THE THEORY OF PROBABILITY

An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability

By HANS REICHENBACH

PROFESSOR OF PHILOSOPHY IN THE UNIVERSITY OF CALIFORNIA AT LOS ANGELES

ENGLISH TRANSLATION BY

ERNEST H. HUTTEN

AND

MARIA REICHENBACH

Second Edition

UNIVERSITY OF CALIFORNIA PRESS

BERKELEY AND LOS ANGELES · 1949

DeVos

Chapter 3

ELEMENTARY CALCULUS OF PROBABILITY

ELEMENTARY CALCULUS OF PROBABILITY

§ 9. The Probability Implication

The investigation of the concept of probability begins with an analysis of the logical structure of probability statements. The problem, which so far has not been given sufficient attention in the mathematical calculus of probability, is amenable to precise solution with symbolic methods. Symbolic logic has devised means of characterizing the logical form of a statement without regard to its content; these methods can be extended to include a characterization of probability statements. The formalization of the probability statement, in fact, is one of the first objectives in the philosophy of probability.

To consider a typical probability statement: when a die is thrown, the appearance of face 1 is to be expected with the probability $\frac{1}{6}$. This statement has the logical form of a relation. It is not asserted unconditionally that face 1 will appear with the probability $\frac{1}{6}$; the assertion, rather, is subject to the condition that the die be thrown. If it is thrown, the occurrence of face 1 is to be expected with the probability $\frac{1}{6}$; this is the form in which the probability statement is asserted. No one would say that the probability of finding a die on the table with face 1 up has the value $\frac{1}{6}$, if the die had not been thrown. Probability statements therefore have the character of an implication; they contain a first term and a second term, and the relation of probability is asserted to hold between these terms. This relation may be called *probability implication.* It is represented by the symbol

$$\underset{p}{\ni}$$

This is the only new symbol that the probability calculus adds to the symbols of the calculus of logic. Its connection with *logical implication* is indicated by the form of the symbol: a bar is drawn across the sign of logical implication. Whereas the logical implication corresponds to statements of the kind, "If a is true, then b is true", the probability implication expresses statements of the kind, "If a is true, then b is probable to the degree p".

The terms between which the probability implication holds are usually events. Let x be the event, "The die is thrown", and y the event, "The die has come to rest on the table"; then a probability implication between the two events is asserted. We recognize at once that this requires a more exact formulation. We speak of a definite probability only when the event is characterized in a certain manner, namely, as an event y in which face 1 is up. This means that the event y is regarded as belonging to a certain class B. We are dealing with a class, since the individual features of the event y are disregarded in the statement. It does not matter on what part of the table the die lies, or in which direction its edges point; only the attribute of having face 1 up is considered. Thus the event y is characterized only as to whether it can be said to belong to the class B. The same applies to the event x, since we do not consider with what force the die is thrown or what angular momentum is imparted to it; we demand only that x be a throw of the die, that it belong to a certain class A. Therefore we write the probability statement in the form

$$x \in A \underset{p}{\ni} y \in B \tag{1}$$

This formulation, however, requires modification. We must express the fact that the elements of the classes are given in a certain order, for instance, in the order of time. In other words, the event x belongs to the discrete sequence of the events $x_1, x_2, \ldots x_i \ldots$, while at the same time the event y belongs to a corresponding sequence $y_1, y_2, \ldots y_i \ldots$ There is a one-one correspondence between the elements of the two sequences, expressed by equality of subscripts, and we assert only a probability implication between the corresponding elements x_i, y_i, so that we write, instead of (1),

$$x_i \in A \underset{p}{\ni} y_i \in B \tag{2}$$

The coördination of the event sequences is necessary for the following reason. We do not wish to say that the probability implication holds, for instance, between the event x_i of throwing the die and the event y_{i+1} of obtaining a certain result. When we merely state that the event x_i of throwing the die occurs, we have not yet asserted that the event x_{i+1} of throwing the die will also occur and that, therefore, a probability for the occurrence of the event y_{i+1} exists.

However, even (2) does not completely represent the form of the probability statement; we must add the assertion that the same probability implication holds for each pair x_i, y_i. This generalization is expressible by two all-operators, meaning, "for all x_i and for all y_i". Using an abbreviation, we can reduce the two all-operators to one by placing only the subscript i in the parentheses of the operator. Thus the probability statement is written

$$(i)\ (x_i \in A \underset{p}{\ni} y_i \in B) \tag{3}$$

This expression represents the final form of the probability statement: *The probability statement is a general implication between statements concerning a class membership of the elements of certain given sequences.*

To illustrate this formulation of the probability statement: a relation of the kind described is employed in dealing with the probability of a case of influenza leading to death. We do not speak unconditionally of the probability of the death of the patient, but only of the probability resulting from the fact that he has contracted influenza. Here again are two classes—the class of influenza cases and the class of fatal cases—and the probability implication is asserted to hold between them. If x_i is interpreted as a result of medical diagnosis, A as an influenza case, y_i as the state of the patient after one week of illness, and B as the death of the patient, then this example of a probability statement from the field of medicine has the form (3).

Another example is the probability of hitting a target during a rifle match. Here x_i represents the single shot, y_i the hit scored at the target, B the class of hits within a certain range, and A the class to which the rifleman belongs according to his ability. The probability of a hit will be different according to the contestant's degree of skill. Here again the probability is determined only when the classes A and B are chosen.

An example from physics is the bombardment of nitrogen by α-rays, or helium nuclei. There is a certain probability that a helium nucleus will eject a hydrogen nucleus from the nitrogen atom. Let A represent the class of α-rays, x_i the hit of an individual helium nucleus, and y_i the event produced by it. The event results in the occasional emission of a hydrogen nucleus, that is, it belongs to the class B. Although it is not possible to observe directly the causal connection between the helium nucleus and the released hydrogen nucleus, we assume, nevertheless, a one-one correspondence between x_i and y_i. Using a very weak radioactive preparation that rarely emits helium nuclei, we can employ the temporal coincidence observed for the α-rays and the hydrogen rays as a criterion of the correspondence.

In the previous examples, x_i and y_i stand in the relation of cause to effect, but other instances can easily be found in which y_i represents the cause and x_i the effect. In this case we carry out a reverse inference, from the effect to the probability of a certain cause, for example, in investigating the cause of a cold. And there are other examples for which the relation x_i to y_i is of a still different type. There exists a probability that a certain position of the barometer indicates rain, but there is no direct causal connection between the two events. In other words, one is not the cause of the other. Rather, the two events are effects produced by a common cause, which leads to their concatenation in terms of probabilities. It is easily seen that these examples also conform to the logical structure of (3).

The analysis presented shows that the probability implication can be regarded as a relation between classes. The class A will be called the *reference class;* the class B, the *attribute class.* It is the probability of the attribute B that is considered with reference to A. It must be added, however, that the probability relation between the two classes A and B is determined only after the elements of the classes are put into a one-one correspondence and ordered in sequences. For instance, the probability implication holding between the birth and the subsequent death of an infant—the rate of infant mortality—differs from one country to another, that is, it differs according to the sequence of events for which the statistics are tabulated. Even for an individual die there exists a particular pair $x_i y_i$ of sequences, and it is an assertion derived from experience that the probability remains the same for different dice. Therefore, strictly speaking, the probability implication must be regarded as a three-term relation between two classes and a sequence pair. The pair of sequences provides the *domain* with respect to which the probability implication assumes a determinate degree. Later the conception is extended to combinations of more than two sequences. The significance of the order of sequences is the subject of chapter 4.

Because of the equivalence that exists between classes and propositional functions, formula (3) may be expressed in a somewhat different way. According to (2, § 7), we may use instead of the statement $x \epsilon A$ the corresponding propositional functional $f(x)$ and, similarly, instead of $y \epsilon B$, the corresponding propositional functional $g(y)$. Then we must express the one-one correspondence between the sequences of x and y by a one-one functional $e(x,y)$ in order to determine for each x the corresponding value y. Thus (3) assumes the form

$$(x)(y)[f(x) . e(x,y) \underset{p}{\Rightarrow} g(y)] \tag{4}$$

In this form it is not necessary to employ the subscript i, if the order of the elements is regarded as understood.

A special kind of probability implication is included in the general form (3) or (4). It may happen that the sequences coincide and that the elements x_i and y_i are identical. The function $e(x,y)$ then reduces to the identity relation. We thus obtain, instead of (3) and (4),

$$(i)\ (x_i \epsilon B \underset{p}{\Rightarrow} x_i \epsilon B_k) \tag{5}$$

$$(x)\ [f(x) \underset{p}{\Rightarrow} g(x)] \tag{6}$$

Since it refers to a probability implication within the same sequence, this form will be called an *internal probability implication*. It is employed in many important problems of probability, particularly in social statistics. Examples are the probability that an inhabitant of Bavaria suffers from goiter, or that a new-born baby is a boy. In such cases x_i is not represented by an event but by a person or an object that may possess the two properties B and B_k simultaneously. In more strictly statistical applications, the internal form of the probability statement prevails to so high a degree that it is usually made the basis of the probability calculus. Yet it would not be advisable to restrict the probability statement to this special form, since there are numerous other cases in which the more general types (3) or (4) are used. In particular, the application of the probability concept to the causal connection of events would be impossible if it were not based on the more general form of the probability statement as given above.

§ 10. The Abbreviated Notation

The form of the probability statement as given in (3, § 9) is rather complicated. An abbreviated notation, therefore, will be used for the development of the calculus of probability. Abbreviation is possible because certain properties of formula (3, § 9) occur in all probability statements in a similar manner, and can be suppressed in a simplified notation.

The probability statement has been written, so far,

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

This formula will be abbreviated to the form

$$(A \underset{p}{\ni} B) \tag{2}$$

The transition from the abbreviated to the detailed notation is controlled by the following rule:

Rule of translation. *For every capital letter K substitute the expression $x_i \epsilon K$, using for different capital letters different variables x_i, y_i . . . , with the subscript i, but the same variable x_i for the capital letters K_1, K_2 . . . . In front of all parentheses containing capital letters place the symbol i within an all-operator.*

The method of abbreviation, as is seen from the rule, amounts to leaving out the specification of the sequence pair, an omission that is permissible because in probability statements the elements of the sequence pair never occur as free, but always as bound, variables. In the abbreviated notation, parentheses play the part of the all-operator; therefore, brackets must be used if generalization is not to be indicated. Furthermore, the difference between the two kinds of negation that exist for general statements is expressed as follows: in one case the negation bar is placed only above the expression written within parentheses; in the other it is extended above the parentheses. We thus define

$$(\overline{A \underset{p}{\ni} B}) =_{Df} (i)\ (\overline{x_i \epsilon A \underset{p}{\ni} y_i \epsilon B}) \tag{3}$$

$$\overline{(A \underset{p}{\ni} B)} =_{Df} \overline{(i)\ (x_i \epsilon A \underset{p}{\ni} y_i \epsilon B)} \tag{4}$$

The use of parentheses for the expression of the generalization applies also to formulas not containing the sign of the probability implication, and allows us to go from a class to a statement. Thus, $A \supset B$ is a class, and $(A \supset B)$ is a statement; according to the rule of translation, this statement has the form (20*b*, § 7) and is therefore identical with $A \subset B$. Adding parentheses to a class symbol means, in this notation, that the class is identical with the universal class and thus leads to the meaning expressed explicitly in (24, § 7).

If compound classes are used, like the class $A \supset B$, the rule of translation leads to the simple result: different capital letters mean narrower couple classes; equal capital letters with different subscripts mean simple classes. Couple classes containing implication or equivalence signs are interpreted by analogy with (13 and 14, § 7). The subscripts headed by circumflexes are dispensable for couple classes because their function is taken over by the difference of the capital letters. Class inclusion for different capital letters, i.e., for narrower couple classes, means a relation similar to the one illustrated in figure 3, § 7, for which the two circles are drawn in different planes, one on top of the other; corresponding points represent the couples of elements. Since for all practical purposes the narrower couple classes behave like simple classes, it is permissible to forget about the distinction for technical manipulations. The treatment of the general probability implication is technically not different from that of the internal probability implication.

A further abbreviation may be introduced. For many applications, particularly in mathematical calculations, we must solve the probability implication (2) for the degree p. We denote the degree p by $P(A,B)$, reading this symbol as "the probability from A to B". Some writers call this "the relative probability of B with respect to A". But in the present notation, the natural order, from the known to the unknown element of the relation, is used, thus introducing the same order of terms that is used in the implication $a \supset b$. The expression "probability from A to B" has the same grammatical form as the geometrical expression "distance from A to B", which also designates a relation. The order shows clearly that probabilities are treated as relations, in correspondence with the definition given in § 9. The calculus of probability in its usual form includes absolute as well as relative probabilities. The word "absolute" must be interpreted merely as an abbreviated notation, applying when the first term, the reference class, is dropped as being understood. Thus when it is said that there is the absolute probability $\frac{1}{6}$ for a face of the die, it is understood that the reference class is represented by the throwing of the die. This suppression of a first term has led to some confusion.

Instead of (2), then, the equation is written

$$P(A,B) = p$$

The p-symbol is a *numerical functor*, that is, a functional variable the special values of which are numbers.[1] It leads to statements only when it is used within mathematical equations. The P-symbol need not be considered as a primitive symbol; it can be reduced to the symbol of the probability implication by the definition

$$[P(A,B) = p] =_{Df} (A \underset{p}{\ni} B) \tag{5}$$

The symbol $P(A,B)$ itself is not defined—only the expression $P(A,B) = p$. This is permissible since the symbol $P(A,B)$ never occurs alone, but only in such equations. Thus a mere *definition in use* is given for $P(A,B)$. The equality sign used with this symbol represents arithmetical equality, i.e., equality between numbers. In the foregoing account of symbolic logic the sign was not explained because the rather complicated connection between logic and arithmetic could not be demonstrated. It may suffice to say that mathematical equality can be reduced to the basic logical operations.[2] The negation of a statement of mathematical equality is denoted by the inequality sign $\neq$. The notation by means of the P-symbol is called *mathematical notation;* that in terms of the $\underset{p}{\ni}$-symbol, *implicational notation.*

Another abbreviation is now introduced. Sometimes we omit the statement of the degree of probability and write

$$(A \ni B) \tag{6}$$

This relation is called *indeterminate probability implication.* Since it is not permissible simply to drop one constituent within a formula, a definition must be used to connect (6) with the symbols previously defined:

$$(A \ni B) =_{Df} (\exists p)(A \underset{p}{\ni} B) \tag{7}$$

The abbreviation (6) therefore means, "There is a p such that there exists between A and B a determinate probability implication of the degree p".

Passing from (6) to the detailed notation we obtain, according to the rule of translation,

$$(A \ni B) =_{Df} (\exists p)(i)(x_i \in A \underset{p}{\ni} y_i \in B) \tag{8}$$

The all-operator is placed after the existential operator, so that (8) represents the stronger form in the sense of (9, § 6).

The value p is often written within separate parentheses behind the probability implication:

$$(\exists q)(A \underset{q}{\ni} B).(q = p) \tag{9}$$

[1] See *ESL*, p. 312.

[2] It is an identity of classes of a higher type. See *ibid.*, § 44.

This is merely a more convenient way of writing and has the same meaning as (2). We need this form because we shall later obtain for the probability degree p expressions that are too involved to be written as subscripts of the symbol of the probability implication. The resulting parentheses in the expression $(q = p)$ do not indicate an all-operator for the detailed notation because they do not contain capital letters.

The abbreviations given in this section will be useful in the following presentation of the theory of probability. In particular, it is an advantage that even in the abbreviated notation the symbols of the propositional operations can be manipulated according to the rules of the propositional calculus, although these symbols are placed between class symbols (that is, between capital letters) and thus represent class operations. This is possible because of the isomorphism of the two calculi (see § 7).

§ 11. The Rule of Existence

The *formal structure* of probability statements has been explained, but nothing has been said so far about their *meaning*. The laws of the probability implication can be completely developed, however, without interpretation. Discussion of the problem of interpretation will be deferred to a later section.

As a consequence, a method cannot yet be provided whereby we can determine whether, if two classes are given, a probability implication holds between them; in other words, we cannot yet ascertain the *existence* of a probability implication. However, this impossibility need not disturb us at this point. We assume the existence of some probability implications to be given; and we deal only with the question of how to derive new probability implications from the given ones. This operation exhausts the purpose of the probability calculus.

The existence of a probability implication I regard, in general, as a synthetic statement that cannot be proved by the calculus. The calculus can only transfer the existence character; with its help we can infer, from the known existence of certain probability implications, the existence of new ones. The property of transference by the calculus is, in part, directly expressed by the form of the axioms; some of the axioms, such as III and IV, directly assert the existence of new probability implications if certain others are given. However, these particular cases of transference do not suffice; for the transfer property will be required in a more general manner, as will be seen later. We must be able to assert that whenever the numerical value of a probability implication is determined by given probability implications, this probability implication does exist. It will become obvious (§ 17) that this existence is not self-evident, but must be asserted separately. The following postulate is therefore introduced.

RULE OF EXISTENCE. *If the numerical value p of a probability implication* $(A \underset{p}{\Rightarrow} B)$, *provided the probability implication exists, is determined by given probability implications according to the rules of the calculus, then this probability implication* $(A \underset{p}{\Rightarrow} B)$ *exists.*

The rule of existence is not an axiom of the calculus; it is a rule formulated in the metalanguage, analogous to the rule of inference or the rule of substitution (see § 5). It must be given an interpretation even in the formal treatment of the calculus. There must exist a formula that can be demonstrated in the calculus and that expresses the probability under consideration as a mathematical function of the given probabilities, with the qualification that the function be unique and free from singularities for the numerical values used. This is what is meant by the expression, "determined according to the rules of the calculus".[1]

§ 12. The Axioms of Univocality and of Normalization

From the discussion of the logical form we turn to the formulation of the laws of the probability implication. As explained above, an interpretation of probability is not required for this purpose. The laws will be formulated as a system of axioms for the probability implication—that is, as a system of logical formulas that, apart from logical symbols, contains only the symbol of the probability implication. Among the logical symbols, the logical implication occurs, and is thus used in formulating the laws of the probability implication.

The system to be constructed is called the system of axioms of the probability calculus. The name is justified by the fact that it is possible to derive from these axioms the formulas that are actually used in all applications of the probability calculus. When, at a later stage, an interpretation of probability is presented by means of statements about statistical frequencies, it will be possible to give another foundation to the axioms, showing that they are derivable from the given interpretation of probability. For the present, however, no use is made of the connection between probabilities and frequencies; instead, the axiom system is regarded as a system of formulas by which the properties of the probability concept are determined. By this procedure the axiomatic system of the probability calculus assumes a function comparable to that of the axiomatic system of geometry, which, in a similar way, determines *implicitly* the properties of the basic concepts of geometry, that is, of the concepts "point", "line", "plane", and so on (see § 8).

[1] The rule of existence can be replaced only incompletely by axioms. See footnote, p. 61.

We begin with the first two groups of axioms:

I. UNIVOCALITY $(p \neq q) \supset [(A \underset{p}{\rightarrow} B).(A \underset{q}{\rightarrow} B) \equiv \overline{(\bar{A})}]$

II. NORMALIZATION 1. $(A \supset B) \supset (\exists p)\, (A \underset{p}{\rightarrow} B).(p = 1)$

2. $\overline{(\bar{A})}.(A \underset{p}{\rightarrow} B) \supset (p \geqq 0)$

Group II will be discussed first. The degree of probability is asserted by II,2 to be a positive number, including 0 as an extreme case. That p cannot be greater than 1 is not incorporated into the axioms because it will be derived as a theorem in § 13. The normalization to values in the interval from 0 to 1, end points included, is restricted to the case where the class A is not empty. The condition is expressed by the term $\overline{(\bar{A})}$, which means, according to the rule of translation (see p. 49), $\overline{(i)(\overline{x_i \in A})}$, or, what is the same, $(\exists i)(x_i \in A)$. The significance of this condition will be explained presently.

Axiom II,1 establishes a connection between the logical implication and the probability implication. Whenever a logical implication exists between A and B, there exists also a probability implication of the degree 1; the converse does not hold, however. It follows from a simple consideration that the reverse relation cannot be maintained. For the demonstration we use the formula corresponding to II,1:

$$(A \supset \bar{B}) \supset (\exists p)\, (A \underset{p}{\rightarrow} B).(p = 0) \tag{1}$$

the necessity of which seems clear, though the exact derivation will be given later.

Formula (1) states that whenever an impossibility exists, a probability implication of the degree 0 exists also. For this case it is easy to illustrate why the reverse condition cannot be required. For instance, if we prick a sheet of paper with a needle, the probability (at least for a mathematical idealization of the problem) of hitting a given point is equal to 0; nevertheless a certain point is hit each time. Thus the probability 0 does not entail impossibility. Consequently, in order to remain free of contradictions, we must assert that certainty does not follow from the probability 1. Rather, certainty and the probability 1 stand in the relation of a narrower to a more comprehensive concept; certainty is a special case of the probability 1 (see § 18).

The relation of the two concepts is thus made clear in a very simple manner; the mysterious conception, which is occasionally voiced, that certainty and the probability 1 are incomparable concepts is untenable. On the contrary, the relation between the logical and the probability implication as expressed by II,1 represents an important relation holding between the two concepts, which connects the logic of the probability implication with classical logic. At this point the axiom system of probability differs from that of geometry. The concepts "point", "line", "plane", and so on, occurring in geometry, are of a type different from that of logical concepts; for that reason they can never assume the meaning of logical concepts, even for a special case.

The formulation of the univocality axiom I is clarified by the preceding remarks on the connection of the logical and the probability implication. It is obvious that the univocality of the degree of probability must be demanded. At first sight we might try to formulate univocality by

$$\overline{(A \underset{p}{\rightarrow} B) \, . \, (A \underset{q}{\rightarrow} B) \, . \, (p \neq q)} \tag{2}$$

However, this formula leads to contradictions. They result from the fact that in II,1 the logical implication was considered to be a special case of the probability implication. Certain properties of the logical implication prevent the assertion of (2) with complete generality. This is due to an above-mentioned property of the logical adjunctive implication, according to which a false proposition implies any proposition. In logic this fact is expressed by the *reductio ad absurdum*

$$(A \supset B) \, . \, (A \supset \bar{B}) \equiv (\bar{A}) \tag{3}$$

Formula (3) is a generalization of (1*g*, § 4). It is proved by transforming the left side of (3) by means of (6*a*, § 4), applying (4*c*, § 4) and using (5*d* and 5*c*, § 4). Addition of the parentheses, meaning extension to an all-statement, is of course always permissible for tautologies. Logic thus admits an ambiguity of logical implication, but this case is restricted to the condition $(\bar{A})$. The ambiguity is transferred to the probability implication, since (3) with II,1 and (1) lead to the relation

$$(\bar{A}) \supset (\exists \, p) \, (\exists \, q) \, (A \underset{p}{\rightarrow} B) \, . \, (A \underset{q}{\rightarrow} B) \, . \, (p = 1) \, . \, (q = 0) \tag{4}$$

In case of $(\bar{A})$ being true, the right side of the formula is valid, in contradiction to (2). Instead of (2) we therefore write axiom I, which brings the ambiguity of the probability implication into a form analogous to the ambiguity of logical implication. The condition $p \neq q$ must be written in front of I, since the expressions in brackets, contrary to (3), do not show whether we are dealing with different probability degrees.

In order to clarify I, it may be remarked that this axiom has the same meaning as the following implications:

$$(A \underset{p}{\rightarrow} B) \, . \, (A \underset{q}{\rightarrow} B) \, . \, (p \neq q) \supset (\bar{A}) \tag{5}$$

$$(\bar{A}) \supset (A \underset{p}{\rightarrow} B) \, . \, (A \underset{q}{\rightarrow} B) \tag{6}$$

These two formulas result when formula (7*a*, § 4) is used to dissolve the equivalence in I into implications going in both directions. In this case the expression $(p \neq q)$ is dropped at the left side of (6); the condition is redundant because (6) holds likewise if the condition is not satisfied, that is, if $p = q$. From (6) is derived

$$(\bar{A}) \supset (A \underset{p}{\ni} B) \tag{7}$$

Since p can be chosen completely at random, the formula states that for the case $(\bar{A})$ any degree of probability may be asserted to hold between A and B. Formula (7) goes beyond (4) so far as it extends the ambiguity to any chosen degree of probability, including even values greater than 1 or smaller than 0.[1]

The ambiguity thus admitted is harmless because it applies only to the case in which the first sequence does not contain a single element x_i belonging to the class A. This follows because, according to the translation rule,

$$(\bar{A}) =_{Df} (i)(\overline{x_i \epsilon A}) \tag{8}$$

In the case $(\bar{A})$, therefore, the probability cannot be used to determine expectations of the events B because the event A is never realized, and so the plurality of values cannot lead to practical inconveniences. It seems reasonable, in such a case, to consider the probability implication between A and B with respect to the sequence pair $x_i y_i$ as not defined at all and, therefore, to allow the assertion of any value for the degree of probability. This generalization of the probability concept extends it beyond practical needs; the extension is required because we wish to incorporate in the probability concept—as a special case—the logical implication as it is formulated in symbolic logic. The univocality, however, is always guaranteed if at least a single element x_i of the sequence belongs to the class A; it does not matter whether the corresponding y_i belongs to the class B. For, using the tautological equivalence provided by the propositional calculus,

$$a.b \supset c \equiv \overline{a.b} \vee c \equiv \bar{a} \vee \bar{b} \vee c \equiv \bar{a} \vee c \vee \bar{b} \equiv \overline{a.\bar{c}} \vee \bar{b} \equiv a.\bar{c} \supset \bar{b} \tag{9}$$

and substituting

$$\begin{aligned} &\text{for } a\text{: } (A \underset{p}{\ni} B).(A \underset{q}{\ni} B) \\ &\text{for } b\text{: } (p \neq q) \\ &\text{for } c\text{: } (\bar{A}) \end{aligned} \tag{10}$$

we derive from (5) the formula

$$(A \underset{p}{\ni} B).(A \underset{q}{\ni} B).\overline{(\bar{A})} \supset (p = q) \tag{11}$$

---

[1] The latter extension is necessary because otherwise the system of axioms would lead to contradictions, as J. C. C. McKinsey and S. C. Kleene have pointed out. See my note on probability implication in *Bull. Amer. Math. Soc.*, Vol. 47, No. 4 (1941), p. 265. It is for this reason that in this article I introduced for axiom II,2 the condition $(\bar{A})$, which the German edition of this book does not contain.

When the double negation is removed and the translation rule (p. 49) and formula (13, § 6) are applied, we obtain

$$(A \underset{p}{\rightarrow} B).(A \underset{q}{\rightarrow} B).(\exists i)\ (x_i \epsilon A) \supset (p = q) \qquad (12)$$

This means that the univocality of the degree of probability is guaranteed if there is at least one element x_i that belongs to the class A.

It is a result of the axioms I and II that the probability implication assumes the function of an extension of logical implication, the general implication introduced in (3, § 6). The latter is to be regarded as a special case of a probability implication, as we may recognize particularly from the form (6, § 9). This conception permits a more precise formulation of the concept of physical law, which was interpreted above as a general implication (§ 6). Closer inspection reveals that general implications that are absolutely certain can occur only if they are tautologies. The uncertainty of synthetic implications originates from the fact that any conceptual formulation of a physical event represents an idealization; the application of the idealized concept can possess only the character of probability (p. 8). The expression, "It follows according to a physical law", must therefore be represented, strictly speaking, not by a general implication but by a probability implication of a high degree (see § 85). Upon this fact rests the great importance of the probability implication: all laws of nature are probability implications.

There is an important difference between logical implication and probability implication. To the general implication $(A \supset B)$ corresponds an individual implication $a \supset b$, as defined by the truth tables 1B (§ 4). For probability implication such an individual relation is not used; the expression $A \underset{p}{\rightarrow} B$, therefore, need not be considered as a meaningful expression. Only in a fictitious sense can the degree of probability, holding for the entire sequence, be transferred to the individual case. Like the meaning of an individual connective implication of the synthetic kind (see § 6), that of an individual probability implication is constructed by a *transfer of meaning from the general to the particular case.* This transfer makes understandable why a frequency interpretation of the degree of probability can be applied to single events, though only in a fictitious sense. The problem will be considered later (see § 72).

§ 13. The Theorem of Addition

A well-known theorem of the probability calculus is that the probability of a logical sum is determined by the arithmetical sum of the individual probabilities, provided the events are mutually exclusive. For instance, the probability of obtaining face 1 or 2 by throwing a die is calculated to be $\frac{1}{6}+\frac{1}{6}=\frac{2}{6}$. For the addition it is essential that only one of the two faces can lie on top;

otherwise this manner of calculating would be unjustified. The theorem is usually called the *theorem of addition*, and it must now be formulated as an axiom.

The condition of exclusion could be written in the form $(B \supset \bar{C})$, but it is sufficient to use the weaker statement

$$(A \,.\, B \supset \bar{C}) \tag{1a}$$

which can be derived from $(B \supset \bar{C})$, whereas the latter formula is not derivable from (1*a*). Although (1*a*) appears to be nonsymmetrical with respect to B and C, this is actually not so; for, because of (6*a* and 5*a*, § 4), formula (1*a*) is equivalent to

$$(A \,.\, C \supset \bar{B}) \tag{1b}$$

By the use of (1*a*) the theorem of addition may be written as follows:

III. THEOREM OF ADDITION

$$(A \underset{p}{\Rightarrow} B) \,.\, (A \underset{q}{\Rightarrow} C) \,.\, (A \,.\, B \supset \bar{C}) \supset (\exists r)\, (A \underset{r}{\Rightarrow} B \vee C) \,.\, (r = p + q)$$

The addition theorem is a formula that expresses the transfer property of the calculus: it states a rule according to which the character of existence is transferred. It asserts the existence of the probability implication for the logical sum, if the individual probability implications are given. Nonetheless, we recognize the indispensability of the rule of existence (§ 11). For it is the existence rule that permits us to reverse the addition theorem; with its help we can derive the theorem

$$(A \underset{p}{\Rightarrow} B) \,.\, (A \underset{r}{\Rightarrow} B \vee C) \,.\, (A \,.\, B \supset \bar{C}) \supset (\exists q)\, (A \underset{q}{\Rightarrow} C) \,.\, (q = r - p) \tag{2}$$

This theorem cannot be obtained from axiom III alone, since the latter asserts existence only if the individual probabilities are given. The implicans of (2) differs from that of the axiom in that it contains only one individual probability and, moreover, the probability of the logical sum. Yet we recognize that the degree q of the probability implication, stated on the right side of (2), is determined by the addition theorem, provided this probability implication exists. Because of the univocality axiom I, the probability q, if it exists, must assume a value that, when added to p, furnishes the value r, that is, $q = r - p$. Now we can apply the existence rule, and the existence of the probability implication $(A \underset{q}{\Rightarrow} C)$ can be asserted.

The form of the relation (2) makes it clear that axiom III can be only partially reversed. The existence of the probability of the logical sum is not sufficient for the reversal; one of the two individual probabilities must also be given. Otherwise the degree of probability, q, would be undetermined, and the existence rule would not be applicable. The restricting condition is necessary because otherwise it would be possible to infer quite generally $(A \underset{q}{\ni} C)$, that is, the existence of a probability implication for any event. The *tertium non datur* (1e, § 4) and the formula $(\exists r)\,(A \underset{r}{\ni} C \vee \bar{C}).(r = 1)$, which is obtained from it by the help of (8c, § 4) and axiom II,1, would give this result. The unwarranted generalization is made impossible by the existence rule, which demands that the probabilities under consideration be determined by those given.

The idea expressed in (2) is of great importance in the logical construction of the probability calculus. It is the validity of reversed formulas like theorem (2) and thus of the existence rule upon which rests the possibility of operating with numerical values of probabilities according to the rules of algebra. When we no longer incorporate the condition of exclusion into the formula, stating it only in the context, we may write, introducing the *P*-notation,

$$P(A,B \vee C) = P(A,B) + P(A,C) \tag{3}$$

With this way of writing we express the fact that the rules by which mathematical equations are manipulated can be applied to probability formulas. Thus it is permissible to proceed from (3) to the formula

$$P(A,C) = P(A,B \vee C) - P(A,B) \tag{4}$$

The admissibility of this step is expressed in theorem (2). We recognize that the mathematical symbolization of the probability calculus is made possible by a particular property of this calculus, a property that requires a special formulation. The property is expressed by the rule of existence in combination with the axiom of univocality.

Certain difficulties arise from the fact that we cannot incorporate into the mathematical symbolization the condition of exclusion, presupposed for (3) and (4), but must add it verbally. A formula that is not dependent on conditions to be added in the context will be developed later (see § 20).

A remark must be made concerning the univocality of the *P*-symbol. Since univocality of a probability $P(A,B)$ is restricted to the case that A is not empty, the *P*-symbol has only in this case the character of a *numerical functor*, a number variable determined by the argument in parentheses. In order to make equations like (3) hold also in the case of an empty class A, the convention is introduced that such equations then represent *existential statements* of the form, "There is a numerical value for the dependent probability that satisfies the equation when the independent probabilities are given". For instance, (3) states for an empty class A that, if for $P(A,B)$ and $P(A,C)$ any values are given, there is a probability value among those holding for $P(A,B \vee C)$ that satisfies (3). All equations, in this case, will represent trivial statements, because, if A is empty, a probability with A in the first term will have all real numbers as its values; the existential statement will therefore be trivially satisfied. The advantage of this convention is that it allows us to drop, for probability equations, the condition stating that A is not empty. The equations also hold in the contrary case, but then they say nothing. For the implicational mode of writing, no such convention is needed, since axiom III and formula (2) are existential statements and lead to univocal values of the probabilities only if A is not empty. The convention as to the P-symbol is therefore in agreement with the rule of translation (p. 49).

In the greater part of this book the mathematical notation will be employed. Except in this section and the next, the axioms formulated in the implicational notation will no longer be used as a basis for further derivations. Their place will be taken by theorems in the P-notation, derived from them. The transition to the P-notation restricts the logical operations to the inner part of the P-symbols. Supplementary remarks will be made in the context whenever other restricting conditions, on which the validity of the formulas depends, are added.

We now derive a few theorems that have been used in the preceding section. Because of the *tertium non datur*, the formula $(A \supset B \vee \bar{B})$ is always true, and we obtain the general formula

$$(\exists\, r)\ (A \underset{r}{\Rightarrow} B \vee \bar{B}) . (r = 1) \tag{5}$$

or, in the P-notation.

$$P(A,B \vee \bar{B}) = 1 \tag{5'}$$

We may therefore add formula (5) to $(A \underset{p}{\Rightarrow} B)$. The conditions of theorem (2) are satisfied if we substitute $\bar{B}$ for C, since $(A . B \supset \bar{\bar{B}})$ also is always valid. We thus obtain the theorem

$$(A \underset{p}{\Rightarrow} B) \supset (\exists\, u)\ (A \underset{u}{\Rightarrow} \bar{B}) . (u = 1 - p) \tag{6}$$

In the P-notation the theorem is written

$$P(A,B) + P(A,\bar{B}) = 1 \tag{7}$$

This formula is called the *rule of the complement.*

We can now demonstrate that the probability degree, for which we postulated in II,2 only the nonnegative character, can never become greater than 1. We can complement the term B by its negation to constitute a complete disjunction. Considering the fact expressed in II,2 that both probabilities occurring in (7) cannot be negative, we obtain from (7) the relation

$$0 \leqq P(A,B) \leqq 1 \tag{8}$$

Furthermore, we have from II,1 and (6) the theorem

$$(A \supset \bar{B}) \supset (\exists\, p)\ (A \underset{p}{\Rightarrow} B) . (p = 0) \tag{9}$$

The mathematical symbolization of the calculus of probability may be illustrated by another problem. Given the three classes B_1, B_2, B_3, which are mutually exclusive but do not form a complete disjunction, and given the three probabilities

$$P(A,B_1 \vee B_2) \quad P(A,B_2 \vee B_3) \quad P(A,B_3 \vee B_1) \tag{10}$$

we wish to infer from them the existence of the three individual probabilities

$$P(A,B_1) \quad P(A,B_2) \quad P(A,B_3) \tag{11}$$

Theorem (2) is not applicable, because none of the individual probabilities is known to exist. However, we obtain from the addition theorem the equations

$$\begin{aligned} P(A,B_1) + P(A,B_2) &= P(A,B_1 \vee B_2) \\ P(A,B_2) + P(A,B_3) &= P(A,B_2 \vee B_3) \\ P(A,B_3) + P(A,B_1) &= P(A,B_3 \vee B_1) \end{aligned} \tag{12}$$

They can be solved for the individual probabilities:

$$\begin{aligned} P(A,B_1) &= \tfrac{1}{2}[P(A,B_1 \vee B_2) + P(A,B_3 \vee B_1) - P(A,B_2 \vee B_3)] \\ P(A,B_2) &= \tfrac{1}{2}[P(A,B_1 \vee B_2) + P(A,B_2 \vee B_3) - P(A,B_3 \vee B_1)] \\ P(A,B_3) &= \tfrac{1}{2}[P(A,B_3 \vee B_1) + P(A,B_2 \vee B_3) - P(A,B_1 \vee B_2)] \end{aligned} \tag{13}$$

The three individual probabilities (11) are therefore determined according to (13) by the or-probabilities (10); and it follows from the rule of existence that when (10) is given, the existence of (11) is also assertable. Owing to the rule of existence, we can apply, in the calculus of probabilities, the procedure of eliminating unknown quantities from a system of equations and use it to find new existing probabilities. Probability equations, therefore, *determine existence*, that is, the existence of any of the probabilities occurring in an equation is secured if all the other probabilities are known to exist.[1]

§ 14. The Theorem of Multiplication

The fourth and last group refers to an axiom that determines the probability of a combination of terms. It is a well-known theorem of the probability calculus that the probability of a combination—that is, the probability of a

[1] I am indebted to E. Tornier for having called my attention to the fact that the problem formulated in (10) and (11) cannot be solved by means of the formulas given in my paper on probability published in *Math. Zs.*, Vol. 34 (1932), p. 568. In that article I did not use the existence rule, but gave special *reversal axioms* that permitted the derivation of such theorems as (2) and, thereby, the application of the calculus of algebraic equations. But it turned out that, in this system, the existence-determining character is not always conserved when variables are eliminated. Equations (12) determine existence for my former system also, but equations (13) do not have this property. This fact led me to replace the reversal axioms by the rule of existence.

logical product—is represented by the arithmetical product of certain individual probabilities. This is the *multiplication theorem* of the probability calculus. The theorem is formulated by the following axiom:

IV. THEOREM OF MULTIPLICATION

$$(A \underset{p}{\ni} B).(A.B \underset{u}{\ni} C) \supset (\exists w)\ (A \underset{w}{\ni} B.C).(w = p \cdot u)$$

For the first time we deal with probability expressions in which the probability implication refers to three different classes, two of them occurring either in the first or in the second term. This does not cause any difficulty, because the translation rule (p. 49) determines the transition to the detailed notation for formulas of this kind also. In this case the domain of the probability implication is a triplet of sequences.

By a procedure of the kind used for the theorem of addition we can derive the converse of the multiplication theorem. We obtain two different conversions, since the three events A,B,C do not occur symmetrically in IV, whereas III is symmetrical with respect to B and C:

$$(A \underset{p}{\ni} B).(A \underset{w}{\ni} B.C) \supset (\exists u)\ (A.B \underset{u}{\ni} C).\left(u = \frac{w}{p}\right) \tag{1}$$

$$(A.B \underset{u}{\ni} C).(A \underset{w}{\ni} B.C) \supset (\exists p)\ (A \underset{p}{\ni} B).\left(p = \frac{w}{u}\right) \tag{2}$$

The proof of the theorems is based on the rule of existence, which applies because it can be demonstrated that the probability implications occurring on the right in (1) and (2) are determined by those on the left. Because of theorems (1) and (2), axiom IV can be replaced by the more comprehensive formula, written in the P-notation,

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

Theorems (1) and (2) mean that formula (3) can be solved according to the rules for mathematical equations for each of the individual probabilities occurring. Here again it is seen that the mathematical formalization of the probability calculus depends on the validity of the existence rule, as explained in §13.

Formula (3) is always true and does not require any restricting condition to be added verbally in the context, as was necessary for (3, § 13). Formula (3) will therefore be used in further discussion of the theorem of multiplication, without going back to axiom IV. The form selected here for theorem (3), characterized by the occurrence of three classes and of a term having two classes in the place of the reference class, has long been applied in the British and the American literature.[1] It has been used in the axiomatic construction in this work because only in this form is the axiom always correct. The probability from A to the logical product $B.C$ can be calculated only if the probability from A to B as well as that from $A.B$ to C is given.

In mathematical presentations the probability $P(A.B,C)$ is usually called "the relative probability of C with respect to B". This notation does not seem advisable because all probabilities are relative, and, furthermore, because the probability under consideration cannot be characterized by B and C alone but requires class A also.

For example, the probability that a person suffering from diphtheria subsequently contracts nephritis and dies is represented by a probability of the form $P(A,B.C)$, A denoting diphtheria; B, nephritis; and C, death. The probability is calculated as the product of the probability that a person suffering from diphtheria contracts nephritis, and the probability that a person dies who gets nephritis after having had diphtheria. The latter probability is different from the one that a person suffering from nephritis will die, since a patient who has had diphtheria is weakened and therefore is in greater peril of losing his life. This consideration shows why the last probability occurring in (3) must be characterized by three classes.

Another example is the probability that a thunderstorm follows a hot summer day with a subsequent change in the weather, which splits up into the product of two probabilities: the probability that a thunderstorm will follow a hot day and the probability that a change in the weather will follow a thunderstorm that was preceded by a hot day. The second probability is smaller than the probability that any thunderstorm brings with it a change in the weather, because the *convective* thunderstorms produced by local heat conditions usually do not result in a change in the weather, in contradistinction to *frontal* thunderstorms. The example illustrates once more the necessity of characterizing by three classes the probability that occurs in the last term of (3).

It must be regarded as a special case if two classes suffice for this term—a case arising when the actual three-class probability is equal to a certain two-class probability. Such specialization results if

$$P(A.B,C) = P(A,C) \tag{4}$$

Then (3) assumes the form of the *special theorem of multiplication*:

$$P(A,B.C) = P(A,B) \cdot P(A,C) \tag{5}$$

[1] In 1878 the form was used by C. S. Peirce. See his *Collected Papers* (Cambridge, Mass., 1932), Vol. II, p. 415. J. M. Keynes also employed the form in *A Treatise on Probability* (London, 1921), chap. XI, p. 6. The use of relative probabilities for the determination of dependent events is, of course, much older. P. S. Laplace gives a corresponding rule in his *Essai philosophique sur les probabilités* (Paris, 1814), chapter on "Principes généraux, quatrième principe." But he uses only two classes, my classes B and C, suppressing the general reference class A.

The condition (4) is paraphrased by the statement: *the events B and C are mutually independent with respect to A* (see also § 23). For example, the probability that a sudden gust of wind will capsize two sailboats is obtained as the product of the probability that the wind overturns one boat by the corresponding probability concerning the other boat. The two probabilities need not be the same, since the two sailboats may be of different construction. It is, however, necessary for (5) that the probability of the second boat's turning over be independent of whether the first boat turns over.

Another specialization of (3) is obtained if A can be represented as the product of two events A_1 and A_2 such that

$$P(A_1.A_2,B) = P(A_1,B) \qquad P(A_1.A_2.B,C) = P(A_2.B,C) \tag{6}$$

In this case (3) leads to

$$P(A_1.A_2,B.C) = P(A_1,B) \cdot P(A_2.B,C) \tag{7}$$

If we add the specialization analogous to (4)

$$P(A_2.B,C) = P(A_2,C) \tag{8}$$

we obtain

$$P(A_1.A_2,B.C) = P(A_1,B) \cdot P(A_2,C) \tag{9}$$

This case may be illustrated by the throwing of two dice: A_1 refers to the throwing of one die and A_2 to the throwing of the other. However, (9) would not be permissible without the conditions (6) and (8).

A third specialization results if

$$P(A.B,C) = P(B,C) \tag{10}$$

Then (3) becomes

$$P(A,B.C) = P(A,B) \cdot P(B,C) \tag{11}$$

Examples of this kind occur in certain causal chains: A may be represented by the occurrence of a storm; B, the falling of a tree; C, an accident caused by the falling tree. For the application of (11), however, we must inquire in each case whether (10) is satisfied.

The preceding discussion reveals that specializations of the multiplication theorem—some of which are used as axioms in representations of the probability calculus—do not provide formulas that are always true. They result from the general form (3) only for special cases. The latter are characterized by the equality of certain probabilities having different references classes, as stated in (4), (6), (8), (10). It follows that the question whether one of the special forms of the multiplication theorem can be applied is reduced to a question of the same type as that of how to determine the numerical value of a probability. It is always known whether two probabilities are equal when the probabilities themselves are known. Using the general form (3), or the form of axiom IV, for the theorem of multiplication eliminates certain logical difficulties that were connected with this theorem in the history of the calculus of probability.

§ 15. Reduction of the Multiplication Theorem to a Weaker Axiom