasoning plain. The longer the series of drawings, if all of them turn out to be white, the more likely it is that the bag from which the drawings are made contains either nothing but white counters or at any rate very nearly 100 % of white counters. But on the first alternative the next drawing *must* be white, and, on the second, it is very highly probable that it will be white.

The formal argument is as follows. We divide the interval between 0 and 1 into the usual set of a very large number μ of very short adjoined sub-intervals, each of length η; and we denote the proposition that the value of $R(x)/Q(x)$ & h lies between $r\eta$ and $(r+1)\eta$ by P_r. We denote the proposition $f_N(R\,;Q) = 1$ as usual by $P_N(1)$. Then, by Lemma V,

$$R(x_{N+1})/P_N(1) \,\&\, k \,.=.\, \overset{r=\mu-1}{\underset{r=0}{S}} P_r/P_N(1) \,\&\, k \times R(x_{N+1})/P_r \,\&\, P_N(1) \,\&\, k. \quad (1)$$

Now, for reasons which were stated at length in the proof of the *Statistical Principle of Greatest Probability* in Part II of this paper, the factor $P_N(1)$ is irrelevant, in conjunction with P_r, in the supposal of $R(x_{N+1})/P_r$ & $P_N(1)$ & k. Such terms as this can therefore be written in the simpler form $R(x_{N+1})/P_r$ & k.

But, by the *Inverse Principle of Great Numbers*, when N tends to infinity such terms as $P_r/P_N(1)$ & k tend to 1 for $r = \mu - 1$ and to 0 for all other values of r, provided only that $P_{\mu-1}/k$ is not zero. Therefore, provided that $P_{\mu-1}/k$ is not zero, the right-hand side of the above equation reduces to $R(x_{N+1})/P_{\mu-1}$ & k as N tends to infinity. But $P_{\mu-1}$ is the proposition that $R(x)/Q(x)$ & h lies in the very small interval between $1 - \eta$ and 1. Therefore as N tends to infinity $R(x_{N+1})/P_{\mu-1}$ & k approaches indefinitely nearly to 1. Therefore the left-hand side of the above equation approaches indefinitely nearly to 1 as N is increased indefinitely, provided only that the probability with respect to k that $R(x)/Q(x)$ & $h = 1$ is not zero. And this is what we had to prove.

(5·2) *The Numerical Rule.*—If N trials of Q's have been made under Bernoullian conditions and all of them have been R, then the probability that the $(N+1)$th Q to be tried will be an R is $\frac{N+1}{N+2}$, provided that all the possible values from 0 to 1 of $R(x)/Q(x)$ & h are equally likely with respect to k.

What we have to prove, then, expressed in our abbreviated notation, is that

$$R(x_{N+1})/P_N(1) \,\&\, k = \frac{N+1}{N+2}$$

provided that P_r/k has the same value for all values of r.

The proof is as follows. If we apply Lemma VII (the *Bayes Principle*) to the factors $P_r/P_N(1)$ & k in the expression on the right-hand side of Equation (1) above, and if we remember that the term $P_N(1)$ can be suppressed in the supposal of the factors

$R(x_{N+1})/P_r$ & $P_N(1)$ & k, we find that Equation (1) can be transformed into

$$R(x_{N+1})/P_N(1) \& k = \frac{\overset{r=\mu-1}{\underset{r=0}{S}} P_r/k \times P_N(1)/P_r \& k \times R(x_{N+1})/P_r \& k}{\overset{r=\mu-1}{\underset{r=0}{S}} P_r/k \times P_N(1)/P_r \& k}. \quad (2)$$

We can evaluate all the terms in this except P_r/k. And, on the supposition that P_r/k has the same value for all values of r, the P_r/k terms in the numerator and the denominator cancel each other out. The terms to be evaluated are $R(x_{N+1})/P_r$ & k and $P_N(1)/P_r$ & k. The first is $\frac{r}{\mu}$; and, since the conditions of the experiment are assumed to be Bernoullian, the second is $\left(\frac{r}{\mu}\right)^N$. Therefore the expression on the right of Equation (2) reduces to

$$\frac{\overset{r=\mu-1}{\underset{r=0}{S}} r^{N+1}}{\mu \overset{r=\mu-1}{\underset{r=0}{S}} r^N}.$$

Now μ is enormously great, for it is the number of very short sub-intervals of length η into which we divided the interval from 0 to 1. And it is easy to prove by elementary algebra that the fraction to which the right-hand side of Equation (2) has just been reduced becomes equal to $\frac{N+1}{N+2}$ as μ is indefinitely increased. So this is the value of $R(x_{N+1})/P_N(1)$ & k, provided that all values of $R(x)/Q(x)$ & k are equally likely with respect to k. Q.E.D.

Theorem 6.—The rate at which the probability of a universal generalisation increases with each additional confirmatory instance diminishes as the number of confirmatory instances is increased.

This is an immediate consequence of the equation at the beginning of the proof of Theorem 4·1 and the Non-numerical Rule of Succession. For the former tells us that the ratio of U/k & $P_{N+1}(1)$ to U/k & $P_N(1)$ is *inversely proportional* to $R(x_{N+1})/P_N(1)$ & k. And the latter tells us that $R(x_{N+1})/P_N(1)$ & k increases to the limiting value 1 as N is indefinitely increased.

Theorem 7.—This is a theorem connecting the relative 'generality' of two universal propositions with their relative antecedent probabilities. It divides into two, but I shall show that the two are logically equivalent.

We can compare two universal propositions in respect of 'generality' if either (i) both have the same predicate, but one has a more restricted subject than the other; or (ii) both have the same subject, but one has a more restricted predicate than the other. An example of the first case would be the two propositions 'All men are liars' and 'All black men are liars'. We can describe the latter as 'super-determinate in respect of its subject' to the former. An example of the second case would be the two propositions 'All men are fools' and 'All men are fools and knaves'. We can describe the latter as 'super-determinate in respect of its predicate' to the former.

Now it is easy to prove the following general proposition: 'If p implies q, then, whatever h may be, q/h is greater than p/h unless either (i) $p/h = 0$ or (ii) $p/q \,\&\, h = 1$'. The proof is as follows.

If $p \supset q$ then $p \equiv p \,\&\, q$.

Therefore, by Lemma III and Postulate (v),

$$q/h = \frac{p/h}{p/q \,\&\, h}.$$

Now, by Postulate (ii) $p/q \,\&\, h$ cannot be greater than 1. Therefore, unless it is equal to 1, it must be less than 1. Therefore, unless $p/h = 0$, the expression on the right-hand side of the equation must be greater than p/h. Therefore q/h is greater than p/h unless $p/h = 0$ or $p/q \,\&\, h = 1$. Q.E.D.

It is of some interest to consider what happens if the condition that p/h is not equal to 0 breaks down. An immediate consequence of Lemma V and Postulate (ii) is that if $p/h = 0$ then either $p/q \,\&\, h = 0$ or $q/h = 0$, whatever q may be. So, if $p/h = 0$, either $q/h = 0$ also or q/h assumes the indeterminate form $\frac{0}{0}$.

Theorems (7·1) and (7·2) are immediate consequences of the general proposition which we have just proved. They may be stated as follows.

(7·1) If p and q are any two universal propositions with the same *subject*, and the predicate of p is super-determinate to that of q, then the probability of p is *less than* that of q with respect to any datum h, unless $p/h = 0$ or $p/q \,\&\, h = 1$.

(7·2) If p and q are any two universal propositions with the same *predicate*, and the subject of p is super-determinate to that of q, then the probability of p is *greater than* that of q with respect to any datum h, unless $q/h = 0$ or $q \,/\, p \,\&\, h = 1$.

The proof is obvious. In the first case (*e.g.*, where p is of the form 'All S is P' and q is of the form 'All S is P or Q') p entails q. In the second case (*e.g.*, where p is of the form 'All S which is P is Q' and q is of the form 'All S is Q') q entails p. The two theorems then follow at once from the general proposition proved above.

Both are in accordance with common-sense. The proposition with the less determinate predicate runs less risk of refutation because of the comparative *vagueness* of what it asserts. The proposition with the more determinate subject runs less risk of refutation because of the comparative *narrowness* of the field within which it asserts the predicate.

It is perhaps worth while to remark that, so far from there being any conflict between the two criteria, they logically entail each other. For suppose we start with the pair 'All S is P' and 'All S is P or Q', in which there is a common *subject* and where the predicate of the first is super-determinate to that of the second. Each of these propositions is logically equivalent to its contrapositive. Now their contrapositives are respectively 'All $\bar{\text{P}}$ is $\bar{\text{S}}$' and 'All $\bar{\text{P}} \,\&\, \bar{\text{Q}}$ is $\bar{\text{S}}$'. These have a common *predicate*, and the subject of the second is super-determinate to that of the first.

Theorem 8. *Curve-fitting.*—I find Hr. von Wright's treatment of this subject very unsatisfactory. In the first place, I think it is vitiated by an elementary mathematical oversight, which I will explain. Secondly, even when this is avoided, as we shall see that it can be, the rest of the argument is to me (and to others far more competent than myself whom I have consulted) extremely obscure. I shall therefore have to construct an argument of my own, which is suggested by Hr. von Wright's obscure statements and leads to the same kind of conclusion as his, but is certainly not to be found in his book. Possibly it is what he has in mind.

First for the mathematical 'howler', as it appears to me to be. The essence of the matter is as follows. Hr. von Wright supposes that we have n pairs of correlated values of two variables, x and y, given by observation, and that we are trying to find a curve which will fit them all exactly. He explicitly confines his attention to curves of the form

$$y = \text{A}_0 x^m + \text{A}_1 x^{m-1} + \ldots \text{A}_m.$$

Now it is of the essence of his argument that there might be *two* curves of this form, *viz.*, polynomials in x, fitting the *same n* points exactly, for *both* of which m is less than n. But it is easy to show that this supposition is logically impossible.

For suppose that all the n points were on a curve of this form, where m was less than n. Consider any other curve of the same form but of higher order $m + p$, where $m + p$ is also less than n. This would cut the former curve in *only* $m + p$ points, *viz.*, those which are given by the roots of the equation

$$A_0x^{m+p} + \ldots A_{p-1}x^{m+1} + (A_p - B_0)x^m + (A_{p+1} - B_1)x^{m-1} + \ldots (A_{p+m} - B_m) \mp 0.$$

Hence only $m + p$ of the n points which are on the curve of the mth order could be also on the curve of the $(m + p)$th order. And, by hypothesis, $m + p$ is less than n.

This objection would not hold if Hr. von Wright were comparing *two* sets of n points, one from one experiment and the other from another, and were supposing that a polynomial of the mth order fitted the former set whilst one of the $(m + p)$th order fitted the latter. But this is not what he says or what his argument presupposes. What he says, and what his argument presupposes, is what I have just shown to be logically impossible.

In order to continue the discussion let us, however, suppose henceforth that we are considering *two* sets of n observations, one from one experiment and the other from another. I propose to substitute for Hr. von Wright's very obscure argument the following reasoning, which is quite clear and, I believe, valid.

Suppose that two experiments are done, each on a different natural phenomenon. In each of them n pairs of correlated values of two variables, x and y, are observed. Suppose that in the first case all the observed values fall on a certain polynomial of order m; and that in the second they all fall on a certain polynomial of order $m + p$, where both m and $m + p$ are less than n. Denote these two propositions respectively by $O(m, n)$ and $O(m + p, n)$. Let $L(m)$ be the proposition 'The law of the phenomena in the first experiment is the polynomial of order m which fits the n observations made in that experiment'. Let $L(m + p)$ have a similar meaning, *mutatis mutandis*, for the second experiment. Let h be any relevant information that we have antecedent to $O(m, n)$ and $O(m + p, n)$. We wish to compare the two probabilities

$$L(m)/O(m, n) \,\&\, h \quad \text{and} \quad L(m + p)/O(m + p, n) \,\&\, h.$$

Consider the former of these. If we bear in mind the fact that $O(m, n)/L(m) \,\&\, h = 1$, by Postulate (iii), since $L(m)$ implies

$O(m, n)$, we can prove at once from Lemma VII (the *Bayes Principle*) that

$$L(m)/O(m, n) \,\&\, h = \frac{L(m)/h}{L(m)/h + \bar{L}(m)/h \times O(m, n)/\bar{L}(m) \,\&\, h}.$$

Now consider the factor $O(m, n)/\bar{L}(m) \,\&\, h$ in the denominator of this. We have used up m out of our n observed values in determining the coefficients in the polynomial. So we are left with $n - m$ which might or might not fall on this curve. Now antecedently it might be that 0 or 1 or . . . $n - m$ of these would fall on the polynomial determined by the remaining m. The possibility that r of them do so covers as many possibilities as there are different ways of choosing r things out of $n - m$, *i.e.*, ${}^{n-m}C_r$. So the total number of such possibilities is $\overset{r=n-m}{\underset{r=0}{S}}\ {}^{n-m}C_r$, *i.e.*, 2^{n-m}. Of these $O(m, n)$ is a single one. So, if all of them are equally probable on the supposition that the law $L(m)$ is *false*, we have $O(m, n)/\bar{L}(m) \,\&\, h = \frac{1}{2^{n-m}}$. Therefore

$$L(m)/O(m, n) \,\&\, h = \frac{L(m)/h}{L(m)/h + [1 - L(m)/h]\frac{1}{2^{n-m}}}.$$

By precisely similar reasoning we can show that

$$L(m+p)/O(m+p, n) \,\&\, h = \frac{L(m+p)/h}{L(m+p)/h + [1 - L(m+p)/h]\frac{1}{2^{n-m-p}}}.$$

Now suppose that the antecedent probability of $L(m)$ and $L(m + p)$ is the same, *i.e.*, that $L(m)/h = L(m + p)/h =$ (say) α. Put $\frac{1 - \alpha}{\alpha} = \lambda$. Then

$$L(m)/O(m, n) \,\&\, h = \frac{1}{1 + \frac{\lambda}{2^{n-m}}}$$

and
$$L(m + p)/O(m + p, n) \,\&\, h = \frac{1}{1 + \frac{\lambda}{2^{n-m-p}}},$$

i.e., the probability that the polynomial of lower order which fits the n observations of the first experiment is the law of the phenomena examined in that experiment is greater than the probability that the polynomial of higher order which fits the *n* observations of the second experiment is the law of the phenomena examined in *that* experiment. (It should be observed that this conclusion has been reached only subject to two assumptions about equi-probability.)

There is one other remark which I will make before leaving this topic. Hr. von Wright discusses the question on the hypothesis that the empirically determined points fall *exactly* on this, that, or the other suggested polynomial curve. It seems to me that this case is hardly worth considering. In real life it is a question of comparing the degree of 'goodness of fit' of a number of alternative curves, none of which exactly fit all the observed points. Any adequate treatment of this problem would involve discussing the Method of Least Squares, about which there is an enormous literature.

This concludes what I have to say about the Formal Analysis of Inductive Probability. It is evident that we are left with two problems. One is the interpretation to be put on probability propositions. The other is the evidence, if such there be, for the truth of the conditions under which the various theorems have been deduced. It seems plain that the former question should be considered before the latter.

(2) *Interpretations of Probability Propositions.*—Hr. von Wright distinguishes the following main interpretations of the formal postulates of the calculus of probability. (i) *The Frequency Interpretation ;* (ii) *The 'Spielraum' Interpretation ;* and (iii) *The Interpretation of Probability as an Indefinable Special Notion.* He subdivides the 'Spielraum' Interpretation first into two forms, which he calls 'logical' and 'empirical'. I intend, for reasons which will appear in due course, to describe them respectively as (ii, *a*) *Purely Quotitative,* and (ii, *b*) *Partly Quantitative.* Lastly, he distinguishes two sub-species of the purely quotitative form of the 'Spielraum' interpretation. I shall call these (ii, *a*, α) the *Intrinsic,* and (ii, *a*, β) the *Extrinsic* forms of the theory.

There is one general remark to be made before we explain these various interpretations in detail. We must remember that an 'interpretation' of a set of postulates, in the technical sense, means no more than a set of entities of any kind which, when substituted for the x's and y's and R's of the postulates, turn the latter into true propositions. They need not be in the least what anyone has in mind when he uses the words in which the axioms are stated. Thus, *e.g.*, a perfectly satisfactory 'interpretation' of the postulates of Euclidean geometry arises if we substitute for the word 'point', wherever it occurs, an ordered triad of any three numbers (x, y, z); and for the phrase 'distance between two points' the square root of the sum of the squares of the differences between the corresponding numbers in two such triads. But no one in his senses would suggest that the former is what a person has in his mind in ordinary life when he uses the word 'point' or that the latter is what he has in mind when he uses the word 'distance'. It is important to face this fact and recognise it where it is obvious, as in the case of geometry, before entering the much more obscure region of probability where it might not be noticed.

(i) *Frequency Interpretation.*—On this interpretation the notion of probability applies strictly and primarily only to *propositional functions* and not to *propositions*. To say that the probability of a thing or event being an instance of R, given that it is an instance of Q, is p means, on this interpretation, that $f_N(R ; Q)$ approaches a limit as N is indefinitely increased, and that this limiting value is p. Some writers, *e.g.*, von Mises, introduce a further condition, *viz.*, that the distribution of instances of R among the instances of Q must be in a certain sense 'random'. Hr. von Wright has dealt with this latter contention in his article on *Probability* in MIND, Vol. XLIX, No. 195, and the reader may be referred to it and to my review of von Mises in MIND, Vol. XLVI, for a discussion of this subject. It is not of importance in relation to the question whether induction can be justified in terms of probability.

It is known that the probability postulates are necessary propositions if probability is interpreted to mean limiting frequency. A proof will be found in Reichenbach's *Wahrscheinlichkeitslehre.* There is just one remark that seems to me to be worth making on this point. So far as I can see, Postulates (v) and (vi), *i.e.*, the Conjunctive and the Disjunctive Postulates, are in a different position on the frequency interpretation. Postulate (v) becomes a purely *algebraic* triviality, depending simply on the identity that $\frac{a}{b} = \frac{a}{c} \times \frac{c}{b}$ where a, b, and c are any *numbers* that you please. But Postulate (vi) depends on a certain necessary proposition about the number of terms in a *disjunctive class, viz.*, that

$$Nc`(\alpha \vee \beta) = Nc`\alpha + Nc`\beta - Nc`(\alpha \,\&\, \beta),$$

where α and β are any *classes* that you please. If I am right in saying this, it seems to cast doubt on whether the frequency interpretation expresses what we ordinarily mean by probability.

For we certainly do not regard these two postulates as being fundamentally different in character.

It is commonly held that the frequency interpretation cannot plausibly be regarded as expressing what we have in mind when we ascribe probabilities (i) to singular propositions, such as 'Mr. Jones, who has just been taken ill with influenza, will recover', and (ii) to general laws or theories, like the Newtonian Theory of Gravitation. In the former it would be claimed by opponents that statistical information about the proportion of recoveries in various classes of cases to which Mr. Jones's case belongs are *evidence for* ascribing such and such a probability to the proposition about Mr. Jones, but are not the whole of the meaning of such an ascription. In the latter it would be said that no plausible interpretation in terms of frequency has ever been suggested.

Hr. von Wright, who is inclined to accept the frequency interpretation as adequate, deals with the question of the probability of universal generalisations in the following way. Consider the generalisation, 'All swans are white'. This is a particular instance of a wider generalisation, 'All birds of the same species have the same kind of pigmentation'. In assigning a meaning to the statement, 'The probability that all swans are white is p', we have to proceed as follows. Instead of considering *individuals of a certain species* (*e.g.*, this, that, and the other swan) and asking what proportion of them have a *certain colour* (*e.g.* white), we have to consider *species of a certain genus* (*e.g.*, the species swan, the species crow, and other species of bird) and ask what proportion of them have *sameness of pigmentation for all members* (*e.g.*, whiteness for all swans, blackness for all crows, etc.). Suppose that the limiting proportion in the latter case is p. Then this is what we mean by saying of any generalisation of the form 'All members of such and such a species of birds (*e.g.*, swans) have such and such a kind of pigmentation (*e.g.*, uniform whiteness)' that its probability is p.

Hr. von Wright realises that the conditions here laid down can never, strictly speaking, be fulfilled. As regards an individual, *e.g.*, a swan, one can decide with certainty by looking at it whether it does or does not have a certain colour all over, *e.g.*, whiteness. But we cannot examine all members, past, present, and future, of *any* species, and therefore cannot know that it has a certain property common to all its members. To this he answers that we may know, with regard to the N species of a genus, instances of which have been examined, that a proportion $1 - p$ *lacked* internal uniformity in respect of a certain generic quality (*e.g.*, colour). And we may know that the remaining pN of them had such uniformity *for all their members examined up to date.* On this kind of basis, which is all that from the nature of the case we could ever have, the most reasonable estimate of the probability of any such generalisation as 'All swans are white' is p. And the *meaning* of such a statement as 'The probability that all swans are white is p' remains unaltered by these further refinements.

I have no doubt whatever that this kind of reference to wider classes is often a most important part of the *evidence for* a universal generalisation. One would be inclined, *e.g.*, to feel less confidence in such a generalisation as 'All swans are white' than in such a one as 'All samples of phosphorus melt at 40° C.' on the ground that experience has shown that pigmentation is a more variable quality among birds of the same species than is melting-point among samples of the same chemical element. But it is not clear to me how the notion of *limiting* frequency, which is Hr. von Wright's proposed interpretation of probability, is to be applied to such a collection as the various species of the genus bird and to such a characteristic as 'sameness of pigmentation within a species'. This notion is clear enough when we have something analogous to the potentially unlimited sequence of drawings and replacements of a counter from a bag. But surely the analogy has worn very thin here.

(ii) *The 'Spielraum' Interpretation.*—Let us call any proposition which is explicitly constructed from one or more other propositions by the single or repeated application of some or all of the processes of negation, conjunction, or disjunction 'explicitly molecular'. A proposition which is not explicitly molecular will be called 'ostensibly atomic'. The ostensibly atomic propositions out of which an explicitly molecular proposition is constructed will be called its 'elements'. Each of the elements of an explicitly molecular proposition may be either true or false. So, if there are n of them, there will be 2^n alternative possible combinations of truth-values for them. Each such alternative will be called an 'elementary truth-value combination'. Any explicitly molecular proposition will be true for certain of the truth-value combinations of its elements and will be false for the rest of them. All forms of 'Spielraum' interpretation depend on the facts just described. In the *purely quotitative* variety of the theory the alternative possibilities are merely counted. They are not regarded as each having a certain kind of magnitude—degree of 'possibility' or of 'probability' —in respect of which they can be judged to be equal or unequal.

In the *partly quantitative* variety the alternatives are compared in respect of such a magnitude before being counted.

(ii, *a*, α) *Intrinsic Form of Purely Quotitative Variety.*—On this form of the theory the 'probability' of any proposition is defined as the ratio of the number of the truth-value combinations of its elements which *make it true* to the *total* number of such combinations. This makes the probability of every ostensibly atomic proposition to be $\frac{1}{2}$. For there are two possibilities, true or false ; and the possibility of its being true is one of them. Next, consider the conjunctive proposition $p \& q$. With two elements there are four elementary truth-value combinations. One only of these, *viz.*, p true and q true, makes $p \& q$ true. Therefore the probability of $p \& q$ is $\frac{1}{4}$. Lastly, consider the disjunctive proposition $p \vee q$. Of the four elementary truth-value combinations all but one, *viz.*, p false and q false, make $p \vee q$ true. Therefore the probability of $p \vee q$ is $\frac{3}{4}$. It is evident from this that the Disjunctive Postulate holds with this interpretation of probability. For, as we have just seen, the probability of $p \vee q$ is $\frac{3}{4}$; that of the two atomic propositions p and q is $\frac{1}{2}$ in each case ; and that of the conjunctive proposition $p \& q$ is $\frac{1}{4}$. And, by simple arithmetic, $\frac{3}{4} = \frac{1}{2} + \frac{1}{2} - \frac{1}{4}$. In a similar way it can be shown that all the other postulates hold.

I call this form of the quotitative variety of the theory 'intrinsic', because it does not make the probability of a proposition to be or to involve a relation to any other proposition.

(ii, *a*, β) *Extrinsic Form of Purely Quotitative Variety.*—Let P be any proposition and H be any other proposition. Either of them may be either ostensibly atomic or explicitly molecular. On the present theory the probability of P given H is defined as the ratio of the number of truth-value combinations of the elements of the conjunctive proposition $P \& H$ which make *that conjunction true* to the number of them which make H *true.* This definition will be made clearer by some examples. In the first place, the probability of any ostensibly atomic proposition p with respect to any other ostensibly atomic proposition h will be $\frac{1}{2}$. For there are four elementary truth-value combinations with the two propositions p and h ; and one of them makes $p \& h$ true, whilst two of them make h true. Again, consider $p \vee q/h$. With three elementary propositions there are eight elementary truth-value combinations. Three of these make the conjunction $p \,.\vee.\, q :\& h$ true, whilst four of them make h true. So the probability of $p \vee q/h$, on the present theory, is $\frac{3}{4}$, provided that p, q, and h are all ostensibly atomic propositions. It is easy to show that all the Postulates are satisfied with this interpretation of P/H.

I call this form of the quotitative variety of the theory 'extrinsic', because it makes the probability of a proposition to be something which is essentially relative to another proposition.

(ii, *b*) *Partly Quantitative Variety.*—The 'Spielraum' theory has very little plausibility as an account of what we commonly have in mind when we talk of 'probability' so long as it remains purely quotitative. To make it plausible we have to add the condition 'provided that all the alternative possibilities considered and counted are *equally* probable'. Hr. von Wright says that at this stage the theory does not differ from the classical Laplacean definition. If it does not, I should say that it is an extremely clumsy way of putting that definition. The normal and straightforward way of putting it would be as follows. According to it the statement '$p/h = \frac{m}{n}$' means that h is a disjunction of n mutually exclusive and equi-probable alternatives whilst p is logically equivalent to a disjunction of m of these. This would, of course, be circular if it were offered as a definition of *probability*. But it is not circular if it is content to take the notion of probability as undefined and merely claims to define the statement that the *degree* of probability of a proposition is *so-and-so*. With this restriction it seems to me to answer exactly to our practice in working out problems in the Theory of Probability. But the proviso that the alternatives shall be equi-probable is absolutely essential, and this notion remains undefined. Moreover, the definition does not suggest any criterion by which we are to judge in any concrete application whether two alternatives are or are not equiprobable. At this point the theory has either to rely on individual intuition, or to appeal to some general principle such as that of Symmetry or that of Insufficient Reason, or to base its judgments of equi-probability on the relative frequency with which the various alternative possibilities have been realised. The first two alternatives are unsatisfactory, whilst the third brings us back to the Frequency Theory and to the problem of Statistical Generalisation and its justification.

(iii) *Probability as an Indefinable Notion.*—Hr. von Wright mentions this alternative and points out that it cannot provide an 'interpretation' of the Postulates in the technical sense, as the Frequency and the 'Spielraum' theories certainly do. That is to say, the Postulates do not become necessary propositions of arithmetic or of the logic of classes, as they do if probability is interpreted either in accordance with the Frequency

Theory or the several varieties of the 'Spielraum' Theory. He does not pursue this alternative further; and, in particular, he does not consider the contention that the probability relation is analogous to but 'weaker than', the relation between the premiss and the conclusion of a valid deductive inference.

(3) *Restatement of the Laws of Great Numbers in Terms of the Frequency Theory.*—Before going further I think it will be very well worth while to do something which Hr. von Wright does not attempt, *viz.*, to restate the Direct and the Inverse Laws of Great Numbers in terms of the Frequency Theory of Probability. For the sake of simplicity and concreteness I shall take the particular examples of drawing counters from bags rather than the general propositions of which these are instances.

(3·1) *Direct Law of Great Numbers.*—The essential point to notice is that we have to consider two different sequences, *viz.*, (i) a sequence of *drawings*, and (ii) a sequence of *sets of drawings*. In the first we are concerned with the limiting frequency of *white drawings*, and in the second with the limiting frequency of *sets containing a certain proportion of white drawings*. This being understood, the Direct Law may be stated as follows:

Suppose that the proportion of *white drawings* in a sequence of N *drawings* made under Bernoullian conditions from a certain bag of counters approaches to the limit p as N is increased indefinitely. Then, however small δ may be, the proportion of *sets of n drawings containing between* $n(p - \delta)$ *and* $n(p + \delta)$ *white drawings* in a sequence of N' *sets of n drawings* from that bag will approach to the limit 1 as N' is increased indefinitely.

(3·2) *Inverse Law of Great Numbers.*—Here there are two additional points to bear in mind. (i) This law is concerned with the probability of a probability. On the Frequency Theory this must be the limiting frequency with which a certain limiting frequency occurs. (ii) The condition under which the Law holds is that it shall not be infinitely improbable that the value of a certain probability shall lie in the immediate neighbourhood of a certain fraction. This condition will have to be expressed in terms of limiting frequencies. With these preliminaries the Law may be stated as follows.

Suppose that from each of N' bags of unknown constitution a sequence of N drawings has been made under Bernoullian conditions, and that in each of these sequences the proportion of white drawings has been p. Suppose that, in such circumstances as those under which the experiment was done, bags containing a proportion p of white counters are not infinitely rare. Then, however small δ and ϵ may be, if N' and N be sufficiently great the proportion of these N' bags in which the proportion of white counters lay between $p - \delta$ and $p + \delta$ cannot have differed from 1 by more than ϵ.

I am afraid that this is a very complicated statement; but I think that it is clear, and I do not see how to express the Inverse Law of Great Numbers in terms of the Frequency Theory in any simpler way.

(4) *The Alleged 'Bridge' between Probability and Frequency.*— It has often been suggested that, even if probability cannot be *defined* in terms of frequency, yet the Direct Law of Great Numbers provides a 'bridge' from the former to the latter. If this means that the Law enables one to start from a premiss which asserts probability and to infer a conclusion which categorically asserts that such and such a frequency will be or has been realised, it is a complete mistake. A moment's inspection of the accurate formulation of the Direct Law will show this. The premiss is that the probability of any *single* event of a certain generic kind turning out in a certain specific way is so-and-so. The conclusion is that the probability of such and such a *proportion of events* of that kind turning out in that way approaches to such and such a limit as the length of the sequence is indefinitely prolonged. Thus the inference is from probability to probability, not from probability to frequency. This is in no way altered by the fact that the frequency which is the subject of the probability-statement in the conclusion is the same fraction as the probability which is asserted of the individual event in the premiss. Nor is it altered by the fact that the probability which is mentioned in the conclusion is there asserted to differ by as little as we please from 1 provided that the sequence is sufficiently prolonged.

Again, the Inverse Law of Great Numbers does not provide a 'bridge' from frequency to *first-order* probability, but only to the probability of another probability having a certain numerical value. The premiss here is that, in a sequence of events of a certain generic kind, a certain proportion have turned out in a certain specific way. The conclusion is that it is to such and such a degree probable that the probability of any single event of this kind turning out in this way was so-and-so. The argument, then, is from a frequency to a *second-order* probability. This is in no way altered by the fact that the first-order probability which forms the subject of the conclusion is there shown to be more likely to have the same numerical value as the frequency which is asserted in the premiss than to have any other value.

Nor is it altered by the fact that the probability of its having that numerical value is there asserted to differ by as little as we please from 1 provided that the sequence mentioned in the premiss was long enough.

Hr. von Wright mentions certain linguistic confusions which he thinks may have caused intelligent persons to make the mistakes which we have been pointing out. Whatever the cause may be, the best cure is simply to state the two Laws of Great Numbers with meticulous accuracy, and then to translate them with equal care in terms of whatever interpretation of 'probability' one may like to adopt. I have tried to do this in the previous Section for the Frequency interpretation.

(5) *Can Inductive Generalisation be justified by the Principles of Probability?*—If this can be done at all, it must be done by means of the Theorems which we have proved above; and all the rest of these depend upon the Direct and the Inverse Laws of Great Numbers. Now we have just seen that, whatever interpretation we may put on 'probability', these laws will at best enable us to infer only that under certain conditions certain statistical or universal generalisations will have a probability which approaches in the limit to 1. So we can confine ourselves to the contention that, in favourable circumstances, inductive generalisation can be justified with high probability. But this means nothing in particular until some interpretation has been put upon the notion of probability. For the present purpose we may divide the possible interpretations into (i) the Frequency Interpretation, and (ii) Non-frequency Interpretations.

On the Frequency Interpretation it is certain that the more probable it is that an event of a given kind will turn out in a certain way the greater is the relative frequency with which events of that kind will turn out in that way if the sequence of such events be sufficiently prolonged. This is certain simply because, on this interpretation of probability and on this alone, it is analytic. On the other hand, the Direct Law of Great Numbers starts with a premiss about the probability of events of a certain kind turning out in a certain way. Now, on the Frequency Interpretation, this is itself a statement about the relative frequency with which events of that kind will turn out in that way as the sequence of such events is indefinitely prolonged. Either this is merely assumed as a hypothesis, or it is taken as a categorical premiss. On the first alternative anything that is inferred from it is shown only to be a consequence of the hypothesis; it cannot be asserted by itself as a categorical conclusion. On the second alternative we are at once faced with the question: What is your evidence for the proposition about limiting frequency which is asserted in your premiss? Plainly no proposition of this kind can be a mere report of what has been perceived, as, *e.g.*, the proposition, 'This swan is white' or 'All the swans in this pond are white', might be. It is in fact evident that the premiss is itself a statistical generalisation from the observed frequency of certain kinds of events in certain limited sequences. To put it briefly. On the frequency interpretation of probability the Theorems will enable you to pass from premisses about limiting frequencies in certain sequences to conclusions about limiting frequencies in certain other sequences related to the former in certain specified ways. Such inferences are of great interest and importance. But, since the premisses are themselves either assumed as hypotheses or established by inductive generalisation, these Theorems cannot supply a justification for inductive generalisation in general.

Suppose, then, that we put some *non-frequency* interpretation on probability. Then the mere fact that the probability of an event of a certain kind turning out in a certain way is p is no *guarantee* that, even in the long run, events of that kind will turn out in that way with the relative frequency p. No doubt, if the run is long enough, it is overwhelmingly probable (in whatever sense 'probable' is being used) that the proportion of such events which turn out in this way will be approximately p. But (except on the frequency interpretation, which we are now excluding) this does not entail that in a very long series of very long equal runs of such events an overwhelming proportion of the runs *would* contain a proportion p of events which had turned out in this way. This too can, no doubt, be shown to be overwhelmingly probable (in whatever sense 'probable' is being used); and so on without end. But at no stage shall we be able to pass from a certain frequency being overwhelmingly probable, in the non-frequency sense, to its being overwhelmingly frequent.

The upshot of the matter is this. The fact that one alternative is more likely to be fulfilled than another is a good reason for acting on the assumption that the former will be realised only if on the whole and in the long run the more probable alternative is the one that is more often realised. This condition is guaranteed analytically on the frequency interpretation, and is not guaranteed at all on any other interpretation, of probability. On the other hand, if we interpret probability in terms of limiting frequency, we must abandon all hope of justifying inductive generalisation by means of the Laws of Great Numbers. For, on that interpretation, these Laws merely enable you to pass from premisses about limiting frequencies in certain sequences to conclusions about limiting frequencies in certain other sequences connected in certain specific ways with the former. The problem: What justification is there for asserting that the limiting frequency for any particular kind of series is so-and-so? falls outside these Theorems, just as the problem of guaranteeing the premisses of a syllogism falls outside the Theory of the Syllogism.

(6) *Induction as a Self-Correcting Process.*—The last topic with which Hr. von Wright deals is Reichenbach's contention that induction is a self-correcting process. I am not at all sure that I understand either Reichenbach's statements in his *Wahrscheinlichkeitslehre* or Hr. von Wright's synopsis of them. I propose therefore to try by means of an example to state what I imagine to be the point. If I am wrong, wiser heads will be able to correct me.

Suppose that we are given a coin, not known to be fair, and that we want to make an estimate of the probability of throwing a head with it.

(i) We throw it n times, and we find that we get m heads. At this stage we estimate the probability of throwing a head with this coin as $\frac{m}{n}$.

(ii) We now make a sequence of n' *sets*, each of n throws of the coin. Let the number of these sets which contain 0, 1, . . . n heads respectively be $n'_0, n'_1, \ldots n'_n$. Then we should have

$$\overset{r=n}{\underset{r=0}{\mathrm{S}}}\, n'_r = n'.$$

Now suppose that the probability of throwing a head were p. Then we can calculate what would be the most probable number of n-fold sets of throws containing exactly r heads in a sequence of n' such sets on this hypothesis. Call this ν'_r. The actual value of ν_r' would be ${}^n\mathrm{C}_r p^r(1 - p)^{n-r}n'$.

I now introduce something which is entirely conjectural, for I find no explicit mention of it in either Reichenbach or Hr. von Wright. This is the well-known function χ^2 as a measure of goodness of fit.

Consider the expression

$$\chi^2 = \overset{r=n}{\underset{r=0}{\mathrm{S}}} \frac{(n'_r - \nu'_r)^2}{\nu'_r}.$$

This measures the goodness of fit of the hypothesis that the probability of throwing a head is p to the actual distribution of heads among the sets in the sequence. The fit is closest when this expression is as small as possible. Since ν_r' is a function of p, χ^2 will also be a function of p. We want to find that value of p which will make χ^2 a minimum. Call this p_0. If p_0 does not differ from $\frac{m}{n}$ we have no reason to revise our original estimate of the probability of throwing a head with this coin. But, if p_0 does differ appreciably from $\frac{m}{n}$, it is reasonable to substitute it for $\frac{m}{n}$ as our estimate of this probability. For this is the hypothesis which best fits the more detailed facts now at our disposal, *viz.*, the actual *distribution* of the various possible numbers of heads in a sequence of n' sets of n throws.

(iii) We now pass to the next stage. We make a sequence of n'' sequences each consisting of n' sets of n throws. Let the number of these sequences which consist of r_0 sets containing 0 heads, r_1 sets containing 1 head, . . . and r_n sets containing n heads be denoted by $n''_{r_0 r_1 \ldots r_n}$. The r's are of course subject to the condition that

$$\overset{k=n}{\underset{k=0}{\mathrm{S}}}\, r_k = n'.$$

Now suppose that the probability of throwing a head were p. Then we could calculate what would be the most probable number of sequences of n' sets of n throws consisting of r_0 sets with 0 heads, r_1 sets with 1 head, . . . and r_n sets with n heads. Call this $\nu''_{r_0 r_1 \ldots r_n}$.

As before we calculate χ^2 for all the possible values of the r's, and we proceed to find that value of p which makes the new χ^2 a minimum. Call this p_0'. If it does not differ from p_0, the estimate reached at the previous stage, there is no reason to revise the latter. If it does differ from p_0, it is reasonable to substitute p_0' for p_0 as our estimate of the probability of throwing a head with this coin.

(iv) The principles of the procedure are now plain, and it could be pursued to as many further stages as we like.

Supposing, for the sake of argument, that what Reichenbach had in mind is something like what I have been trying to describe, the advantages of the procedure would seem to be the following. It is true that the three stages which I have been describing might be taken to consist simply of a run of n throws, a run of

$n'.n$ throws, and a run of $n''.n'.n$ throws, and that we might simply have taken the ratio of the number of heads to the number of throws in each case as successive estimates of the probability of throwing a head with this coin. If we had done so, the three estimates would have been in ascending order of reliability because they are based on increasingly long sequences of throws. But by merely doing this we throw away much detailed information which is relevant to question at issue. We take no account of any information that may be available about the distribution of the various possible proportions of heads among successive equal sets of throws, or about the still more complex facts of distribution among successive equal sequences of successive equal sets of throws. It is intelligible that, by taking into account such information in some such way as I have suggested, we might reach in a shorter total sequence of throws as accurate an estimate as we could reach in a much longer sequence by cruder methods which ignore these details.

All this may be completely beside the mark; but even if it be, it is of some interest in itself, and so I give it for what it may be worth.

CONCLUSION.—As I hope that Hr. von Wright's book may have a wide circulation in the new New Jerusalem, when the world has been made still safer for democracy, I shall end with a list of the misprints or traces of imperfect English which I have noticed in it. The chief of these are as follows:—

P. 48, last line but one, for *phosporus* read *phosphorus*.

P. 80, line 3 of par. 5, for *it* read *if*.

P. 149, line 5 of par. 5, for *others* read *other*.

P. 233, last line of Note 19, for *Braitwaite* read *Braithwaite*.

These are the main misprints. The English needs to be amended in the following passages:—

P. 100, l. 4, for *depending upon what* read *according to what*.

P. 105, l. 6 of par. 3, for *inefficiencies* read *defects*.

P. 106, last line, for *inclusive* read *including*.

P. 108, l. 3 of par. 4, for *arithmetics and the analysis* read *arithmetic and analysis*.

P. 137, l. 1 and l. 25, for *indirectly* read *inversely*.

P. 152, penultimate line, for *in average* read *on the average*.

P. 174, l. 5, for *constance* read *constancy*.

P. 233, Sect. 6, Note 4, for *supersede* read *exceed*.