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

The theorem of multiplication is not independent of the other axioms; it can be reduced to a weaker assumption. In order to show this dependence I shall make use of the fact that the multiplication theorem can be split into two separate assertions. The first partial assertion states that the probability $P(A,B.C)$ *is determined* by $P(A,B)$ and by $P(A.B,C)$; the second assertion is that $P(A,B.C)$ is obtained, in particular, by the *arithmetical multiplication* of the two probabilities. The second assertion need not be stated explicitly as an axiom, but can be derived from the calculus with the use of the other axioms.

To prove this contention, multiplication theorem IV is replaced by the weaker axiom

IVa. $$(A \underset{p}{\Rightarrow} B).(A.B \underset{u}{\Rightarrow} C) \supset (\exists w)\,(A \underset{w}{\Rightarrow} B.C).[w = f(p,u)]$$

Here f stands for a mathematical function, temporarily undefined, that is to determine for any values p,u the corresponding w and, conversely, is required to be solvable unambiguously for p and u. Similarly to (1 and 2, § 14), it can be shown that the probability implication written at the right in these theorems assumes the degree of probability corresponding to the solution of $w = f(p,u)$ for p and u respectively; in these theorems the probability on the right side is replaced by

$$p = f'(w,u) \text{ and } u = f''(w,p), \text{ respectively,} \tag{1}$$

where f' and f'' represent the functions obtained by the solution. In this way it can be shown analogous to (3, § 14) that we may write

$$P(A,B.C) = f[P(A,B), P(A.B,C)] \tag{2}$$

The function f is the function occurring in IV*a*, and the comma between the probability symbols separates the two arguments of this function; that is, it serves as the comma between the arguments of a mathematical function.

In order to infer the form of f from (2), we substitute for C the disjunction of two mutually exclusive events C and D; then (2) becomes

$$P(A,B.[C \vee D]) = f[P(A,B), P(A.B,C \vee D)] \tag{3}$$

According to the first distributive law (4*a*, § 4), we dissolve

$$(B.[C \vee D] \equiv B.C \vee B.D) \tag{4}$$

and apply to both sides of equation (3) the addition theorem (3, § 13):

$$P(A,B.[C \vee D]) = P(A,B.C \vee B.D) = P(A,B.C) + P(A,B.D) \quad (5a)$$

$$P(A.B,C \vee D) = P(A.B,C) + P(A.B,D) \quad (5b)$$

The probabilities of the logical products occurring in (5a) are dissolved again according to (2):

$$P(A,B.C) = f[P(A,B), P(A.B,C)]$$

$$P(A,B.D) = f[P(A,B), P(A.B,D)] \quad (6)$$

Thus (3) is transformed into

$$f[P(A,B), P(A.B,C)] + f[P(A,B), P(A.B,D)]$$

$$= f[P(A,B), P(A.B,C) + P(A.B,D)] \quad (7)$$

Using the abbreviations

$$P(A,B) = p \qquad P(A.B,C) = u \qquad P(A.B,D) = v \quad (8)$$

we can write (7) as

$$f[p,u] + f[p,v] = f[p,u + v] \quad (9)$$

This is a functional equation for f; if it is to be valid for any values u and v the function f must have the form

$$f[p,u] = g(p) \cdot u \quad (10)$$

where $g(p)$ represents a function of p alone, which remains undetermined for the time being.[1]

In (2) we now substitute $[C \vee \bar{C}]$ for C; then (2) becomes

$$P(A,B.[C \vee \bar{C}]) = f[P(A,B),P(A.B,C \vee \bar{C})] \quad (11)$$

According to (5c, § 4), we have

$$(B.[C \vee \bar{C}] \equiv B) \quad (12)$$

and therefore

$$P(A,B.[C \vee \bar{C}]) = P(A,B) = p \qquad P(A.B,C \vee \bar{C}) = 1 \quad (13)$$

[1] I refer to a well-known theorem of mathematics. It may be proved as follows: we put $u = 0$; then we derive from (9) that $f(p,0) = 0$. Assuming v to be the differential increase du, we write (9):

$$f[p,0 + du] - f[p,0] = f[p,u + du] - f[p,u]$$

Dividing by du, we obtain for the limit $du = 0$ the differential equation

$$\left(\frac{\partial f[p,u]}{\partial u}\right)_0 = \left(\frac{\partial f[p,u]}{\partial u}\right)_u$$

The subscript marks the argument-place at which the differential quotient is to be formed. Since u can be chosen at random, the equation states the differential quotient for u to be constant; that is, the function f is linear with respect to u. It is even possible to drop the assumption that the function f is differentiable and continuous, but the proof will then be more complicated.

Using these results in combination with (10), we transform (11) into

$$p = f[p,1] = g(p) \cdot 1 = g(p) \tag{14}$$

With this determination of $g(p)$, the relation (10) assumes the form

$$f(p,u) = p \cdot u \tag{15}$$

Because of (2) and (8) this means

$$P(A,B.C) = P(A,B) \cdot P(A.B,C) \tag{16}$$

Thus we have proved the multiplication theorem (3, § 14).

It is seen from this demonstration that the theorem of multiplication represents a necessary formula within the frame of the calculus of probability. That the probability of the logical product is given by an arithmetical product is a consequence of the fact that the probability of a logical sum is given by an arithmetical sum, in combination with the first distributive law of logic.

The result enables us to introduce a new definition of the property of independence, defined in (4, § 14) or (5, § 14). Combining (4, § 14) with (2), we may define independence as follows.[2] Two events are independent with respect to A if the probability from A to their logical product is a function of their individual probabilities with respect to A alone, that is, if

$$P(A,B.C) = f[P(A,B),P(A,C)] \tag{17}$$

It then follows that f assumes the form of the arithmetical product. This characterization of independence is very instructive; it states that the probability of the combination of independent events is determined whenever the probabilities of the separate events are given. For instance, the probability $\frac{1}{6}$ for each of two dice determines the probability $\frac{1}{36}$ for the combination of any two faces.

§ 16. The Frequency Interpretation

Axioms I to IV suffice to derive all the theorems of the calculus in which probability sequences occur as wholes the structure of which is not considered. The totality of these theorems is called the *elementary calculus of probability.* With the given axioms we therefore control the *formal structure* of the elementary calculus of probability. But before developing the theorems of this calculus we wish to give the probability concept an interpretation over and above the characterization of its formal structure (see § 8).

This leads to a problem that has been under much discussion. The formal structure of the probability calculus that I have developed might be conceded

[2] I am indebted to Kurt Grelling for the suggestion that independence can be characterized in this manner; he thereby directed my attention to the foregoing proof for the product form of the function f.

by adherents of the most diverse theories about probability. But the question of the interpretation of the probability concept can be answered only on the basis of painstaking philosophical investigations, and different theories have answered it in different ways. It will be treated, therefore, in more detail later (see chap. 9).

The laws of the calculus of probability are difficult to understand, however, if one does not envisage a definite interpretation. Thus, for didactic reasons, an interpretation of the probability concept must be added, at this point, to the axiomatic construction. But this method will not prejudice later investigations of the problem. The interpretation is employed merely as a means of illustrating the system of formal laws of the probability concept, and it will always be possible to separate the conceptual system from the interpretation, because, for the derivation of theorems, the axioms will be used in the sense of merely formal statements, without reference to the interpretation.

This presentation follows a method applied in the teaching of geometry, where the conceptual formulation of geometrical axioms is always accompanied by spatial imagery. Although logical precision requires that the premises of the inferences be restricted to the meaning given in the conceptual formulation, the interpretation is used as a parallel meaning in order to make the conceptual part easier to understand. The method of teaching thus follows the historical path of the development of geometry, since, historically speaking, the separation of the conceptual system of geometry from its interpretation is a later discovery. The history of the calculus of probability has followed a similar path. The mathematicians who developed the laws of this calculus in the seventeenth and eighteenth centuries always had in mind an interpretation of probability, usually the frequency interpretation, though it was sometimes accompanied by other interpretations.

In order to develop the frequency interpretation, we define probability as the *limit of a frequency* within an infinite sequence. The definition follows a path that was pointed out by S. D. Poisson[1] in 1837. In 1854 it was used by George Boole,[2] and in recent times it was brought to the fore by Richard von Mises,[3] who defended it successfully against critical objections.

The following notation will be used for the formulation of the frequency interpretation. In order to secure sufficient generality for the definition, we shall not yet assume that all elements x_i of the sequence belong to the class A. We assume, therefore, that the sequence is *interspersed* with elements x_i of a different kind. For instance, the sequence of throws of a coin may be interspersed with throws of a second coin. In this case only certain elements x_i

[1] *Recherches sur la probabilité des jugements en matière criminelle et en matière civile . . .* (Paris, 1837).

[2] *The Laws of Thought* (London, 1854), p. 295.

[3] "Grundlagen der Wahrscheinlichkeitsrechnung," in *Math. Zs.*, Vol. V (1919), p. 52, and later publications.

will belong to the class A, if the class is defined as representing the throws of one of the coins only. Similarly, only some among the elements y_i will belong to the class B, which may signify the occurrence of tails lying up. It may happen that y_i represents a case of tails up, whereas the corresponding x_i does not belong to the class A, that is, the event of tails lying up is produced by the second coin. When the frequency is counted out in such a sequence pair, the result is expressed by the symbol

$$\overset{n}{\underset{i=1}{N}} (x_i \epsilon A) \tag{1a}$$

which means the number of such x_i between 1 and n that satisfy $x_i \epsilon A$. The symbol is extended correspondingly to apply to different variables and to different classes and also to a pair, a triplet, and so on, of variables. For instance, the expression

$$\overset{n}{\underset{i=1}{N}} (x_i \epsilon A).(y_i \epsilon B) \tag{1b}$$

represents the number of pairs x_i,y_i such that x_i belongs to A and simultaneously y_i belongs to B; it signifies the number of pairs x_i,y_i that are elements of the common class A and B. To abbreviate the notation, the following symbol is introduced:

$$N^n(A) =_{Df} \overset{n}{\underset{i=1}{N}} (x_i \epsilon A) \qquad N^n(A.B) =_{Df} \overset{n}{\underset{i=1}{N}} (x_i \epsilon A).(y_i \epsilon B) \tag{2}$$

Furthermore, the *relative frequency* $F^n(A,B)$ is defined by

$$F^n(A,B) = \frac{N^n(A.B)}{N^n(A)} \tag{3}$$

In the special case in which all elements x_i belong to the class A, that is, when the sequence x_i is *compact*, the denominator of the fraction is equal to n, whereas in the numerator the expression A may be dropped; then (3) assumes the simpler form

$$F^n(A,B) = \frac{1}{n} \cdot N^n(B) \tag{4}$$

With the help of the concept of relative frequency, the frequency interpretation of the concept of probability may be formulated:

If for a sequence pair x_iy_i the relative frequency $F^n(A,B)$ goes toward a limit p for $n \to \infty$, the limit p is called the probability from A to B within the sequence pair. In other words, the following coördinative definition is introduced:

$$P(A,B) = \lim_{n \to \infty} F^n(A,B) \tag{5}$$

No further statement is required concerning the properties of probability sequences. In particular, randomness (see § 30) need not be postulated.

§ 17. The Origin of Probability Statements

So long as we regard the probability calculus as a formal calculus by means of which formulas are manipulated, that is, so long as we do not speak of the meaning of the formulas, the origin of probability statements presents no problem. The question whether the individual probability statement is true or false, then, is not a problem of the calculus, as was explained above. The calculus deals solely with transformations of probability statements; and the statements of the mathematical calculus, therefore, represent exclusively tautological implications of the type, "If certain probability implications $a_1, \ldots a_n$ exist, then certain other probability implications $b_1, \ldots b_n$ exist also". I agree here with a conception emphasized by von Mises.

But it would be a shortsighted attitude if mathematicians were induced by this conception to regard the question of the origin of probability statements as unreasonable. With the given definition of the probability calculus, the question is merely shifted to another field. At the very moment at which an interpretation is assigned to the probability statement, there arises the question how to know whether, in a given instance, a probability statement holds. It follows from the nature of the interpretation that the question is equivalent to the question how to ascertain the existence of a limit of an infinite sequence.

Here an important distinction must be made. First, probability sequences may be regarded as mathematically given sequences, that is, as sequences that are defined by a rule. For instance, a probability sequence can be defined by means of an infinite decimal fraction in which every even number is regarded as the case B and every odd number as the case $\bar{B}$. Whether such a sequence has a frequency limit and what the limit is, is a question of purely mathematical nature to be answered by means of the usual mathematical methods. It is important that we have at our disposal such mathematically given sequences representing the frequency interpretation; on occasion they will be used as models (see §§ 30 and 66). In the practical application of the probability calculus, however, they do not play a part.

Second, sequences provided by events in nature may be considered. For such sequences, which include all practical applications of the calculus of probability, we do not know a definite law regarding the succession of their elements. Instead of a defining rule, we have a finite initial section of the sequence; therefore we cannot know, strictly speaking, toward what limit such a sequence will proceed. We assume, however, that the observed frequency will persist, within certain limits of exactness, for the infinite rest of the sequence. This inference, which is called *inductive inference*, leads to very difficult logical problems; and it will be one of the most important problems of this investigation to find a satisfactory explanation of the inference. For the present, however, the inference will not be questioned. Suffice it to say that the inference is actually used—sometimes under the name *a posteriori determination of a probability*—by statisticians as well as in everyday life. We shall therefore use it, too, in problems of the application of the formulas constructed.

It may sometimes be expedient, for mathematical reasons, to imagine a fictitious observer who can count out an infinite sequence and thus is able to determine its limit. But the picture serves only to illustrate certain logical relations and cannot replace the inductive inference where physical reality is concerned.

To summarize: for the present we shall regard as verifiable an assertion stating that there exists a probability sequence of a determinate degree of probability. The verification may be derived either mathematically, from the defining rule of the sequence, or by means of an inductive inference.

The given interpretation will now be used to elucidate some properties of the axiom system that so far, perhaps, have not been made sufficiently clear. First, we realize why the existence of an indeterminate probability implication has been regarded as a synthetic statement requiring empirical proof. The assertion that there exists a limit of the frequency, even without specification of the degree, represents a definite statement that is certainly not satisfied for every sequence pair $x_i y_i$. For this reason the rule of existence is necessary within our formal system; when interpreted, it expresses the assertion that a limit of the frequency exists in the cases concerned.

Second, we recognize that the indeterminate probability implication $(A \twoheadrightarrow B)$ states more than the existence of a mere possibility relation, which we write as $(\overline{A \supset \bar{B}})$.[1] The added meaning consists in the fact that the first statement asserts a certain regularity in the repetition of events. When a die is thrown upon a table, it is possible that a sudden thunderbolt may happen simultaneously; but such a statement of possibility does not mean that a probability implication exists between the two events. I do not wish to say that the probability is very small; I mean, rather, that it is not permissible to assert a definite regularity with respect to the occurrence of thunder when the die is thrown repeatedly. The illustration will make it clear that the existence of a probability cannot be inferred from the possibility of an event. But neither does the converse hold. From (1, § 12) it is seen that the possibility of an event cannot be inferred from the existence of a probability. The probability can be equal to zero, and the probability zero may or may not represent impossibility. In neither direction does an implication hold between the two statements $(A \twoheadrightarrow B)$ and $(\overline{A \supset \bar{B}})$. Probability and possibility are disparate concepts, that is, their extensions overlap.

If we were to assert that a frequency limit must exist for any two repetitive events observed for a sufficiently long time, we would commit ourselves to a

[1] This is the extensional possibility of § 80.

far-reaching hypothesis. On this assumption it would be possible to drop the existence rule; but, instead, we should have to introduce into the calculus an axiom of the form, "For all A and C, $(A \ni C)$ is valid". Obviously this addition would mean an extraordinary extension of the content of the calculus, with which we do not wish to burden the axiom system.

I therefore consider the assertion of a determinate as well as of an indeterminate probability implication to be a synthetic statement, the validity of which can be ascertained, when physical events are concerned, by means of statistics in combination with inductive inferences. This method of ascertainment will not be questioned throughout the mathematical part of the investigation, because the frequency interpretation does not enter into the content of the probability calculus to be developed. It constitutes only an illustrative addition and will not be used for the derivation of theorems.

§ 18. Derivation of the Axioms from the Frequency Interpretation

It will now be shown that all axioms of the calculus of probability can be derived from the frequency interpretation, that is, they are tautologies if the frequency definition of probability is assumed.

We start with the univocality axiom I. The case $(\bar{A})$, to which this axiom refers, signifies that the relative frequency F^n assumes the indeterminate form $\frac{0}{0}$, since the summation N^n in (3, § 16) leads to 0 for numerator as well as denominator. Therefore we also have $P(A,B) = \frac{0}{0}$, that is, the probability does not possess a determinate value. This result represents one assertion of the axiom. If the case $(\bar{A})$ does not hold, however, a definite limit exists; since there can be only one limit, the other assertion of the axiom is likewise satisfied. Notice that a limit exists even when only a finite number of elements x_i belong to A; the value of the frequency for the last element is then regarded as the limit. This trivial case is included in the interpretation and does not create any difficulty in the fulfillment of this or the following axioms.

Axiom II,1 concerns the case in which each element of the form $(x_i \epsilon A)$ is followed by an element $(y_i \epsilon B)$, since this is what the logical implication asserts. In this case all $F^n = 1$, a result following immediately from (3, § 16), so that II,1 is satisfied. The major implication in the axiom can be directed toward only one side, since the probability 1 can be obtained, also, if there are some cases in which $x_i \epsilon A$ is followed by $y_i \epsilon \bar{B}$. These cases, however, must be distributed so sparsely that the limit F^n becomes equal to 1, though every individual F^n may be smaller than 1. An example is given by a compact sequence A accompanied by a sequence B that has a $\bar{B}$ in all elements whose subscript i is the square of a whole number but which has a B in all other elements. Thus the frequency interpretation makes it clear why the probability 1 represents a wider concept than the logical implication.

This consideration shows also that the probability implication of the degree p represents a generalization of the general implication of symbolic logic. Whereas the general implication demands all elements $x_i \epsilon A$ to be followed by a $y_i \epsilon B$, the probability implication includes the case in which certain $x_i \epsilon A$ are followed by a $y_i \epsilon \bar{B}$, with the qualification, however, that between the numbers of the elements there must exist a frequency ratio that goes in the limit toward a determinate value. The probability implication, itself representing a general implication, therefore constitutes the generalization of the usual general implication for sequences in which the individual implication occurs only in a certain number of places. Instead of demanding the individual implication to be valid without exceptions, we require only a frequency ratio.

That II,2 is satisfied follows directly from the fact that the relative frequency F^n is a positive number (including 0). The condition, expressed in (8, § 13), that the probability degree cannot be greater than 1 likewise follows from the definition of the relative frequency.

We turn now to the addition theorem III. In order to prove this axiom, we form first

$$F^n(A,B \vee C) = \frac{N^n(A\,.[B \vee C])}{N^n(A)} \tag{1a}$$

If $(A\,.B \supset \bar{C})$ is valid, this is equal to

$$\frac{N^n(A\,.B)}{N^n(A)} + \frac{N^n(A\,.C)}{N^n(A)} \tag{1b}$$

and we obtain

$$F^n(A,B \vee C) = F^n(A,B) + F^n(A,C) \tag{2}$$

The equation remains unchanged in the transition to the limit, and for mutually exclusive events we have

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

The exclusion condition suffices for the addition of probabilities having the same first term. We need not presuppose, in such a case, that the terms B and C belong to the same sequence; this represents a special case for which, of course, the theorem is also valid.

The given proof can be made clearer by the following consideration. We write the three sequences below one another, each in one row; however, we do not write the elements x_i, y_i, z_i, but only the classes A, B, C, to which the elements belong. For the sake of simplicity we shall assume that the sequence x_i consists only of the elements $x_i \epsilon A$ and thus is compact. We thereby arrive at the following arrangement:

$$\begin{array}{l} A\ A\ A\ A\ A\ A\ A\ A\ \ldots \\ B\ B\ \bar{B}\ B\ \bar{B}\ B\ B\ \bar{B}\ \ldots \\ \bar{C}\ \bar{C}\ C\ \bar{C}\ \bar{C}\ \bar{C}\ \bar{C}\ C\ \ldots \end{array} \tag{4}$$

The frequency $F^n(A,B \vee C)$ expresses the relative frequency of the A under which a B or a C is found. Because of the condition of exclusion, a B and a C can never stand simultaneously under the same A, and thus the relative frequencies of B and C add up to that of $B \vee C$.

The multiplication theorem IV, also, can be derived from the frequency interpretation. We obtain from (3, § 16)

$$F^n(A,B.C) = \frac{N^n(A.B.C)}{N^n(A)} = \frac{N^n(A.B)}{N^n(A)} \cdot \frac{N^n(A.B.C)}{N^n(A.B)}$$
$$= F^n(A,B) \cdot F^n(A.B,C) \tag{5}$$

The equation remains valid for the transition to the limit, if the individual limits exist, and we have with the use of (5, § 16)

$$P(A,B.C) = P(A,B) \cdot P(A.B,C) \tag{6}$$

We thus arrive at the general theorem of multiplication (3, § 14). We now see why this form, which we used for the theorem, is always valid. Only in this form does the multiplication theorem represent a tautology in the frequency interpretation.

This proof, too, may be illustrated by a schema as used above:

$$\begin{matrix} A & A & A & A & A & A & A & A & \dots \\ B & \bar{B} & \bar{B} & B & \bar{B} & B & \bar{B} & B & \dots \\ \underline{\underline{C}} & C & \bar{C} & \underline{\underline{C}} & C & \underline{\underline{\bar{C}}} & C & \underline{\underline{\bar{C}}} & \dots \end{matrix} \tag{7}$$

The frequency $F^n(A,B.C)$ represents the frequency of the couples $B.C$; the first of the expressions standing on the right side of (5), $F^n(A,B)$, counts the frequency of B. Now B selects from the sequence of C's a subsequence, the elements of which are marked by a lower double bar in (7); this subsequence, of course, contains elements C as well as $\bar{C}$. The number of elements of this subsequence is given by $N^n(A.B)$; therefore $F^n(A.B,C)$ means the relative frequency of C in the subsequence. The consideration is always applicable: if a term is added before the comma within a probability expression, the frequency is counted within the subsequence that is selected by this term. Formula (5) states that the desired frequency of the pair $B.C$ can be represented as the product of the frequency of B by the frequency of C counted within the subsequence selected by B.

These considerations lead to an instructive interpretation of the independence relation defined in (4, § 14). The definition

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

states that, within the subsequence selected by B from the C-sequence, C has the same relative frequency as in the main sequence. This characterization reveals the meaning of the independence relation; that B does not influence C means that a selection by B from the C-sequence does not change the relative frequency.[1] For instance, when we throw with two dice and consider, within the sequence produced by the second die, only the subsequence of throws in which the first die simultaneously gives the result of face 6 lying up, we shall find, too, the relative frequency $\frac{1}{6}$ for any face of the second die.

Finally, it remains to prove that the rule of existence is derivable from the frequency interpretation. Since each of the axioms represents a tautological relation between frequencies, which holds strictly even before the transition to the limit, every probability formula derivable from the axioms will correspond also to a tautological relation between frequencies; and this relation will be strictly valid before the transition to the limit. Every such relation can be written in the form

$$f_m^n = r(f_1^n \ldots f_{m-1}^n) \tag{9}$$

In this formula the f_i^n stand for frequency expressions of the form

$$f_i^n = F^n(A_i, B_i) \tag{10}$$

The subscripts in (9) and (10) indicate the fact that we are dealing here with frequency quantities that belong to different events $A, B \ldots$. According to the existence rule, r is a single-valued function, free from singularities at this place. Passing to the limit $n \to \infty$, we derive from the laws governing the formation of a limit that, whenever the $f_1^n \ldots f_{m-1}^n$ go toward limits $p_1 \ldots p_{m-1}$, the f_m^n also must approach a limit p_m. In other words, the probability p_m must exist whenever the probabilities $p_1 \ldots p_{m-1}$ exist. This is the assertion made by the rule of existence.

At the same time we recognize why the existence of a probability is bound by the condition that it be determined by given probabilities. Assume that it is unknown in (9) for two quantities, say, f_m^n and f_{m-1}^n, whether they go toward a limit. Then we cannot infer, from the fact that the other quantities $f_1^n \ldots f_{m-2}^n$ approach certain limits, that the two residual quantities f_m^n and f_{m-1}^n go toward a limit. For instance, if the probability of a logical sum is given, the sum f_3^n of the two frequencies

$$f_1^n + f_2^n = f_3^n$$

approaches a limit p_3. Yet the individual frequencies f_1^n and f_2^n need not go toward a limit. A convergence can be inferred only when it is known that, apart from f_3^n, at least one of the other quantities, say f_2^n, approaches a limit.

This concludes the proof that all the axioms of the probability calculus follow logically from the frequency interpretation. The result holds not only for infinite but also for finite sequences, provided that in this case we regard the limit of the frequency as given by the value of $F^n(A,B)$ taken for the last

[1] R. von Mises has made this idea the starting point of his probability theory. See § 30.

element. All the axioms are satisfied tautologically, and are strictly, not only approximately, valid even before the transition to the limit.

The given proof guarantees that the frequency interpretation is an admissible interpretation of the theorems derivable from the axiom system. The interpretation will be applied in the examples used to illustrate the derived formulas.

§ 19. The Rule of Elimination

We may now proceed to the derivation of individual theorems of the probability calculus from the axiom system.

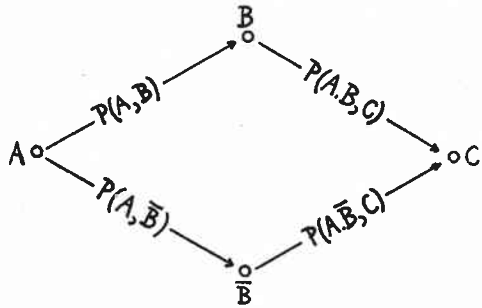

Fig. 4. Schema for rule of elimination, according to (2).

Many practical cases present the problem of calculating the probability from A to C, when C is linked to A by an intermediate term B and only the intermediary probabilities are given. Figure 4 may serve to illustrate the problem.

It represents the *divergent* probabilities $P(A,B)$ and $P(A,\bar{B})$, having the first term in common, and the *convergent* probabilities $P(A.B,C)$ and $P(A.\bar{B},C)$, which possess a common term after the comma. When the divergent and convergent probabilities are given, it is possible to calculate $P(A,C)$. For this purpose we use the logical equivalence

$$([B \vee \bar{B}].C \equiv C) \tag{1}$$

and thus obtain the relations

$$P(A,C) = P(A,[B \vee \bar{B}].C) = P(A,B.C \vee \bar{B}.C)$$
$$= P(A,B.C) + P(A,\bar{B}.C)$$

In the last equality the addition theorem has been applied because the terms are mutually exclusive. The use of the multiplication theorem gives the result

$$P(A,C) = P(A,B) \cdot P(A.B,C) + P(A,\bar{B}) \cdot P(A.\bar{B},C) \tag{2}$$

This formula is called the *rule of elimination.* It permits the elimination of a term B that is interpolated between the terms A and C, and the establishment of a direct probability from A to C. The rule of elimination performs with respect to probability implication the function that is performed for the logical implication by its transitivity (8l, § 4). But here the logical structure is much more complicated than it is for a transitivity. The elimination of B can be achieved, according to (2), only when $P(A.\bar{B},C)$ is known, apart from $P(A,B)$ and $P(A.B,C)$. The probability $P(A,\bar{B})$ is determined by $1-P(A,B)$, but $P(A.\bar{B},C)$ represents an independent probability that is not determined by the other quantities written at the right of (2). The convergent probabilities $P(A.B,C)$ and $P(A.\bar{B},C)$ will be called *nonbound probabilities,* since their sum can be greater or smaller than 1; the divergent probabilities $P(A,B)$ and $P(A,\bar{B})$ are *bound probabilities,* that is, they must add up to the value 1.

The theorem may be illustrated by an example previously used. Let A denote a hot summer day; B, the occurrence of a thunderstorm; C, a change in the weather. The probability of a change in weather occurring on a hot day can be calculated from the intermediary probability concerning the thunderstorm; but we must know the probability of the occurrence of a thunderstorm, the probability of a change in the weather on a hot day after a thunderstorm has occurred, and the probability of a change in the weather on a hot day on which no thunderstorm occurs.

In the frequency interpretation, (2) can easily be made clear: the number of C's to which a B is coördinated, and the number of C's to which a $\bar{B}$ is coördinated, add up to the total number of C's.

The rule of elimination contains some interesting special cases. First, we may have

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

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

Then (2) assumes the form

$$P(A,C) = P(A,B)\cdot P(B,C) + P(A,\bar{B})\cdot P(\bar{B},C) \tag{4}$$

We can illustrate this form by choosing for B and $\bar{B}$ two bowls that contain black and white balls in different ratios, and for A another bowl containing, say, numerous tickets on which is written B or $\bar{B}$. The ticket drawn from A decides whether the second draw should be made from B or $\bar{B}$. By C we understand the event of a black ball being obtained.

A further specialization results for

$$P(A.\bar{B},C) = 0 \tag{5}$$

We then have

$$P(A,C) = P(A,B)\cdot P(A.B,C) \tag{6}$$

If the specialization (3*a*) is added, we obtain

$$P(A,C) = P(A,B) \cdot P(B,C) \tag{7}$$

Only in this very specialized case does the rule of elimination assume the form of a transitivity, in which the degrees of probability are simply multiplied. The case may be illustrated by the example above, with the qualification that the bowl $\bar{B}$ does not contain any black balls. Other examples are given in causal chains: for instance, when A means the presence of a hot day in summer; B, the occurrence of a thunderstorm; C, a flash of lightning hitting a house. In the special case where $P(A,B) = 1$ and $P(B,C) = 1$, the relation (7) determines also $P(A,C) = 1$; here the condition (5) is no longer required, since the second term in (2) drops out because of $P(A,\bar{B}) = 0$.[1] These relations are satisfied for logical implications of the form $(A \supset B)$ and $(B \supset C)$. The relation (3*a*), too, must hold in this case because with $(B \supset C)$ we have also $(A . B \supset C)$. This is why the logical implication follows a general rule of transitivity that is not restricted by any conditions. It is seen, further, that the transitivity (7), in general, produces a decrease in the degree of probability. If the intermediary probabilities written at the right in (7) are smaller than 1, the total probability at the left in (7) will be smaller than any of the intermediary probabilities. A corresponding statement cannot be made for the general case (2); here $P(A,C)$ represents a certain mean value between the other probabilities.

A third specialization results by the assumption

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

Introducing this condition into (2) and using the relation $P(A,B) + P(A,\bar{B}) = 1$, we obtain

$$P(A,C) = P(A.B,C) = P(A.\bar{B},C) \tag{9}$$

Comparison with (4, § 14) shows that this means the independence of B and C with respect to A. In the frequency interpretation, (9) means that if the subsequences selected from the C-sequence by B and $\bar{B}$, respectively, contain C with equal relative frequencies, this frequency is the same as in the main sequence.

It has been pointed out that $P(A.\bar{B},C)$ is not determined by $P(A,B)$ and $P(A.B,C)$; but (2) states that a determination results if $P(A,C)$ is added. This connection is expressed by the solution of (2) for $P(A.\bar{B},C)$:

$$P(A.\bar{B},C) = \frac{P(A,C) - P(A,B) \cdot P(A.B,C)}{1 - P(A,B)} \tag{10}$$

[1] If it is known that $P(B,A) > 0$, even the condition (3*a*) can be omitted, because this condition then follows from $P(B,C) = 1$. See (6, § 25).

The relation shows how a probability containing a negation in the first term is calculated from the probabilities of nonnegative reference. We must except the case $P(A,B) = 1$, since in this case the value of (10) is indeterminate; this condition is also understood for the relations (11), (12), and (14), to be derived presently.

As before, some important special cases must be considered. We see that with

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

we also have

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

in correspondence to (9); that is, the converse of the relation leading from (8) to (9) is valid. Furthermore, we infer that, if $P(A.B,C) > P(A,C)$, we have $P(A.\bar{B},C) < P(A,C)$, and, similarly, if $P(A.B,C) < P(A,C)$, we have $P(A.\bar{B},C) > P(A,C)$. This result follows because for $P(A.B,C) = P(A,C)$ the relation (10) supplies $P(A.\bar{B},C) = P(A,C)$, and this value is diminished or increased according as $P(A.B,C)$ is larger or smaller than $P(A,C)$.

For mutually exclusive events B and C, that is, $P(A.B,C) = 0$, relation (10) assumes the simple form

$$P(A.\bar{B},C) = \frac{P(A,C)}{1 - P(A,B)} \tag{12}$$

Another special case arises for

$$P(A,B) = P(A,C) \tag{13}$$

Then (10) is transformed into

$$\frac{P(A.\bar{B},C)}{P(A.B,\bar{C})} = \frac{P(A,B)}{P(A,\bar{B})} = \frac{P(A,C)}{P(A,\bar{C})} \tag{14}$$

From (10) we can derive two important inequalities that restrict the choice of the probabilities to be given. Since $P(A.\bar{B},C)$ is bound by the normalization (8, § 13), the expression on the right side of (10) must lie between 0 and 1 (with inclusion of the limits). This leads to the two inequalities

$$1 - \frac{1 - P(A,C)}{P(A,B)} \leqq P(A.B,C) \leqq \frac{P(A,C)}{P(A,B)} \tag{15}$$

The inequality on the left side results from transformation of the condition that (10) must not be greater than 1; the inequality on the right side arises from a transformation of the condition that the numerator of (10) must not be smaller than 0. The double inequality is not necessarily satisfied for given values $P(A,B)$ and $P(A,C)$, even if $P(A.B,C)$ is chosen according to the normalization (8, § 13). The relation (15) formulates an additional condition, which prescribes a narrower domain for $P(A.B,C)$ whenever we have

$1 - P(A,C) < P(A,B)$ or $P(A,C) < P(A,B)$. It can be shown that for independent events B and C, that is, for $P(A.B,C) = P(A,C)$, (15) is always fulfilled.[2] It is permissible, therefore, to give two events as independent, regardless of the values of their probabilities. But if two events are given as dependent, the degree of dependence must be kept within the limits defined by (15). The occurrence of such inequalities in regard to the choice of probabilities may be compared to the occurrence of similar inequalities in geometry. A triangle, for instance, can be constructed from three given determinations only when their values satisfy certain numerical restrictions. Notice that the inequalities (15) hold also for the case $P(A,B) = 1$, which had to be excepted for (10), since in this case the numerator of (10) must be $= 0$ in order to make possible a finite value of $P(A.\bar{B},C)$, and thus the conditions leading to (15) are satisfied. For mutually exclusive events B and C, that is, $P(A.B,C) = 0$, (15) leads to the trivial condition $P(A,B) + P(A,C) \leqq 1$.

We turn now to an extension of the rule of elimination to disjunctions of more than two terms. There are special kinds of such *many-term disjunctions* $B_1 \vee \ldots \vee B_r$ that play a particularly important role in the calculus of probabilities: disjunctions that are both *complete* and *exclusive*. A disjunction is called *complete* if it is true; it then follows that at least one of its terms is true. A disjunction is called *exclusive* if not more than one of its terms is true. These concepts, as applied to probability sequences, are used in an extended sense: the disjunction must have these properties for all elements of the sequence. Thus completeness, in this sense, is formulated by the statement

$$(B_1 \vee \ldots \vee B_r) \tag{16}$$

The parentheses express, according to the convention given in §§ 10, 12, the condition that the disjunction is true for all elements of the sequence; and it would be more correct to speak of completeness and exclusiveness *with respect to the sequence*. The latter qualification is always understood when the terms "complete" and "exclusive" are used in probability considerations.

The combination of the two conditions of completeness and exclusiveness is expressed by the following r formulas, which are all-statements:[3]

$$\begin{gathered}(B_1 \equiv \overline{B_2} \,.\, \overline{B_3} \ldots \overline{B_r}) \\ (B_2 \equiv \overline{B_1} \,.\, \overline{B_3} \ldots \overline{B_r}) \\ \cdots\cdots\cdots \\ (B_r \equiv \overline{B_1} \,.\, \overline{B_2} \ldots \overline{B_{r-1}})\end{gathered} \tag{17}$$

The equivalence signs of the relations can be conceived as representing two mutual implications, according to (7a, § 4). The implication running from left to right expresses exclusiveness; the implication running from right to

[2] This is easily seen for the right-hand inequality. The proof for the left-hand inequality follows from the relation (5, § 23).

[3] The exclusive "or" cannot be used to express these conditions. See *ESL*, p. 45.

left expresses completeness. It can easily be shown that statement (16) is derivable from the relations (17).

For most of the following considerations it will be sufficient if the disjunctions are complete and exclusive with respect to A, that is, with respect to the subsequence selected by A. The symbolic expression is given by the formulas

$$A \supset \begin{cases} B_1 \equiv \overline{B_2} \, . \, \overline{B_3} \, . \, . \, . \, \overline{B_r} \\ B_2 \equiv \overline{B_1} \, . \, \overline{B_3} \, . \, . \, . \, \overline{B_r} \\ \quad . \quad . \quad . \quad . \\ B_r \equiv \overline{B_1} \, . \, \overline{B_2} \, . \, . \, . \, \overline{B_{r-1}} \end{cases} \tag{18}$$

From these formulas the statement of completeness relative to A is derivable:

$$(A \supset B_1 \vee \, . \, . \, . \, \vee B_r) \tag{19}$$

The condition (18) can be used to replace the stronger condition (17) in all cases in which only probabilities containing A in the first term are concerned. Thus when a die is thrown, the six possible results given by the six faces of the die constitute a disjunction that is complete and exclusive with respect to the sequence of events A represented by the throwing of the die. For the sake of simplicity, the condition (17) will always be used, leaving the reader to construct similar proofs on the basis of the weaker condition (18).

The introduction of many-term disjunctions in the rule of elimination is made in the same way as was used for the derivation of (2). Corresponding to (1), we have the relation

$$([B_1 \vee \, . \, . \, . \, \vee B_r] . C \equiv C) \tag{20}$$

Applying the inference leading to (2), we derive for many-term disjunctions the *extended rule of elimination*:

$$P(A,C) = \sum_{k=1}^{r} P(A,B_k) \cdot P(A \, . \, B_k,C) \tag{21}$$

Figure 5 (p. 82) may serve as an illustration. The divergent probabilities again are bound probabilities, so that

$$\sum_{k=1}^{r} P(A,B_k) = 1 \tag{22}$$

is valid; the convergent probabilities, however, are nonbound.

A schematized example for figure 5 is found in games of chance. Let $B_1 \, . \, . \, . \, B_r$ represent bowls containing black and white balls, each in a different ratio. Let C be the drawing of a black ball, and A an auxiliary bowl containing numerous tickets, each carrying one of the numbers $1 \, . \, . \, . \, r$. If there are more than r tickets in the bowl and each number occurs repeatedly, each number has a determinate probability of being drawn from the bowl. We draw first from the auxiliary bowl and determine from which of the other bowls we are to draw next. Repeating the two actions again and again, we obtain a statistical relation between A and C, the frequency of which is determined by $P(A,C)$ according to (21).

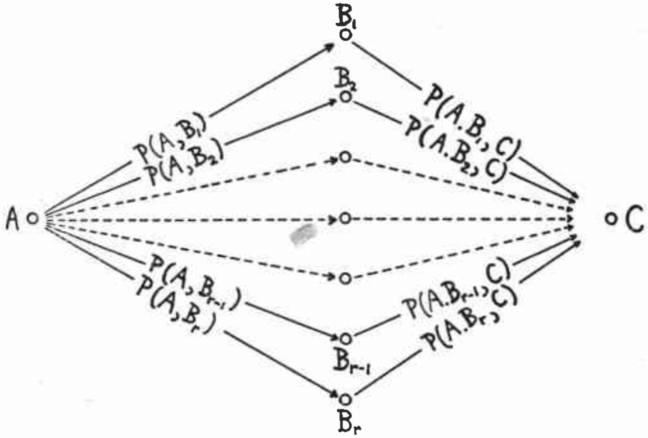

Fig. 5. Schema for extended rule of elimination, according to (21).

Another example results by taking for A the throwing of two dice, for C the occurrence of face 1 of the second die, for B_k the occurrence of face k of the first die. Then (21) means that the probability of obtaining 1 with the second die can be divided, additively, into the probabilities of the combinations in which this result is accompanied by one side k of the other die.

Both examples represent special cases of (21), namely, cases of such a kind that, for the first example, $P(A\,.\,B_k,C) = P(B_k,C)$ holds; for the second example, $P(A\,.\,B_k,C) = P(A,C)$. This corresponds to the causal conception of the problem, according to which, in the first example, B_k is the cause of C; in the second example, A is the cause of C. However, this is irrelevant to the treatment of the problem within probability theory; the lines in figure 5 represent probabilities, but not necessarily causal chains. The statement of the causal relationships requires specific investigation.

§ 20. The General Theorem of Addition

We shall now investigate the question how to calculate the probability of a disjunction if the terms of the disjunction do not mutually exclude one another, that is, if we are dealing with a nonexclusive disjunction. If, for example, two coins are thrown, what is the probability of obtaining tails with either coin, of obtaining *at least one* event of tails lying up? A simple addition would give $\frac{1}{2} + \frac{1}{2} = 1$—which obviously is a wrong result. But the conditions for applying the addition theorem are not satisfied, because it is possible to obtain tails simultaneously with both coins. In order to calculate the desired probability we must, therefore, transform the question into a form suitable for the application of the theorem of addition. Several such methods may be demonstrated.

We can start from the equivalence

$$(B \vee C \equiv B.C \vee B.\bar{C} \vee \bar{B}.C) \tag{1}$$

which leads to mutually exclusive terms and thus permits us to apply the theorem of addition:

$$\begin{aligned} P(A,B \vee C) &= P(A,B.C \vee B.\bar{C} \vee \bar{B}.C) \\ &= P(A,B.C) + P(A,B.\bar{C}) + P(A,\bar{B}.C) \end{aligned} \tag{2}$$

In the example of the two coins, the formula gives $P(A,B \vee C) = \frac{3}{4}$, because each of the probabilities of the combinations is equal to $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

In practice, other methods may be used to solve the problem. Occasionally it is possible, using material thinking (see § 5), to contract certain steps that are made separately in the calculus. The following method may be used: (1) B occurs; then it is immaterial whether or not C also occurs. The probability for this case is $P(A,B)$. (2) B does not occur; then C must occur. The probability for this case is $P(A,\bar{B}.C)$. Since the cases (1) and (2) are mutually exclusive, the theorem of addition is applicable, and we obtain

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

a result that is identical with (2) because of $P(A,B) = P(A,B.C \vee B.\bar{C})$. This method differs from the former one in that the first two cases of the disjunction (1) are collected in one case by the help of material thinking. This thinking can also be formalized: in (5e, § 4) we have a formula that leads directly to (3).

A third method starts from the equivalence

$$(B \vee C \equiv \overline{\bar{B}.\bar{C}}) \tag{4}$$

which leads with (7, § 13) to the simple result:

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

Here the probability of the opposite case is calculated and then is subtracted from 1. For the example with the two coins, the probability of obtaining heads with both coins is equal to $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Because in any other case at least one event of tails must happen, the desired probability is calculated to be $1 - \frac{1}{4} = \frac{3}{4}$.

We now establish for such probabilities a fourth formula that seems very convenient for technical reasons. It can be derived directly from the calculus without the aid of material thinking. Because of

$$\begin{aligned} P(A,B) &= P(A,B.C) + P(A,B.\bar{C}) \\ P(A,C) &= P(A,B.C) + P(A,\bar{B}.C) \end{aligned} \tag{6}$$

we can write, together with (2), the three formulas

$$\begin{aligned} P(A,B \vee C) &= P(A,B.C) + P(A,B.\bar{C}) + P(A,\bar{B}.C) \\ 0 &= -P(A,B.C) - P(A,B.\bar{C}) + P(A,B) \\ 0 &= -P(A,B.C) + P(A,C) - P(A,\bar{B}.C) \end{aligned} \tag{7}$$

Adding the three formulas, we obtain

$$P(A,B \vee C) = P(A,B) + P(A,C) - P(A,B.C) \tag{8}$$

This formula is called the *general theorem of addition*. It is a generalization of the addition theorem (3, § 13), applying to nonexclusive terms. In case $P(A,B.C) = 0$ it becomes identical with the *special theorem of addition* (3, § 13). In contradistinction to the latter, (8) represents an always-true formula because it is not contingent upon any conditions to be expressed in the context. The condition of exclusion, which had to be added verbally to the P-notation (3, § 13) as a logical condition, is *formalized mathematically* in (8); it is expressed by the case that a mathematical quantity assumes the value 0.

In the frequency interpretation, (8) can easily be made comprehensible. In dealing with the nonexclusive cases B and C, the couples $B.C$ will occur according to, say, the following schema:

$$\begin{matrix} A & A & A & A & A & A & \ldots \\ B & \bar{B} & \bar{B} & B & \bar{B} & B & \ldots \\ C & C & \bar{C} & \bar{C} & C & C & \ldots \end{matrix} \tag{9}$$

Adding the frequencies B and C, we shall have counted the couples $B.C$ twice; therefore, to form $P(A,B \vee C)$, the frequency of the couples $B.C$ is to be subtracted once. This fact is expressed in (8).

It need not be expressed as a condition that the probability of the disjunction, as given by (8), satisfy the normalization of probabilities; this follows from the double inequality (15, § 19) previously established. After a simple transformation by means of the theorem of multiplication, the inequality on the left side of (15, § 19) leads to

$$P(A,B) + P(A,C) - P(A,B.C) \leq 1 \tag{10}$$

Now the inequality on the right gives the result

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

By interchanging B and C we obtain

$$P(A,B.C) \leqq P(A,B) \tag{12}$$

Therefore the following inequalities are satisfied:

$$\begin{aligned} P(A,B \vee C) &\geqq P(A,B) \\ P(A,B \vee C) &\geqq P(A,C) \end{aligned} \tag{13}$$

The probability of a disjunction is never smaller and, in general, is even greater than the probability of its individual terms. Thereby the character of the disjunction as a logical sum is clearly expressed. Addition of a term connected by "or" signifies an increase in probability, and only in the limiting case does the probability remain the same.

Some examples may illustrate the general theorem of addition. The testing of a mechanical appliance reveals, on the average, 2% rejections because of material defects and 3% rejections because of defects in assembling the parts. What is the average rejection on the whole? Here the probabilities are given statistically, as is usual in practice. But we must not assume as total rejection $3\% + 2\% = 5\%$, since the two sources of defect are not mutually exclusive. An appliance that is faulty because of material defects may also show a defect owing to assembling. We know from experience that we are dealing here with independent probabilities; thus we can apply the special theorem of multiplication. The probability of both defects occurring simultaneously is given by the product $3\% \cdot 2\% = 0.06\%$. Then (8) provides as average frequency for the total rejection $3\% + 2\% - 0.06\% = 4.94\%$.

Another example is a firm that sells its products partly through traveling salesmen and partly through advertisements. The statistics on customers reveal that 80% of all products are sold by salesmen and 60% by advertisements. What is the percentage of customers won by advertisements as well as by salesmen? Since here $P(A,B \vee C) = 1$ (we assume that all products are sold only in these two ways), it follows that $P(A,B.C) = 80\% + 60\% - 100\% = 40\%$, that is, 40% of the customers are won by both means together.

Formula (8) permits a general calculation of the or-probability, but in applying it we must be sure that the case considered possesses the logical structure of the theorem of addition. Mistakes of this kind may be illustrated by two examples that were given by Richard von Mises[1] with the intention of showing that the addition must not be carried out uncritically, even for

[1] *Wahrscheinlichkeit, Statistik und Wahrheit* (Berlin and Vienna, 1928), p. 40.

mutually exclusive events. He wishes to restrict the theorem of addition to events belonging to the same "collective", that is, the same sequence. My formulation of the theorem is somewhat more general, since the theorem is not restricted to events belonging to the same sequence. Instead, another condition is used, specifying that the probabilities have the same reference class, or first term. I shall now show that my formulas are applicable to the examples given by von Mises, and permit the use of the "or" in a reasonable sense.

Assume that a tennis player has the probability 0.8 of winning in a tournament in Berlin; he may have the probability 0.7 of winning in a tournament played the same day at New York. The events are mutually exclusive; thus one might infer that the probability of the player winning in the one *or* in the other tournament was given by the addition of the probabilities, which would result in $0.8 + 0.7 = 1.5$. This is certainly a nonsensical result.

We are dealing here with a question of interpretation. A problem given in conversational language is to be translated into the strict language of the calculus; one cannot expect unambiguous rules to be available for such a translation. To assume that the special theorem of addition is applicable would be to interpret the problem in the form

$$P(A,B) = 0.8 \qquad P(A,C) = 0.7 \qquad P(A,B.C) = 0 \tag{14}$$

A representing the general situation before the tournaments; B, the victory in Berlin; C, in New York. It is obvious that the numerical values used in the interpretation violate the inequality (15, § 19), because $P(A,B.C) = 0$ implies $P(A.B,C) = 0$, whereas the expression on the left of the inequality assumes the value $\frac{5}{8}$. This illustrates the fact that the condition of exclusion represents a high degree of dependence and therefore can be combined only with suitable numerical values of the other given probabilities. It follows that (14) is not an admissible interpretation of the problem.

An interpretation that comes closer to what is intended by the formulation of the problem can be given. We consider the probability 0.8 of winning in Berlin as referring to the first term B_1, "if the player participates in Berlin"; and the probability 0.7 of winning in New York as referring to the first term B_2, "if the player participates in New York". If C represents "winning", we then can set down

$$P(B_1,C) = 0.8 \qquad P(B_2,C) = 0.7 \tag{15}$$

The two probabilities do not differ by their second term, as do the expressions (4), but by their first term. It is obvious that the probabilities do not permit the application of formula (8). The general condition A holding before the tournaments take place appears as a reference class in the sense of the theorem of elimination (fig. 5, p. 82), representing the fact that the player may decide to participate in one or the other of the tournaments; and the condition of exclusion must then be written

$$P(A,B_1.B_2) = 0 \tag{16}$$

When we wish to derive from these conditions the probability of winning, that is, $P(A,C)$, the two further probabilities

$$P(A,B_1) \qquad P(A,B_2) \tag{17}$$

must be given. This means that the probability of winning depends on the probabilities of the player deciding, respectively, to participate in New York or in Berlin.

In this interpretation the problem is solved, since $P(A,\overline{B_1 \vee B_2}.C) = 0$, by the equations

$$\begin{aligned} P(A,C) &= P(A,[B_1 \vee B_2 \vee \overline{B_1 \vee B_2}].C) \\ &= P(A,B_1.C) + P(A,B_2.C) \\ &= P(A,B_1) \cdot P(A.B_1,C) + P(A,B_2) \cdot P(A.B_2,C) \\ &= P(A,B_1) \cdot P(B_1,C) + P(A,B_2) \cdot P(B_2,C) \end{aligned} \tag{18}$$

because we may assume (10, § 14). That we cannot carry out the calculation numerically is due to the fact that the probabilities (17) are not given, but the failure to obtain a solution does not result from an inadmissible use of the "or". It is clear, furthermore, that in this interpretation the sum of $P(B_1,C)$ and $P(B_2,C)$ can be greater than 1, since these values represent nonbound probabilities (see § 19).

Von Mises presents another example that is supposed to demonstrate the use of an unreasonable "or". Let 0.011 be the probability that a man 40 years of age will die between his 40th and 41st birthdays; and let the probability that a man 41 years old marries in that year be 0.009. Both events are exclusive for one individual. If we now want to find the probability that a man 40 years of age either dies within the current year or marries in the following year, it may occur to us to add the given numbers, thus obtaining the result, $0.011 + 0.009 = 0.020$. Von Mises is right in asserting that this is a nonsensical result.

For the conception of the or-probability developed in this section, however, the problem is not meaningless. The probability of a man 40 years old dying this year or marrying next year can be interpreted to have a definite meaning. It may be expressed statistically: after a lapse of two years, we count among all the original quadragenarians those who died within the first year or married in the second year. These numbers may indeed be added, in agreement with (8). However, we must not add the numerical values given; the second value cannot be used because it states, not the probability that a man 40 years of age will marry in his 41st to 42d year, but the probability that a man 41 years of age will marry in that period. The probabilities are not the same, because some of the men will have died within the year. The value 0.009, therefore, is to be interpreted as the probability that a man 40 years old who reaches his 41st year will marry in his 41st to 42d year. This probability is represented by $P(A.\bar{B},C)$, if A stands for the class of quadragenarians, B for the class of deaths among them, and C for the class of men 41 years old who marry. We have, therefore,

$$P(A,B) = 0.011 \qquad P(A,B.C) = 0 \qquad P(A.\bar{B},C) = 0.009 \tag{19}$$

and obtain

$$\begin{aligned} P(A,B \vee C) &= P(A,B) + P(A,C) \\ &= P(A,B) + P(A,[B \vee \bar{B}].C) \\ &= P(A,B) + P(A,B.C) + P(A,\bar{B}.C) \\ &= P(A,B) + P(A,\bar{B}) \cdot P(A.\bar{B},C) \\ &= 0.011 + (1 - 0.011) \cdot 0.009 = 0.0199 \end{aligned} \tag{20}$$

This represents the probability that a man 40 years of age either will die in his 40th to 41st year or will marry in his 41st to 42d year.

In criticizing these examples I do not wish to deny that the probability calculus of von Mises supplies equally correct solutions. I intend merely to show that we can dispense with the relatively complicated operations of constructing new collectives, which von Mises has introduced, and that the desired probabilities can be conceived reasonably as or-probabilities.

We shall now derive from the general theorem of addition some consequences for later use. We can calculate a probability of the form $P(A,B \supset C)$ by resolving the implication into $\bar{B} \vee C$ according to (6*a*, § 4) and then applying the general theorem of addition. We obtain

$$\begin{aligned} P(A,B \supset C) &= P(A,\bar{B} \vee C) \\ &= P(A,\bar{B}) + P(A,C) - P(A,\bar{B}.C) \\ &= P(A,\bar{B}) + P(A,C) - P(A,\bar{B}) \cdot P(A.\bar{B},C) \end{aligned} \tag{21}$$

By the use of (10, § 19) we arrive at

$$P(A,B \supset C) = 1 - P(A,B) + P(A,B) \cdot P(A.B,C) \tag{22}$$

In a similar way we obtain for the equivalence, by the dissolution $([B \equiv C] \equiv [B.C \vee \bar{B}.\bar{C}])$, according to (7*b*, § 4), and with (10, § 19),

$$\begin{aligned} P(A,B \equiv C) &= P(A,B.C \vee \bar{B}.\bar{C}) \\ &= P(A,B.C) + P(A,\bar{B}.\bar{C}) \\ &= P(A,B) \cdot P(A.B,C) + P(A,\bar{B}) \cdot P(A.\bar{B},\bar{C}) \\ &= 1 - P(A,B) - P(A,C) + 2\,P(A,B) \cdot P(A.B,C) \\ &= 1 + P(A,B.C) - P(A,B \vee C) \end{aligned} \tag{23}$$

A formula containing an exclusive "or" will now be constructed. According to (1, § 4), this operation can be defined as

$$b \wedge c =_{Df} (b \vee c).\overline{b.c} \tag{24}$$

Because of the equivalence

$$(b \vee c).(\overline{b.c}) \equiv (b \vee c).(\bar{b} \vee \bar{c}) \equiv b.\bar{c} \vee \bar{b}.c \tag{25}$$

we can write, using (7*b* and 7*c*, § 4),

$$b \wedge c \equiv \overline{b \equiv c} \tag{26}$$

The symbol of the exclusive "or" can be used also in the class calculus. The class $B \wedge C$ represents, according to (24), the common class of $B \vee C$ and $\overline{B.C}$, that is, the part of the joint class of B and C that results by subtracting the common class of B and C. Because of the relation (26) we have

$$P(A,B \wedge C) = P(A,\overline{B \equiv C}) = 1 - P(A,B \equiv C) \tag{27}$$

With the use of (23) we obtain, applying (8),

$$P(A,B \wedge C) = P(A,B) + P(A,C) - 2\,P(A,B.C) \tag{28}$$

Although we have thus derived a formula dissolving an exclusive "or", the result shows that it is not possible, for the special theorem of addition, to eliminate the condition of exclusion by the use of a symbol for the exclusive "or". The formula

$$P(A,B \wedge C) = P(A,B) + P(A,C) \tag{29}$$

is false if it is conceived as holding for all B and C; it holds only if $P(A,B.C) = 0$, that is, if B and C are mutually exclusive. But if this condition must again be added, the introduction of the symbol of the exclusive "or" is useless. The aim of expressing the addition theorem completely in the mathematical notation is achieved, instead, in the general theorem of addition formulated in (8).