PHILOSOPHICAL FOUNDATIONS OF PROBABILITY

HANS REICHENBACH

UNIVERSITY OF CALIFORNIA, LOS ANGELES

I

In sections I–V we deal with the formal structure of probability; in sections VI–XI we investigate the meaning and the assertability of probability statements.[1]

The concept of probability refers to a relation. If we cast a die, the probability of a certain face is 1/6; the condition introduced by "if" is necessary for this instance of a probability relation as well as for all others. When the condition is omitted the statement must be regarded as elliptic; such omission is possible if it is obvious from the context what condition is understood. We therefore regard probability as having the logical form of an implication, which we call the *probability implication.*

This implication, however, holds not between individuals but between classes. Thus the phrase "cast the die" defines a class A of events, and similarly the phrase "face 6 turns up" defines a class B of events. The class A is called the *reference class,* the class B is named the *attribute class.* Furthermore, the events x_i and y_i belonging respectively to these classes are regarded as given in a certain order and in such a way that a one-to-one correspondence between the elements of the sequences is known, which we express by the use of the same subscript "i". Since the probability statement refers to all events belonging to the classes A and B, it can be written in the form of an all-statement:

$$(i)\ \ (x_i \epsilon A \underset{p}{\supset\!\!\!-} y_i \epsilon B) \tag{1}$$

The symbol "(i)" is the all-operator of logistics; the symbol "ϵ", as usual, denotes the relation of class membership. The real number p is the degree of probability.

Instead of the *implicational notation* presented in (1) it is convenient to introduce a *mathematical notation,* or *functor notation.* We write

$$P(A,B) = p \tag{2}$$

The symbol "$P(\ \ )$" is a functor, meaning "the probability of". Expression (2) has the same meaning as (1) and can be regarded as an abbreviation in which the reference to the sequences x_i and y_i is not expressed. We read (2) in the form "the probability from A to B is p".

[1] For a detailed account of the following ideas we refer the reader to the author's *Wahrscheinlichkeitslehre* (Leiden, 1935), A. W. Sijthoff. A summary in the French language was published under the title "Les fondements logiques du calcul des probabilités," *Annales de l'Institut Henri Poincaré,* t. VII, fasc. v (Paris, 1937), pp. 267–348. The general ideas of secs. VI–XI are presented in chap. v of the author's *Experience and Prediction* (Chicago, 1938).

For operations inside the P-symbol we apply the rules of logistics, using the signs " $\vee$ " for the inclusive "or", "." for the "and", and a line on top of the letter for the negation. Thus "$P(A,\bar{B})$" means the probability from A to non-B; "$P(A,B \vee C)$" means the probability from A to B or C. For every expression it is permissible to substitute tautologically equivalent expressions; this rule allows for manipulations of the kind used in logistics. The P-symbol as a whole is regarded as a mathematical variable used in equations like (2) or of a more general form.

Like all mathematical systems the calculus of the P-symbol is employed in two conceptions. In the *formal conception* we do not give any meaning to the P-symbol, but set up formal relations connecting various forms of expressions. In other words, we define the P-symbol implicitly by a set of axioms. In the *material conception* or *interpretation* we introduce a meaning for the P-symbol in terms of other mathematical or physical concepts. We then have to show that the *coördinative definition* so introduced satisfies the set of axioms. The latter condition restricts the class of admissible interpretations, but it does not single out one interpretation as the only admissible one.

The axioms which we introduce are the following ones:

α. Normalization:

1. $P(A,A \vee B) = 1$
2. $P(A,B.\bar{B}) = 0$
3. $0 \leqq P(A,B)$

β. Addition:

$$P(A,B \vee C) = P(A,B) + P(A,C) - P(A,B.C)$$

γ. Multiplication:

$$P(A,B.C) = P(A,B) \cdot P(A.B,C)$$

The variables expressed by small letters are real numbers; a restriction to the values 0 and 1 limits included follows from axiom α, 3. To this axiom, however, we add the condition (not expressed in the symbolic notation) that it holds only in case the class A is not empty. The general restriction, therefore, is bound to the same condition.

We do not claim that between any two classes A and B a probability exists, but regard it as an empirical question whether there is such a probability. On the other hand, we set up the rule that a probability exists if it is numerically determined by given probabilities in terms of the axioms or derivable theorems (*rule of existence*). This rule permits us to solve a probability equation for any of its variables.

Axiom β is called the *general theorem of addition*. For exclusive events B and C the term $P(A,B.C)$ drops out, and the formula then becomes the *special* *theorem of addition.* Similarly we call axiom γ the *general theorem of multiplication.* For independent events we have

$$P(A.B,C) = P(A,C) \tag{3a}$$

The theorem then assumes the form

$$P(A,B.C) = P(A,B) \cdot P(A,C) \tag{3b}$$

which we call the *special theorem of multiplication.*

The frequency interpretation can be written in the form

$$P(A,B) = \lim_{n \to \infty} \frac{N^n(A.B)}{N^n(A)} \tag{4}$$

where the symbol "$N^n(X)$" means the number of those elements of the sequence up to the nth element which belong to the class X. It can be shown that if the frequency interpretation (4) is used the axioms follow tautologically. The frequency interpretation is therefore admissible. For the formal manipulations within the calculus of probability, however, the interpretation (4) is not necessary. All theorems of the calculus, for instance the rule of Bayes, can be formally derived from the axioms. The calculus defined by the axioms $\alpha - \gamma$ we call the *elementary calculus of probability.*

II

We have introduced probability as a property of sequences; if we use the frequency interpretation this property is identified with the limit of the frequency. The only condition a sequence must have in order to be a probability sequence, therefore, is that it has a probability, or limit of the frequency.

This definition leaves open the structure of probability sequences. We include among probability sequences the special forms that are usually called random sequences, but we do not hesitate likewise to include such strictly ordered forms as a sequence in which B and $\bar{B}$ alternate regularly. It cannot be the task of the mathematician to deny the name "probability sequence" to certain types of sequences; it should rather be regarded as his task to define various types of order and to derive the laws holding for them. This task is undertaken by *the theory of the order of probability sequences,* which represents the second chapter of the calculus of probability.

A definition of types of order is achieved by certain means of structural characterization. They are based on the study of the probability in subsequences resulting from the original sequence by means of selections. This means of structural characterization was introduced by v. Mises; we apply it, however, in a generalized form since we do not always require that the probabilities of the selected subsequences are equal to that of the main sequence. For every sequence there will exist a number of selections resulting in subsequences of the same probability as the main sequence; the class of selections satisfying this (and a further) condition is called the *domain of invariance*. A certain type of sequence will be characterized by rules which set up minimum requirements for the domain of invariance. This method of minimum requirements has the advantage that for the derivation of theorems we do not introduce more postulates than are necessary for the derivation.

We extend our symbolism by using *phase superscripts*. Assume, for instance, that in a sequence of events B and $\overline{B}$ we select all events preceded by a B. This probability will be written in the implicational notation

$$(i)\ (x_i \epsilon A \,.\, y_i \epsilon B \underset{p}{\ni} y_{i+1} \epsilon B) \tag{5}$$

In the functor notation we use phase superscripts, and write instead of (5)

$$P(A.B,B^1) \tag{6}$$

The condition that a selection by the predecessor leads to the same probability as the main sequence is then expressed by the equation

$$P(A.B,B^1) = P(A,B) \tag{7}$$

By a sequence *free from aftereffect* we understand the condition

$$P(A,B_{i_1}{}^1 \dots B_{i_{\nu-1}}{}^{\nu-1},B_{i_\nu}{}^{\nu}) = P(A,B_{i_\nu}) \qquad (\nu = 1, 2, 3, \dots) \tag{8}$$

Another form of selection is given by the arithmetical progressions. We say that an element x_i belongs to the selection $S_{\lambda\kappa}$ when

$$i = \kappa + (m - 1) \cdot \lambda \qquad \begin{matrix} m = 1, 2, 3, \dots \\ \kappa = 1, 2, \dots \lambda \end{matrix} \tag{9}$$

A selection $S_{\lambda\kappa}$ is called a *regular division*.

The *domain of invariance* is defined as the class of all selections that leave unchanged not only the probability of the major sequence but also all phase probabilities. We define *normal sequences* by the condition that the regular divisions and the selections by predecessors belong to the domain of invariance. The first condition is expressed by the following equations:

$$P(A.S_{\lambda\kappa},B_i) = P(A,B_i) \tag{10}$$

$$P(A.S_{\lambda\kappa}{}^{a}.B_{i_1}{}^1 \dots B_{i_{\nu-1}}{}^{\nu-1},B_{i_\nu}{}^{\nu}) = P(A.B_{i_1}{}^1 \dots B_{i_{\nu-1}}{}^{\nu-1},B_{i_\nu}{}^{\nu})$$

$$\kappa = 1, 2, \dots, \lambda;\ a = 1, 2, \dots, \lambda;\ \nu = 1, 2, \dots, \lambda;\ \lambda = 1, 2, 3, \dots)$$

The second condition is given by (8); it is identical with the condition that the sequence be free from aftereffect. It can be shown that for normal sequences so defined the special theorem of multiplication holds for any succession of consecutive elements and that they therefore satisfy the Bernoulli theorem. The normal sequences of our definition are identical with the admissible numbers introduced by A. Copeland.[2] The postulates introduced by R. v. Mises for his collectives are much stronger than our postulates; every collective is a normal sequence, but not vice versa.

A randomness which is not restricted to an arbitrary class of selections, and which makes it impossible to select deviating subsequences by arithmetical methods that do not refer to the attribute, may be called a *logical randomness*. The plan of defining a logical randomness was set forth by R. v. Mises[3] and has been continued by A. Wald[4] and A. Church.[5] The results of Church, in particular, are interesting for the logician. But it should not be forgotten that the random sequences represent only a special case of probability sequences and that therefore the logical problems of probability are independent of the definition of randomness.

Among the sequences with aftereffect we meet with an interesting type in which the aftereffect depends only on the immediate predecessor. Such sequences were studied by Markoff. We speak here of transfer of probability. This condition is expressed by the following equations:

$$P(A.B_{i_1}{}^{1} \ldots B_{i_{\nu-1}}{}^{\nu-1}, B_{i_\nu}{}^{\nu}) = P(A.B_{i_{\nu-1}}{}^{\nu-1}, B_{i_\nu}{}^{\nu}) \quad (11)$$

The types of order so far considered refer to individual sequences. In many cases, however, we are concerned with sequences of sequences, i.e., with a *sequence lattice*. The probabilities of horizontal and vertical sequences must here be studied individually; they are logically independent of each other. If a dependence—for instance, equality of all limits—is assumed, it must be introduced by a special postulate defining a certain type of sequence lattice. An important type of this kind is given by *normal sequences in the narrower sense*. Another type, which combines properties of probability transfer with certain properties of normal sequences, is important for the analysis of such physical phenomena as occur in the mixture of liquids.

The definition of types of order is not limited; it will be advisable, however, to construct the definitions in such a way that they lend themselves to practical applications.

III

The probability sequences so far considered possess only a finite number of attributes, in the simplest case only B and $\overline{B}$. A more general form is given by sequences possessing an infinite number of attributes. Translating a term introduced by v. Mises, we speak here of an *attribute space*. Such sequences may be called *primitive probability sequences*. A division of the attribute space into areas will produce probability sequences of a finite number of attributes.

[2] A. H. Copeland, "Admissible numbers in the theory of probability," *Amer. Jour. Math.*, vol. 50, no. 4 (1928), p. 535.

[3] R. v. Mises, *Wahrscheinlichkeitsrechnung* (Leipzig, 1931), and earlier publications.

[4] A. Wald, *Die Widerspruchsfreiheit des Kollektivbegriffs, Ergebnisse eines mathematischen Kolloquiums*, Heft 8 (Wien, 1937).

[5] Alonzo Church, *Bull. Amer. Math. Soc.*, vol. 46 (1940), p. 130.

That probability can be defined as a ratio of areas in the attribute space is made possible by the fact that the formal system of axioms admits of a geometrical interpretation. For many mathematical purposes it is convenient to use only the geometrical interpretation; and some theories of probability therefore read like chapters of set theory. It should not be forgotten, however, that the probability concept of applications is always the frequency concept, not the geometrical concept. That it is permissible to apply the results of geometrical probabilities to practical statistics derives from the isomorphism between the two interpretations.

For the treatment of the continuous attribute space a condition of complete additivity is required which states that the probability of the limit of an infinite set of classes is equal to the limit of the probabilities of these classes.

IV

The conception of probability developed in the preceding considerations may be called the *mathematical conception*. In it probability is conceived as a property of sequences of events or of other physical objects. We now turn to the *logical conception* of probability, which regards probability as a property of sentences, comparable to truth. The first to see this duality of interpretation was G. Boole.[6]

In order to make probability the analogue of truth we shall omit the general reference class A in the notation, regarding this class as understood. In the frequency interpretation it is given by the subscript of the symbols of the elements. We thus write "$P(B)$" instead of "$P(A,B)$".

The frequency interpretation can be transferred to the logical conception by the device of counting sentences about events instead of counting events. Since the number of true sentences of the form "$x_i \epsilon B$" corresponds to the number of events B, the two interpretations are isomorphous. The logical interpretation, however, offers certain advantages since it is required for the understanding of linguistic forms in which probability is used as the analogue of truth. Thus we say it is probable that it will rain tomorrow; or we speak of the probability of a certain hypothetical assumption. For the present we shall study only the logical structure of such forms, postponing the question of interpretation.

Two-valued logic is based on the truth tables (12). In addition to the symbols introduced above we employ here the signs "$\supset$" for implication and "$\equiv$" for equivalence.

TRUTH TABLES OF TWO-VALUED LOGIC

a	$\bar{a}$
T	F
F	T

a	b	$a \vee b$	$a.b$	$a \supset b$	$a \equiv b$
T	T	T	T	T	T
T	F	T	F	F	F
F	T	T	F	T	F
F	F	F	F	T	T

(12)

[6] G. Boole, *An Investigation of the Laws of Thought* (London, 1854), pp. 247–248.

The truth tables can be read in two directions. The first direction goes from right to left, i.e., from the compound proposition to the elementary propositions. Thus if "$a \vee b$" is true we know that "$a.b$" is true or that "$\bar{a}.b$" is true or that "$a.\bar{b}$" is true. The second direction goes from left to right, i.e., from the elementary propositions to the compound proposition. Thus if "$a.b$" is true we regard the statement "$a \supset b$" as verified. When we use only the first direction we follow a *connective interpretation*; when we use both directions we apply an *adjunctive interpretation*. In conversational language the implication is usually interpreted connectively; the adjunctive implication, therefore, is often regarded as "unreasonable." The "or" is used in both interpretations; an adjunctive "or" therefore appears "reasonable." For logical purposes the adjunctive operations appear preferable; the definition of connective operations can then be achieved by means of the metalanguage.[7] For "adjunctive" the word "extensional" has been used; we prefer the use of "adjunctive" because the word "extensional" has been used in several meanings.

The truth tables of probability logic can be derived from the calculus of probability; they are given by theorems concerning the probability of certain compound classes. We present them in the table (13).[8] The symbols "b" and "c" used in these tables may be regarded as standing for individual sentences or for sequences of sentences, depending on the interpretation employed.

TRUTH TABLES OF PROBABILITY LOGIC

$P(b)$	$P(\bar{b})$
p	$1-p$

(13a)

$P(b)$	$P(c)$	$P(b,c)$	$P(b \vee c)$	$P(b.c)$	$P(b \supset c)$	$P(b \equiv c)$	$P(c,b)$
p	q	u	$p+q-p \cdot u$	$p \cdot u$	$1-p+p \cdot u$	$1-p-q+2p \cdot u$	$\frac{p \cdot u}{q}$

(13b)

The main difference between the tables (13) and (12) is that in table (13) we need a third argument column. The probability of the compound sentence is not determined by the probabilities of the individual sentences; we need a third probability, the relative probability from "b" to "c", which may be regarded as a degree of coupling between two sentences. Some logicians have objected that for this reason probability logic is not extensional. This is true only for a very narrow meaning of the word "extensional." The truth tables of probability logic are certainly adjunctive in the sense defined above. That the probability of a compound sentence is a function of three parameters, and not

[7] Cf. the author's *Elements of Symbolic Logic* (New York, 1947), § 7 and § 9, and chap. viii.

[8] These truth tables were first published by the author in *Sitzungsber. d. Preuss. Akad., Phys.-math. Kl.* (Berlin, 1932, p. 476). The truth tables (20) of the quantitative negation were added in the present publication.

of two, expresses a generalization of a kind well known in mathematics. Thus in Euclidean geometry the third angle of a triangle is determined by the two other angles, whereas in a non-Euclidean geometry of a given constant curvature this determination requires a third parameter, the area.

Incidentally, we could use in probability logic, instead of the relative probability from "b" to "c", the probability of the compound "$b.c$" (or of another compound) as the third independent parameter; the probabilities of the other compounds would then be determined. We shall regard the expression "a,b" also as a compound proposition; it has properties similar to those of the other compounds. The expression "$P(a,b)$" then is of the same form as the expression "$P(a.b)$" or "$P(a \vee b)$".

In addition to the truth tables we have for the three independent parameters the inequality

$$\frac{p+q-1}{p} \leqq u \leqq \frac{q}{p} \tag{14}$$

This relation is derived from the postulate that not only the fundamental probabilities but also all probabilities derivable from them are subject to the normalization condition of being numbers between 0 and 1 limits included. For the value $p = 1$ we derive from (14) that $u = q$. In this case, therefore, the third parameter becomes a function of the two other ones. It can be shown that by the use of this condition the truth tables (12) of two-valued logic can be derived from the truth tables (13) of probability logic. Two-valued logic appears therefore as a special case of probability logic; it is even a degenerate case in which the general dependence on three parameters is eliminated and replaced by a dependence on two parameters only.

A problem that requires particular investigation is the problem of assertion in probability logic. In two-valued logic we follow the rule that only true sentences can be asserted. If a sentence "a" is false, however, we need not resort to the metalinguistic statement " "a" is false", but can express the falsehood by asserting the negation "$\bar{a}$". In fact it is one of the major functions of the negation that it allows us to express falsehood in the object language.

In a logic of the three truth values t_1, t_2, t_3, assertability is restricted to one of these values, for which we may choose t_3. In order to express the other two truth values in the object language we introduce a cyclical negation by the tables (15)

a	$\sim a$
t_3	t_2
t_2	t_1
t_1	t_3

(15)

When we wish to assert that a sentence "a" has the truth value t_1 we assert the sentence

$$\sim a \tag{16}$$

When the sentence "a" has the truth value t_2 we assert the sentence

$$\sim\sim a \tag{17}$$

We see that the subscript of the truth value corresponds to the number of negations in the assertable sentence. This holds also for the value t_3 since the true sentence can also be written in the form

$$\sim\sim\sim a \tag{18}$$

This procedure can be transferred to probability logic. Usually the probability p of a sentence "a" is expressed by the metalinguistic statement

$$P(a) = p \tag{19}$$

In order to construct an object equivalent of this metalinguistic sentence we introduce a *quantitative negation.* We negate a sentence "a" to the degree w by putting the symbol "$\lceil w \rceil$" before the sentence; we thus obtain the sentence "$\lceil w \rceil a$". The quantitative negation is defined by the truth tables (20):

$P(a)$	$P(\lceil w \rceil a)$
p	$p - w + \delta_{p-w}$
	$\delta_{p-w} = \begin{cases} +1 & \text{for } p - w \leqq 0 \\ 0 & \text{for } 0 < p - w < 1 \\ -1 & \text{for } p - w = 1 \end{cases}$

(20)

w is a real number between 0 and 1 limits included. We see from the table (20) that if we negate the sentence "a" to the degree p, i.e., if $w = p$, the resulting sentence has the probability 1. Following the rule that only sentences of the probability 1 can be asserted, we can express the degree p of a statement "a" by asserting the statement

$$\lceil p \rceil a \tag{21}$$

This sentence of the object language takes the place of the metalinguistic statement (19). It is easily seen from the truth tables (20) that the statement (21) has the probability 1 if and only if p is the probability of "a". We see, furthermore, that a negation to the degree $w = 1$ leaves the truth value of the statement unchanged. A negation to the degree $w = 0$ in general leaves also the truth value unchanged, except for the case $p = 0$ or $p = 1$, where this negation reverses the truth value.

By means of the quantitative negation we can write derivations in the object language. Thus we have the inferential schema:

$$\begin{array}{l} \lceil p \rceil a \\ \lceil u \rceil (a,b) \\ \lceil v \rceil (\bar{a},b) \\ \hline \lceil p\cdot u + (1 - p)\cdot v \rceil\, b \end{array} \tag{22}$$

For the case $p = 1$ the schema assumes the simple form

$$\frac{\begin{array}{l} a \\ [u]\,(a,b) \end{array}}{[u]b} \tag{23}$$

The two schemas (22) and (23) can be regarded as generalizations of the modus ponens of two-valued logic.

The concept of tautology can be transferred to probability logic by the definition that tautologies are formulas which have the probability 1 for all probability values of their components. It can be shown that all tautologies of two-valued logic remain tautologies within probability logic. In addition, however, we can construct in probability logic tautologies that cannot be written in two-valued logic. This is achieved by means of the quantitative negation. Thus the schema (22) furnishes the tautology

$$[p]a\,.\,[u](a,b)\,.\,[v](\bar{a},b) \rightarrow [p\cdot u + (1-p)\cdot v]b \tag{24}$$

The arrow indicates a particular form of implication which we do not specify here.

These considerations show that probability logic is a generalization of two-valued logic containing the latter as a special case. Probability logic, therefore, will be applicable to those forms of knowledge in which truth is replaced by probability.

V

Let us summarize the two major results of our analysis.

1. We constructed the calculus of probability in axiomatic form. The calculus so constructed is of a general kind including all types of sequences; the random sequences represent only a special type. The axioms of this general calculus of probability follow from the frequency interpretation.

2. We transformed the calculus of probability into a probability logic by means of truth tables that appear as a generalization of those of two-valued logic.

VI

After presenting the formal system of probability we now turn to the analysis of the application of the formal system to physical reality.

The problem of application may be divided into two parts. In the first part we analyze the meaning of the concept of probability underlying applications to physical reality. In the second part we inquire with what title probability statements about physical objects can be asserted. The two problems, that of *meaning* and that of *assertability*, are closely connected since the assertability will depend largely on the meaning assumed for the probability statement.

Turning to the problem of meaning, let us repeat that the formal system does not prescribe one interpretation as the only admissible one. We said that a set of axioms can be regarded as a set of implicit definitions; they delimit the fundamental concepts only to a certain extent, leaving open a class of admissible interpretations. The meaning of the applied concept of probability is therefore not determined by the formal system. This system will rule out some interpretations as inadmissible; but which of the admissible interpretations is to be used for the applied concept must be determined by considerations outside the formal system.

The problem of meaning may be subdivided into two parts, depending on the type of object to which the concept of probability is applied. The first type of interpretation is given when the object to which the concept of probability is applied consists in a *sequence*, which may be either a sequence of physical events or of propositions about events (propositional sequence). The second type is given by an interpretation which refers to individual events or individual propositions. Let us discuss the two types of interpretation in this order.

VII

As long as sequences are regarded as the objects of probability statements the question of interpretation is not controversial. Practically all logicians are agreed that for such objects only the limit of the relative frequency supplies an adequate interpretation. Let us, therefore, restrict our discussion of the first type of interpretation to the frequency interpretation and study its advantages and difficulties.

The problem of meaning is here identical with the question of whether a statement about the limit of the frequency can be regarded as meaningful. Now it is obvious that there is no difficulty with limit statements if they refer to an *intensionally given* sequence, i.e., a sequence given by a mathematical rule. For instance, to use an example suggested by Poincaré, that the relative frequency of odd and even numbers in the last digit of a table of logarithms converges to the limit 1/2 is mathematically demonstrable; the meaning of the statement is therefore not controversial. Now it is an advantage of our construction of the calculus of probability that intensionally given sequences are admissible interpretations; even normal sequences can be constructed by means of mathematical rules, as has been shown by Arthur Copeland.[9] But we know that in physical applications we are usually concerned with *extensionally given* sequences, i.e., sequences given element by element. Since it is impossible to enumerate an infinite sequence, it has been questioned whether a statement about the limit in such a sequence is meaningful. We cannot determine, for such sequences, the relation between the width ϵ of convergence and the number n of the element, i.e., we cannot say for which n a convergence within ϵ is reached.

This difficulty, however, does not seem to me so serious as it has appeared to other logicians. That in applied statistics we speak of infinite sequences must be regarded as an idealization which we use for the sake of convenience although we know that all applications are restricted to finite sequences. The idealization may be compared to the one used in geometry, where we speak of

[9] As cited in note 2 above.

lines and points without width although we know that physical objects do not strictly satisfy this requirement. It would be possible to replace idealized geometry by a geometry in which points are small areas and lines have a narrow width; in this geometry the usual theorems would hold only approximately. Similarly, we could construct a *finitized* calculus of probability which deals with sequences of a finite length possessing some sort of convergence toward a limit within an interval ϵ which could be defined by suitable methods. While the theorems of the elementary calculus of probability would even hold strictly for that calculus, the theorems of the theory of order would hold only approximately and would have to be carefully worded in a way that would include reference to the interval of convergence and the length of the subsequences. There is no doubt that by the construction of such a calculus all postulates of the *finitist* could be strictly satisfied. When we prefer to speak of infinite sequences we do so because of the great simplification introduced by this idealization. If, however, the question of meaning is under consideration, we may always refer to the fact that for all purposes of application finite sequences will suffice. In fact, the "limit" of the applied calculus is meant to be a *practical limit.* A sequence has a practical limit if, for a finite number of elements, large enough and yet accessible to human experience, it shows properties of convergence. This definition excludes sequences that converge so late that for all human observations they behave like nonconvergent series; on the other hand, it admits sequences that diverge after the section of practical convergence and have no limit when continued to infinity. The concept of practical limit will answer the question of meaning for the probability sequences of applications.

Let us now turn to the problem of assertability. It may be subdivided into the problem of the assertability of probability laws and the problem of the ascertainment of the degree of probability. The first problem is easily answered: since the axioms of the calculus follow from the frequency interpretation, the laws of probability are guaranteed, for this interpretation, by deductive logic. This is a great advantage of the frequency interpretation; it is made possible by the generalized form of the calculus of probability, in which the inner structure of probability sequences is regarded as unessential. The definition of various types of order is given within a special chapter, the theory of order. We saw that all these definitions are constructed in the form of postulates stating that certain probabilities are equal. The determination of a type of order, therefore, requires no more than the ascertainment of a degree of probability.

It is the second problem, the ascertainment of the degree of probability, that leads into difficulties for extensionally given sequences. While for an intensionally given sequence the value of the limit of the frequency can be derived from the defining rule, there exists for extensionally given sequences only the method of enumeration: we have to count the whole sequence in order to know the limit. For infinite sequences this is certainly impossible. We should realize, however, that the difficulties of this problem are not removed by a finitization. Although a finite sequence can be enumerated, in principle, this possibility does not help us because, in all practical applications, we wish to know the value of the limit before the total sequence is observed. We resort here to the method of counting an initial section of the sequence and then assuming that the observed frequency will persist on further prolongation of the sequence. This procedure represents the *inductive inference*. The frequency interpretation of probability, whether employed for infinite or for finite sequences, is therefore burdened with the problem of induction. A satisfactory solution of the problem of applicability can be given only when it is possible to solve the problem of induction.

It is here that we meet with a specific difficulty of the calculus of probability. For other axiomatic systems there exists no problem of application; the axioms can be applied if the physical objects under consideration have the properties required in the axioms, and instead of a problem of applicability we have the requirement of a *suitable choice of the objects*. With respect to probability sequences this method breaks down because it is the very question whether the physical objects possess the necessary properties that cannot be answered. If we could wait until the total probability sequence is observed we could apply the method of the suitable choice of the object, rejecting as unsuitable any sequence that does not have the required property. In practical statistics, however, we cannot wait until the whole sequence is observed for the reason that we wish to use probability values for predictions. It is the predictive nature of the applied calculus of probabilities that leads into the difficulties of the problem of induction.

VIII

Let us now turn to the investigation of applications in which the concept of probability refers to an individual event or to an individual proposition. It is on this ground that the frequency interpretation has been questioned. Some logicians have argued that we are concerned here with a different notion of probability not reducible to frequencies. Let us inquire whether the contention of the existence of two disparate notions of probability is tenable.

At first sight, indeed, it appears as though a probability applied to a single case has nothing to do with a frequency. We say, "it is probable that it will rain tomorrow"; "it is improbable that Julius Caesar was in Great Britain"; etc.; and thus refer probability to a single event, or in the logical conception to a single proposition. What does it help us to know that in a certain percentage of days of certain meteorological conditions it will rain, when we wish to know the probability for rain on one individual day? Similarly the example of Julius Caesar's stay in Britain has often been quoted as denying a frequency interpretation. Let us analyze the various meanings that can be suggested for such a second concept of probability.

In the first interpretation, the degree of probability is regarded as a measure of the intensity of expectation with which we anticipate a future event. This interpretation, however, leads into difficulties because the feeling of expectation varies from person to person; we rather use probabilities as a standard of what the intensity of expectation *should be*, but not as a measure of what it *is*.

Thus the optimist is controlled by too high an expectation if the event expected is desirable; the pessimist, on the other hand, suffers from too low an expectation. But if probability is a standard of what expectation should be it cannot be identified with the intensity of a psychological status. Furthermore, the validity of the laws of probability is by no means warranted by such an interpretation.

The second and third interpretations to be considered derive from a problem which historically speaking constituted the first philosophical issue by which the theory of probability was confronted. The historical origin of the calculus of probability from the study of the games of chance has led to the conception that a degree of probability can be ascertained by means of an a priori method that is attached to a disjunction of equally probable cases. Throwing a die, for instance, we frequently argue that the six possible cases should be equally probable because we have no reason to prefer one face to the other. The principle of this inference was called the *principle of indifference*, or of *no reason to the contrary*. It was introduced by Laplace; and although mathematicians have long since abandoned this principle, it has haunted the field of philosophical inquiry into the nature of probabilities. Up to our day it is defended fervently by some philosophers who hope by such means to secure for themselves a fenced-off area, a reservation, safe from the precision of mathematical methods.

That the principle of indifference is logically untenable has been sufficiently demonstrated. Maybe we have no reason to prefer one face of the die to the other; but then we have no reason to assume that the faces are equally probable, either. To transform the absence of a reason into a positive reason represents a feat of oratorical art worthy of an attorney of the defense but not permissible in the court of logic. Moreover, it has been demonstrated clearly enough that the principle of indifference leads to contradictions when applied to geometrical probabilities connected by a nonlinear measure transformation. I should like to classify the principle of indifference as a *fallacy of incomplete schematization*. It leads to a true conclusion in the case of a die possessing geometrical symmetry, but to false conclusions in other cases. Where it leads to true conclusions it can do so only because more is known than is stated in the principle. In the case of the die, for instance, we can derive the equiprobability of the faces by a more complicated schema of inference that contains in its premises a very general empirical statement about probability functions.[10] Unfortunately, the incomplete schematization formulated in the principle of indifference has misled some philosophers into construing the concept of probability in such a way that the principle can be maintained.

The first attempt to save the principle of indifference is based on the logical *principle of retrogression*. This principle plays a part in the theory of meaning; it states that the meaning of a sentence is given by the method of its verification. Since, according to the adherents of the a priori conception of probability, we determine a probability by counting the terms of an exclusive disjunction in which we have no reason to prefer one term, the principle of retrogression

[10] Cf. the author's *Wahrscheinlichkeitslehre* (Leiden, 1935), § 65.

furnishes the result that the meaning of probability is given by reference to such a disjunction. We thus arrive at a *retrogressive interpretation* of probability. For instance, that the probability of obtaining face 6 with a die is 1/6 means, according to this interpretation, that face 6 is a term of a disjunction of six terms and that we have no reason to prefer one of the terms.[11] It is obvious that with this interpretation the use of the principle of indifference is justified since then the probability statement states no more than what is assumed as the premise of that principle; but it is equally obvious that with this interpretation the probability statement has lost its predictional value. Why should we bet on the occurrence of the event "non-six" rather than on the occurrence of "six"? The retrogressive interpretation narrows down the meaning of probability statements in such a way that the assertion of the statement is justified; but in the transition from probability statements to bets, or advices to action, there reappears the very problem that the retrogressive interpretation was intended to evade and that the principle of indifference cannot solve.

The retrogressive interpretation has been used in a second form to cover another fallacious application of probability inferences, called the *inference by confirmation*.[12] It is claimed by some logicians that the probability of hypothetical assumptions is derived by means of the following inference: When a certain consequence of the assumption is verified, such observation is regarded as conferring a certain degree of probability to the assumption. This means an inference of the form

$$\begin{array}{ll} \text{"}a \supset b\text{"} & \text{is proved} \\ \text{"}b\text{"} & \text{is verified} \\ \hline \text{"}a\text{"} & \text{is probable} \end{array}$$

The theory of probability knows no such inference. Whenever such inference is successfully applied, we can show that it represents an incomplete schematization, that much more is known than is expressed in the inference. In fact, the inferences made in the theory of indirect evidence and in the verification of scientific theories must be construed as inferences in terms of the rule of Bayes. They can be made only when the values of some other probabilities are known or, at least, can be roughly estimated.

In order to save the inference by confirmation an attempt has been made to construct a retrogressive interpretation of probability in such a way that the meaning of probability is given by the premises of the inference. This kind of probability is called *degree of confirmation*. Although, of course, such an interpretation would justify the inference, it is clear that the degree of confirmation so defined possesses no predictional value and cannot account for the reliability of hypothetical assumptions.

---

[11] As far as I know, the first to use this interpretation was K. Stumpff in a paper published in *Berichte d. Bayer. Akad., Philos. Kl.* (Munich, 1892). The interpretation was taken up in our day by some logicians under the influence of ideas of L. Wittgenstein, *Tractatus Logico-philosophicus* (London, 1922), p. 113; thus by A. Waisman, *Erkenntnis*, V, 1 (1930), 229.

[12] Cf. R. Carnap, "Testability and meaning," *Philosophy of Science*, III (1937), 420. Carnap's ideas on confirmation were carried on chiefly by E. Nagel and C. Hempel.

Finally we may mention here a retrogressive interpretation of the principle of induction. According to this interpretation, the meaning of the conclusion of the inductive inference is given by the statement of the premises. Thus when we infer from past observations that the sun will rise tomorrow we mean by this conclusion, according to the interpretation, that we have seen the sun rising in past observations. This interpretation is so absurd that it has scarcely ever been maintained. It is obvious that with it the principle of induction would lose its predictional value.

In view of the deficiencies of the retrogressive interpretation, the adherents of the a priori determination of probability degrees have introduced another interpretation, which abandons the principle of retrogression and regards probability as a *primitive concept* not capable of further definition. According to this conception, the statement that the probability of an expected event is 1/6 has a meaning of its own, comparable to the meaning of the primitive notions of logic; and we cannot interpret this meaning as a frequency or a report about terms in a disjunction. This conception is sometimes stated in the form that probability is a *rational belief*, that the laws of probability constitute a *quantitative logic* based on a self-evidence comparable to those of ordinary logic.[13] As far as I see, the primitive-concept interpretation is not always clearly distinguished from the retrogressive interpretation by its adherents; it appears that some logicians vacillate between the two interpretations and use sometimes the one, sometimes the other, depending upon what they wish to prove.

The difficulties of the primitive-concept interpretation appear to me so overwhelming that I cannot see how a logician can commit himself to this interpretation. First, the degree of probability remains unverifiable. When the event expected with a probability 5/6 is observed, does this observation verify the probability statement? Obviously not, since the nonoccurrence of the event is also compatible with the probability statement. A numerical value of a probability cannot be ascertained by one observation. We do not escape this predicament by restricting probability statements to relations of order stating that a probability is higher or lower than another; such relations cannot be verified by one observation, either. It is sometimes argued that the verification of the degree of probability is obtained not by the observation of the event but by other methods such as used in the principle of indifference. But such procedure can be regarded as a verification only if a retrogressive interpretation of probability is adopted. With this turn, however, the primitive-concept interpretation is abandoned and the interpretation loses its predictional value.

Another difficulty of the primitive-concept interpretation is the justification of the laws of probability. In fact, the whole calculus of probability and its application to physical objects appears here as a system based on *synthetic self-evidence*. A philosophy of probability that would commit itself to interpret

[13] This conception is represented by the ideas of J. M. Keynes, *Treatise on Probability* (London, 1921), and was continued by H. Jeffreys, *Theory of Probability* (Oxford, 1939). I could not say to what extent it is present also in the ideas of Carnap and others about confirmation. If the so-called degree of confirmation is meant to be a measure of reliability and an advice to action, it falls under this category.

probability as a primitive notion would lead logic back to rationalism, to a *synthetic a priori;* in other words, to a metaphysics that claims an intrinsic correspondence between reason and physical reality.

IX

The analysis of meaning has suffered from too close an attachment to psychological considerations. The meaning of a sentence has been identified with the pictures and representations associated with the utterance of the sentence. Such conception will lead to meanings varying from person to person; and it will not help us to find out the meaning a man would adopt if he had a clear insight into the implications of his words. Logic is interested not in what a man means but in what he *should* mean; that is, in that meaning which, if assumed for his words, would make his words compatible with his actions. When we analyze the meaning of probability statements about single events by the use of this objective criterion, we find that the frequency interpretation can be applied to this case too, and that we need not resort to one of the questionable interpretations based on the reconstruction of subjective psychological intentions, discussed above.

Assume that the frequency of an event B in a sequence is $= 5/6$. Confronted by the question of whether an individual event B will happen, we shall prefer to answer in the affirmative because, if we do so repeatedly, we shall be right in 5/6 of the cases. We shall not claim that the individual assertion is true; we shall assert it in the sense of a *posit*, i.e., in the attitude of a man who lays a bet. A posit is a statement with which we deal as true although we have no knowledge about its truth. The greater a probability in the frequency sense, the more favorable will it be to posit the individual statement because on repetition we shall have a greater number of successful predictions. The probability appears as a *rating* of the posit, which we call its *weight.*

According to this conception, the probability of an individual statement appears as a fictitious property resulting from a *transfer of meaning from the general to the particular case.* Strictly speaking, it has no meaning that in an individual case the probability of casting "non-six" with a die is $= 5/6$; but when we coördinate to this statement a fictitious meaning it will lead to a behavior that, in repeated applications, will be the most successful one. The frequency interpretation allows us to construct a fictitious meaning for the probability of individual events, or propositions, of such a kind that it makes our words compatible with our actions.

It is not necessary for this conception that the events to be repeated are all of the same kind. Events of various kinds and probabilities may be connected into a sequence such that always positing the more probable event will lead to the greater number of successes. The sequences of insignificant events of everyday life furnish large enough numbers admitting this application of the frequency interpretation.

An apparent difficulty results from the fact that, given an individual event, we often do not know in what reference class we should incorporate it. Conversational language is none too precise in this respect; we speak of the probability of the death of a certain person, of the probability of an expected political event, etc., without explicitly stating a reference class for which the probability is to be constructed. In such cases the statement may be understood to mean: the probability of the event with respect to the best reference class available. This reference class may be defined as the narrowest reference class for which we have reliable statistics. Thus a physician, when asked for the probability of the death of a certain person, will know into which reference class the case should be incorporated. That a suitable choice of the reference class for political events is so difficult indicates that the statistical laws of politics are none too well known.

The logical form of a language that deals with probabilities as truth values of individual sentences is the probability logic presented above. Since conversational as well as scientific language, to a great extent, is of this type, probability logic constitutes the form of a great part of actual language. In fact, the use of two-valued logic in statements about physical reality must be regarded as a degenerate form of probability logic, in which only high and low probabilities are employed, while intermediate values are omitted. That the meaning of the term "probable" in statements such as "Peter will probably come," or "the enemy will probably accept the ultimatum," can be assumed to be the same as the one used in mathematical statistics, is guaranteed by the fact that the truth tables of probability logic are derivable from the calculus of probability.

X

The concept of weight may also be applied on a higher level, i.e., with reference not to propositions about facts but to propositions stating the probabilities of other propositions. From the calculus of probability it is known that probabilities for the occurrence of certain frequencies can be derived; thus the Bernoulli theorem furnishes probabilities for the convergence of the frequency toward a certain limit. If the limit of the frequency under consideration is regarded as a probability of the first level, the Bernoulli probability is a probability of the second level. In those cases where probabilities of the second level can be computed we can employ them as weights determining a rating for an inductive inference; they tell us the reliability of the assumption that the observed frequency will persist. In fact, most inductive inferences are made, not as isolated inferences, but within a network of other inductions. The theory of indirect evidence can be constructed in terms of such a network; the inferences connecting the individual inductions are covered by the theorems of the calculus of probability. The concatenation of inductions thus achieved improves greatly the reliability of inductive methods; on the other hand, the analysis of such network inferences allows us to account for more complicated forms of induction which do not directly possess the form of induction by enumeration.

In view of the fact that all axioms of the calculus of probability are derivable from the frequency interpretation, we can now formulate the two following results:

1. All probability inferences are reducible to induction by enumeration with the addition of deductive inferences.

2. Although some inductive inferences can be given a rating expressed by probabilities, there will always remain other inductive inferences whose weight is unknown. The general theory of induction does not constitute a chapter of probability theory, but must be given without the use of probability considerations.

XI

We now turn to the discussion of the one remaining difficulty connected with the frequency interpretation—of the general problem of induction.

The inference of induction has found its critic in David Hume. Although Hume attached his famous criticism to the particular case that the relative frequency of the event is $=1$, his results hold likewise for the general case of the statistical inference in which the persistence of an observed percentage of events is assumed. What Hume[14] has shown is:

1) The inductive inference is not logically a priori, i.e., the conclusion is not a necessary consequence of the premises.

2) The inductive inference is not logically a posteriori. Any attempt to explain the inductive inference as a result of past experiences in which the inference was used successfully is circular reasoning because the inductive inference would be used for its own justification.

The truth of these results is unquestionable. Does it follow, however, that no justification of induction can be given?

Hume believed that this consequence is inescapable. His theory that induction is a habit is not meant to be a justification; and it is no way out of the difficulty, because it refers to a psychological fact that is logically irrelevant. There are good habits and bad habits; the logical problem is whether induction is a good habit.

I think that it is possible to give a positive answer to this question. The inductive inference can be justified; induction can be shown to be a good habit. This proof, however, requires a reinterpretation of language; scientific and other statements must be regarded, not as *assertions* claimed to be *true*, but as *posits* claimed to be *the best posits* we can make.

If this revision of the claims of language is accepted, the justification of induction is rather easily given. We speak of a justification if it can be shown that for the pursuit of a certain aim it is advisable to apply a certain means; a justification always concerns a means with respect to an end. Let us formulate the aim of making predictions, which is common to science and everyday life, in the form that we wish to find limits of the relative frequency in sequences. When we now regard the inductive inference as a rule for constructing posits, to be applied repeatedly in the sense of a trial subject to later correction, we can show that, if the sequence has a limit, the inductive inference will lead to this limit within an interval ϵ of exactness in a finite number of steps. This result follows from the definition of the limit. If, on the other hand, the se-

[14] David Hume, *Enquiry concerning Human Understanding* (1748).

quence has no limit of the frequency, the inductive rule will not find it—but then, no other method will find it, either. The use of the rule of induction, therefore, can be regarded as the fulfillment of a necessary condition of success in a situation in which a sufficient condition is unknown to us. To speak in such a case of a justified use of a rule appears in agreement with linguistic usage concerning the word "justification." Thus we call Magellan's enterprise justified because, if he wanted to find a thoroughfare through the Americas, he had to sail along the coast until he found one—that there was a thoroughfare was by no means guaranteed. He could act only on the basis of necessary conditions of success; that they would turn out sufficient was unknowable to him.

This simple consideration solves the problem of induction. It removes the last difficulty connected with the frequency interpretation of probability. It requires, on the other hand, a renunciation of a rationalistic attitude toward knowledge; we have to give up the quest for certainty if we wish to account for the use of probability methods. Such reconversion of emotional attitudes is not always easy; but once it is achieved it offers us the greatest reward that philosophical analysis can ever find: it supplies a proof that our methods of knowledge are the best instrument of finding predictions, if predictions can be found. I could not think of a better justification of scientific method than the proof that to apply such method is the best we can do.