V.

THE FOUNDATION OF THE THEORY OF PROBABILITY—I.

(From the Dublin Institute for Advanced Studies.)

By ERWIN SCHRÖDINGER.

[Read 24 June, 1946. Published 23 January, 1947.]

The preface to a monograph in French on the Basis of the Theory of Probability by Pius Servien[1] has induced me to collect my own thoughts about it. It is said there that the unsatisfactory state of probability theory is bound to tell on modern physics, in which the former plays such a prominent part. I do not think that the case is very serious, but I do admit, that the readiness with which we physicists and many others who use the theory in practice adopt Frequency as the basis, does mean taking things a little too easy. There are grave objections to this, mainly that we thereby cut ourselves off from ever applying rational probability considerations to a single event.[2]

I attempt here to keep aloof from the frequency definition of probability and yet to reduce the volume of *axiomatic* statements to a minimum. In this first paper only the product rule retains this rank, everything else is simplifying convention, a normalization of the measure of likelihood. In a second paper, by completing this normalization, the axiomatic basis will be reduced to—very little, next to nothing.

For the sake of succinctness pure common logic is taken for granted. It complicates matters enormously, to treat probability theory as sort of a generalization thereof. It deals with facts, not with propositions.

2. Events and our Knowledge about them.

By event *we understand for the present purpose a simple or arbitrarily complicated individual state of affairs (or fact or occurrence or happening) which either does or conceivably might obtain in the real world around us and of which we are given a description in words, clear and accurate enough to leave us no doubt, that by taking (or having taken at the time or times in question) sufficient cognizance of the*

[1] Paris, Hermann & Cie., 1942.

[2] A lucid discussion of the frequency theory of probability is found in Ch. VIII, p. 92, of J. M. Keynes, A Treatise on Probability. London, MacMillan and Co., 1921;

relevant part of the world it would be possible to decide unambiguously, whether this particular fact (or state of affairs, etc.) actually obtains or not, any third possibility being excluded.

The meaning is not, that it must be practically possible to procure the necessary knowledge. The event may refer to the past and may have taken place (if it has taken place) without witnesses and without leaving fully convincing traces. It may, by the way, cover a span reaching from the past to the future, e.g. the verbal description might read: A certain person, now alive, has either already had an attack of tuberculosis or will have one in future. What we mean is, that the verbal description must be sufficiently precise and adequate, so that according to the sum total of our knowledge of the world it needs must either agree or disagree with actual facts, past, present or future.

As verbal descriptions not fulfilling the requirement and thus, in my opinion, not specifying an event, let me mention:

"The distance between the towns D. and G. is between 157·357124 and 157·357125 miles."

"The ghost of the late Duke has appeared to the castellan last night."

The first case is obvious. In the second case we would first of all have to be told, whether a vivid hallucination on the side of the castellan is to be included as constituting the coming true of the event. But even if we exclude that and take it that an "actual apparition" is meant, our general knowledge about the connexion between an actual apparition and the personality of a deceased is at any rate not such as to make the statement unambiguous.

On the other hand, it is hardly necessary to say, that by pointing to "the real world around us" we do not exlude that in studying the theory of our subject and dealing with examples for exercise we frequently refer to merely imagined events. Yet we have to imagine them actually to take place; they are not just constructions in the mind, as a straight line, a triangle or a surface of the second order. In the same way, the efforts of an engineer who constructs a new locomotive on the drawing board must be said to refer to the real world, no matter whether his design will actually be put into practice or eventually rejected.

An event can thus either be true or untrue, and according to our definition it must be one or the other. In addition to this, and quite independently from it, one of the following two alternatives can obtain: we know which of the two is the case or we do not know it.

In the latter case it can happen, that we have some partial evidence or clue to make conjectures about it.

Plain as it is, there is reason to emphasize the distinction between the intrinsic or objective certainty an event must exhibit to become a worthy object of probability considerations, and the subjective uncertainty about it. The idea that such considerations apply in a non-trivial way *only* to events with an inherent innate uncertainty is not quite unfamiliar. But surely only a fool would find it worth his while to make conjectures about an uncertain event. A historian who gathers evidence as to whether a certain person has died before or after A.D. 635, does so in the conviction that this person has lived, and is not a legendary figure. And even when in an example for exercise we utter an opinion about whether the next cast of the die will be even or odd, this refers to the case that the die *will be cast,* otherwise it is void.

Doubt has been raised whether probability considerations could duly be applied to events with a verbal description of the following kind:—

"The 23[rd] decimal figure of π is 4."

The case is hardly covered by the definition of an event we gave above, which would become rather clumsy if we wished to include also such "internal events" as is the functioning of the mind in calculating this figure. But to exclude such questions on the ground, that they can be decided by pure logical reasoning with full certainty, is to my mind not justified. On the contrary, in this example the postulate of intrinsic certainty is fulfilled in the most ideal way.

3. Preliminary Definition. Symmetry Principle.

Given the state of our knowledge about everything that could possibly have any bearing on the coming true[3] *of a certain event* (*thus* in dubio: *of the sum total of our knowledge*), *the numerical probability p of this event is to be a real number by the indication of which we try* in some cases *to set up a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true.*

Considering the unlimited manifold of events and again the manifold of thinkable states of our knowledge, it is patently impossible to condense into one brief definition the functional connexion between the number p and the state of our knowledge. This connexion can only gain shape gradually in the following considerations on the basis of certain *Conventions* (see sect. 4) and of Axioms (see sect. 5).

Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event. Indeed, it may be that one state of knowledge about the same event allows the application of quantitative

[3] While in ordinary speech "to come true" usually refers to an event that is envisaged before it has happened, we use it here in the general sense, that the verbal description turns out to agree with actual facts.

measure, while another one refuses to be subject to it. Thus whenever we speak loosely of the "probability of an event," it is always to be understood: probability with regard to a certain given state of knowledge.

Even before tightening our programmatic definition, we can infer from it the very important

SYMMETRY PRINCIPLE: *It can happen, that our knowledge is perfectly symmetric with respect to two or more events of the same kind. In this case we must attach numerical probabilities either to none of them or to all of them, and in the latter case obviously the same to each of them, since this number is to depend only on the state of our knowledge, which we suppose to be the same for any two of them.*

It is known that many, if not most, mathematical theories hinge on the invariance with respect to a group of transformations, e.g. all geometries do. On the other hand, even in experimental science we frequently build an apparatus (for instance, the chemist's weighing scale) as nearly as possible symmetric, in order to facilitate certain conclusions. Nay, even in the considerations attached to a frequent repetition of the same measurement or to the "blind check" (repeating the whole performance with the body or agent, to be measured, left out but all other circumstances unaltered) the principle of symmetry is involved. Thus it is not astonishing, that in *this* branch of applied mathematics the invariance with respect to a symmetrical group of permutations (or sometimes several such groups) plays a fundamental rôle, and it is difficult to see why the concept of "equal *a priori* probabilities" has given rise to so much altercation.

But it is to be emphasized, that the application of the principle requires real symmetry, of which a necessary, but not sufficient, condition is that the events be exactly of the same kind. Also, complete lack of knowledge, or equally scanty knowledge, in both or all cases does not yet establish symmetry. For instance, it is *not* equally probable that a female of whom we know nothing but the name is married or is a spinster. Indeed our *general* knowledge contains a lot about the manner by which a female gets into either of these states, and in this respect there is a complete lack of symmetry. On the other hand, we *would* regard it as equally probable that the said female was born on a Thursday or on a Friday.

We have purposely termed the considerations of this section a principle, not an axiom, because we believe them to be a necessary consequence of our programmatic definition.

Wherever in the following we speak of probability, we mean the numerical probability, and the enouncement shall always refer only to the case, that the state of our knowledge justifies the attachment of a real number to the "degree of our anticipation"—justifies it according to the principles we put forward.

4. Conventions. The probability of $\bar{\alpha}$.

We have hitherto only said that our measure of likelihood is to be a real number. One might think of setting up a discontinuous scale as is done for estimating the strength of the wind. For many cases that would suffice and prove quite satisfactory. For refined consideration it would entail complication, just as the Beaufort scale would, if you used it for investigating the dependence of the velocity of the wind on the pressure gradient and the geographical latitude.

We thus decide to allow the variable p a continuous *uninterrupted* range between certain limits. I say uninterrupted, for if there were a "forbidden" interval, we could hardly avoid attaching the same meaning to its two limits and the measure would cease to be unique. The converse inconvenience, namely, that sometimes different degrees of likelihood were expressed by the same number, would occur, if we did not declare our intention, that greater likelihood shall always be expressed by a larger p. One might suspect that this stipulation is void, on the ground that we are not given a primordial measure of likelihood, we are only just about to forge one. But it is certainly not entirely void. It demands, for instance, that p_β must not be smaller than p_α if the event α includes the event β conceptionally according to their verbal descriptions.

The two limits of the range of p must then correspond to the highest and to the lowest thinkable degrees of anticipation, that is to certainty and to patent impossibility (or certainty of the event not being true).

What real interval we choose for the range of p is still at our discretion. But since any two of them can easily and simply be mapped on each other, the customary choice $0 \rightarrow 1$ has from the point of view taken here no deep significance.

We condense these considerations to form our

FIRST CONVENTION: *The probability p shall range continuously from* 0 *to* 1, *whereby greater likelihood* (*more confident anticipation*) *shall be expressed by a larger p, $p = 1$ indicating certainty and $p = 0$ indicating the "certainly not".*

As regards the limiting cases, it is convenient (as is sometimes done in mathematical disciplines) to supplement the manifold of events, properly speaking, by two ideal events by admitting also a self-contradictory verbal description (impossible event) and a description that is actually or virtually empty (it might be called nil event). The simplest examples are, for the first: both α and $\bar{\alpha}$ (meaning non-α) are in agreement with facts; for the second: either α or $\bar{\alpha}$ agrees with facts. These ideal events are safe representatives of the cases $p = 0$ and $p = 1$, while in general considerations we shall consider both p and $1 - p$ different from zero.

Since a conjecture about an event coming true amounts to the same as a conjecture about its not coming true, we needs must attach a probability to $\bar{a}$ whenever we attach one to a, and $p_{\bar{a}}$ must be a single-valued, monotonously decreasing universal function of p_a, the two referring, of course, to the same state of knowledge. This function must take the value 0 for argument 1 and vice versâ. Moreover, since there is no possible distinction between anything like "positive" and "negative" events, the connexion between p_a and $p_{\bar{a}}$ must be symmetric; thus we can express it in the form

$$F(p_a, p_{\bar{a}}) = 0 \qquad (1)$$

where F is a symmetric function of its two arguments.

Let the solution of (1) be

$$p_{\bar{a}} = \phi(p_a) \qquad \text{or also} \qquad p_a = \phi(p_{\bar{a}})$$

(on account of the symmetry). Since $p_{\bar{a}}$ goes monotonically from 1 to 0 when p_a goes from 0 to 1, they must meet once and only once, say at g:

$$g = \phi(g).$$

Now let us explain a monotonically increasing function $f(x)$ for $0 \leqslant x \leqslant 1$ thus

$$\left.\begin{aligned} f(0) &= 0 \\ f(g) &= 1/2 \end{aligned}\right\} \quad \begin{aligned} &\text{otherwise arbitrary,} \\ &\text{for } 0 \leqslant x \leqslant g \end{aligned}$$

$$f(x) = 1 - f[\phi(x)] \quad \text{for } g \leqslant x \leqslant 1.$$

If then we put *quite generally* $p' = f(p)$, then we have $p'_a + p'_{\bar{a}} = 1$. We formulate this as our

Second Convention: *The measure of probability p shall be normalized from the outset so as to satisfy the relation*

$$p_a + p_{\bar{a}} = 1 \qquad (2)$$

between complementary probabilities.

We shall see in sect. 6, that thanks to this convention the Addition Rule *follows* from the Multiplication Rule. Since the normalization we have performed is far from rigid, allowing still a monotonical transformation arbitrary within half the range, one suspects that by tightening it, even more can be achieved in the way of reducing axioms to mere conventions. That is so. But the considerations in question are conceptionally rather involved and require particular care, in order to avoid *petitiones principii.* They would unduly interrupt our present simple trend of ideas, and so we leave them to a subsequent paper. At the moment we acquiesce in introducing the Multiplication Rule as an axiom, the only one that is still left in the present outline of the theory.

5. The Multiplication Axiom.

Probabilities that appear in the same equation or in the course of the same calculation without comment shall always all refer to the same state of our knowledge. However, the notation, e.g. $p_\alpha(\beta^+)$ shall mean the probability the event α would acquire, if in addition to that general state of knowledge we knew the event β to be true, i.e. to agree with facts. In the same way $p_\alpha(\beta^-)$ refers to the addition "if we knew β to be untrue, i.e. to disagree with facts." Similarly with $p_\alpha(\beta^+, \gamma^-, \ldots)$, etc.

By $p_{\alpha\beta\gamma\ldots}$ we mean the probability of the event whose verbal description is obtained by joining together the verbal descriptions of the events α, β γ . . ., in other words the probability of α *and* β *and* γ *and* . . . being in agreement with facts. The result of this "joining together" may, of course, require quite an intricate logical analysis, since we must not exclude the possibility, that one of the verbal descriptions refers to the coming true or not of some of the other events. However if words mean something, the analysis must be possible. The only thing we do exclude is, that they should refer explicitly to the *order* in which we put them together—for we are not going to tell anybody in advance! It then follows that e.g. $p_{\beta\alpha}$ has the same meaning as $p_{\alpha\beta}$, and similarly with more than two "factors."

THE PRODUCT AXIOM:

$$\begin{aligned} p_{\alpha\beta} &= p_\alpha\, p_\beta(\alpha^+) \\ &= p_\beta\, p_\alpha(\beta^+)\,. \end{aligned}$$

Here we must add an ample commentary. First fix your attention to the first equation only. The axiom does not only mean to state, that if these three probabilities exist, they bear the said relation to each other, but that if two of them exist, the existence of the third can be inferred. Apart from this the second relation is trivial, given the first, and need not be included in the *axiom.*

Suppose we had two equal urns and know each to contain 60 balls, which for one of them have been selected by a Prussian to represent the colours of his country, i.e. it contains 30 white and 30 black balls, while the other one has been filled according to the same principle by a person, selected by lot from three Europeans of different, but entirely unknown, nationality. (E.g. a Swede would have put in 30 yellow and 30 blue balls, an Irishman 20 green, 20 white and 20 orange balls, etc.) We select one urn and we draw out of it one ball. Envisage the two events—

(α) We have selected the "Prussian" urn.

(β) We have drawn a white ball.

Here p_α is 1/2 and p_β (α^+) is 1/2, hence from the first line of our axiom $p_{\alpha\beta}$ exists and equals 1/4. But the second line is inapplicable, because neither p_β nor $p_\alpha(\beta^+)$ exist.—

It is known, that our "axiom" becomes selfevident, if the frequency definition of probability is adopted: in a wide statistics (actually in *any* statistics) the fraction of cases of α and β both happening, is, of course, identically equal to the fraction of β^+ cases among the α^+ cases, multiplied by the fraction of α^+ cases among the whole lot.—From the point of view adopted for the moment, the axiom is not at all selfevident. It is to be regarded as an attempt to complete the normalization of p, and we ought indeed to make sure, that it is compatible with previous conventions.

Since the event $\alpha\beta$ includes the event α, the p_α cannot be smaller than $p_{\alpha\beta}$ so that the second factor on the right (of the first equation) cannot turn out greater than 1, when the two other probabilities are given. Neither can p_α, when the other two are given. For p_β (α^+) is logically the same as $p_{\alpha\beta}$ (α^+). And surely our expectation that both α and β come true, cannot be diminished by the additional knowledge, that α *is* true.

I can see no way in which the axiom could clash with the convention about complementary probabilities, but also no direct way of showing that it does not. This flaw will be removed by completing the normalization explicitly (see the remark at the end of the previous section).

Whenever both lines of the axiom apply, we get useful information on the mutual influence that knowledge about one event has on the probability of the other. To begin with

$$\frac{p_\alpha(\beta^+)}{p_\alpha} = \frac{p_\beta(\alpha^+)}{p_\beta}\,. \tag{3}$$

In words: The knowledge that β is true changes (increases or diminishes) the probability of α in the same ratio, in which, conversely, the probability of β is changed by the knowledge, that α is true.

On replacing α and β by $\bar{\alpha}$ and $\bar{\beta}$ respectively, we get nothing essentially new, but, of course, the equality of ratios now holds for the complementary probabilities $p_{\bar{\alpha}} = 1 - p_\alpha$ and $p_{\bar{\beta}} = 1 - p_\beta$.

Applying (3) to α and $\bar{\beta}$ we easily obtain

$$p_\alpha(\beta^-) = \frac{p_\alpha(1 - p_\beta(\alpha^+))}{1 - p_\beta}\,. \tag{4}$$

In particular, *if* α *and* β *are incompatible*

$$p_\alpha(\beta^-) = \frac{p_\alpha}{1 - p_\beta} . \tag{5}$$

The right-hand side cannot become greater than 1, because in this case α includes $\bar{\beta}$.

In general we see that, given p_α and p_β it suffices to know one of the four quantities

$$p_\alpha(\beta^+), \quad p_\alpha(\beta^-), \quad p_\beta(\alpha^+), \quad p_\beta(\alpha^-),$$

in order to calculate the other three. In particular, if the additional knowledge expressed in the bracket, is irrelevant in one of the four cases, it is so in all of them. We then call the two events *independent* with respect to the *general* state of knowledge, to which the consideration refers. (In the Appendix we give a simple non-trivial example of two events becoming dependent by a change in the general state of our knowledge.)

6. The Summation Rule.

Let the event which comes true (this is to be its verbal description!) if and only if at least one of the events α, β, γ . . . comes true, be indicated symbolically by $\alpha + \beta + \gamma + \ldots$ and its probability by $p_{\alpha+\beta+\gamma+\ldots}$; then from the Second Convention

$$p_{\alpha+\beta+\gamma+\ldots} = 1 - p_{\bar{\alpha}\bar{\beta}\bar{\gamma}\ldots} .$$

Applying this at first to two events we get from the axiom and the 2nd Convention:

$$\begin{aligned} p_{\alpha+\beta} &= 1 - p_{\bar{\alpha}\bar{\beta}} = 1 - p_{\bar{\alpha}}\, p_{\bar{\beta}}(\alpha^-) \\ &= 1 - (1 - p_\alpha)(1 - p_\beta(\alpha^-)) \\ &= p_\alpha + (1 - p_\alpha)\, p_\beta(\alpha^-) . \end{aligned}$$

If here we use (4) (with the roles of α and β exchanged in it), we find

$$\begin{aligned} p_{\alpha+\beta} &= p_\alpha + p_\beta - p_\beta\, p_\alpha(\beta^+) \\ &= p_\alpha + p_\beta - p_{\alpha\beta} \\ &\leqslant p_\alpha + p_\beta . \end{aligned} \tag{6}$$

Applying this to the events $\alpha + \beta$ and γ, we get

$$\begin{aligned} p_{\alpha+\beta+\gamma} &= p_\alpha + p_\beta - p_{\alpha\beta} + p_\gamma - p_\gamma\,[p_\alpha(\gamma^+) + p_\beta(\gamma^+) - p_{\alpha\beta}(\gamma^+)] \\ &\leqslant p_\alpha + p_\beta + p_\gamma \end{aligned}$$

and finally

$$p_{\alpha+\beta+\gamma} = p_\alpha + p_\beta + p_\gamma - p_{\alpha\beta} - p_{\beta\gamma} - p_{\gamma\alpha} + p_{\alpha\beta\gamma} \\ \leqslant p_\alpha + p_\beta + p_\gamma . \tag{7}$$

The generalisation for any number of events is most succinctly written thus

$$p_{\alpha+\beta+\gamma+\ldots} = 1 - (1 - v_\alpha)(1 - v_\beta)(1 - v_\gamma)\ldots \\ \leqslant p_\alpha + p_\beta + p_\gamma + \ldots, \tag{8}$$

where however the v's are not numbers, only algebraic symbols. After multiplying out, the power-product $v_\alpha v_\delta v_\epsilon \ldots$ is to be interpreted as $p_{\alpha\delta\epsilon\ldots}$.

It is easy to see, that in the inequalities (6), (7) and (8) the equality sign holds if and only if any pair of events is mutually exclusive. In particular it follows, that the truth of any two out of the three following statements entails the truth of the third:

(i) the events are mutually exclusive,

(ii) their logical sum is the nil-event,

(iii) the sum of their probabilities is 1.

That the application of the Symmetry Principle to events of this kind is the principal source for obtaining numerical probabilities of certain basic events, to serve as working material for calculating the probabilities of others, is well known and we need not enlarge upon it.

7. The Frequency in a Series of Trials.

If one introduces the notion of numerical probability by conventions and axioms rather than by referring it directly to frequency, one has to trace its bearing on statistics afterwards. The connexion hinges, of course, on the Bernoulli Theorem, which states that if a generic event of probability p is N times offered the opportunity of taking place, the probability of its being realized precisely m times out of the N is the term of the binomial series.

$$\binom{N}{m} p^m q^{N-m},$$

where $q = 1 - p$. From this it can be shown, that for N and m large the probability of $\frac{m}{N}$ deviating appreciably from p becomes vanishingly small.

The dominating rôle of this theorem is, I think, responsible for the habit of regarding the numerical probability p as characteristic of the event or the kind of event rather than of our state of knowledge about it. For how should a merely subjective character result in palpable, well-nigh predictable frequencies? But one must not forget, that there is an intimate connexion and mutual influence between our knowledge and the way the statistical data are collected.[4] For a new born baby to reach the age of 70 there is a certain probability (of the order of 0·3). For a high-court-judge the probability is considerably higher. Yet the baby may become a high-court-judge and the judge certainly has once been born. So there is no difference in the nature of the *event*.

Most people agree, that the familiar way of determining probability in practice from statistical frequency is vindicated not directly by the Bernoulli Theorem, but by something like its inversion, about which there has been much controversy. We take from statistical data, that out of 10,000 people aged fifty 157 have died during the following year, and we say 0·0157 is the probability of this happening with a man of fifty, say with the 10,001st one. It thus appears that our anticipation concerning the latter is strongly influenced by our knowledge about the 10,000 previous cases. That strictly contradicts the *independence* between the single events, on which the Bernoulli Theorem is founded. One feels obliged to tackle the question without this assumption.

We contemplate N events of exactly the same kind,

$$a_1, \quad a_2, \quad a_3 \; . \; . \; . \; , \quad a_N, \tag{9}$$

and we assume that our basic knowledge about them is exactly symmetric. (This includes, of course, that if there is a known time order between them, it is disregarded. Very often, e.g. in most of the data taken from statistical tables, there is none.)

Now envisage the 2^N product events of the type

$$a_1 \bar{a}_2 \bar{a}_3 a_4 \; . \; . \; . \; . \; a_{N-1} \bar{a}_N \; .$$

They are mutually exclusive and their "sum" is the nil-event. Hence

$$1 = \sum_{(2^N)} p_{a_1 \ldots \bar{a}_N} \, .$$

We think of these p's as formed by repeated use of the product axiom. But on account of the symmetry only the number m of successful events can be relevant. So we shall label p just by this number. Or rather we shall label it $p_{m,\,N}$, in order to indicate also the dependence on N.

---

[4] See in the Appendix the second example, dealing with the "Ace of Spades".

Thus

$$1 = \sum_{m=0}^{N} \binom{N}{m} p_{m,N}. \tag{10}$$

From the product axiom the general term can be written

$$\binom{N}{m} p_{m,N} = \binom{N}{m} p_{m-1,\ N-1.}\ p_{\alpha}(m-1,\ N-1) \tag{11}$$
$$(m \geqslant 1)$$

and the preceding one can be written

$$\binom{N}{m-1} p_{m-1,\ N} = \binom{N}{m-1} p_{m-1,\ N-1}\ p_{\bar{\alpha}}(m-1,\ N-1). \tag{12}$$

Here $p_{\alpha}(m-1,\ N-1)$ means the probability any one of the single events (9) would acquire, if it were known that $m-1$ out of the $N-1$ others agree with facts, or in more familiar language, the probability of success of the N^{th} trial, it being known that $N-1$ trials have produced $m-1$ successes. The $p_{\bar{\alpha}}(m-1, N-1)$ is the complementary probability, viz. $1 - p_{\alpha}(m-1, N-1)$. Calling r_m the ratio of (11) to (12), we have

$$r_m = \frac{N-m+1}{m} \frac{p_{\alpha}(m-1,\ N-1)}{p_{\bar{\alpha}}(m-1,\ N-1)},$$

which gives

$$p_{\alpha}(m-1,\ N-1) = \frac{m\, r_m}{N+1+m(r_m-1)}.$$

Now supposing that for a particular m the r_m equalled 1 (as it very nearly does near the maximum term of a Bernoulli series), then for this m we should have

$$p_{\alpha}(m-1,\ N-1) = \frac{m}{N+1}. \tag{13}$$

This is known as the Rule of Succession, and is sometimes claimed to be valid for *any* m and N.

How are we to interpret it? Are we justified in assuming $r_m = 1$, identically in m? Hardly. Taken literally, it would mean, that all the terms of the series (10) are equal, in other words that every m is allotted the same *a priori probability,* i.e. probability before anything is known about the successes of the series of trials.

We are not inclined to accept this argument, which by the way produces gratuitous results for small N, such as

$$p_{\alpha}(0,\ 0) = \frac{1}{2}, \quad p_{\alpha}(1,\ 1) = \frac{2}{3}, \text{ etc.}$$

which cannot be taken seriously.

In my opinion (13) can only be taken seriously for at least fairly large N, m, and $N - m$, and must then be vindicated as follows. The idea is that at the outset we know too little about the events (9) for attaching a numerical probability to them. The very object of the series of trials is to furnish one. Now this lack of information amounts to something one might call near-symmetry or quasi-symmetry between two events whose probabilities are successive terms of the series (10), provided that N is large and m is not too near its limits 0 and N. To fix the ideas take $N = 50$. The compound events $m = 34$ and $m = 35$ respectively are then so nearly the same thing, that we feel justified in putting r_{35} very nearly equal to 1. To the bold assumption of equal probabilities for all m our interpretation is superior, in as much as even quite a strong initial bias is compatible with r_m *very nearly* equal to 1, provided N is large and the bias is not too sharp and exclusive.

I do not claim logical rigour for these considerations, and I grant that they gather psychologic strength from the tacit surmise, that the compound event be governed by a Bernoulli series, which as we know gives to m outside the region where $r_m \approx 1$ a probability that vanishes in the limit $N \to \infty$.

APPENDIX.

The gold and silver chests:

The following is a simple non-trivial example of two events at first independent, becoming dependent with respect to an increased state of knowledge. The scheme of the chests is not my invention, I remembered it from Professor Mertens' lectures in Vienna, about 40 years ago.

Three equal little chests of two drawers each. One chest contains a gold coin in each of its drawers, another one a silver coin in each, the third one a silver coin in one, a gold coin in the other drawer:

No. 1	No. 2	No. 3
G G	G S	S S

The "main player" A and the "auxiliary player" B each take one of the chests by random choice and keep it.

This we call the *first state of knowledge*. We envisage the two events

(α) A has got the No. 2 chest.

(β) When A opens one drawer of the chest, he will find gold.

We have

$$p_\alpha = \frac{1}{3}, \qquad p_\beta = \frac{1}{2}.$$

The first is obvious. To fall in with the second, consider that what A will have done is after all only to open at random one of the six drawers.—Moreover

$$p_\beta(\alpha^+) = \frac{1}{2}.$$

Hence α and β are independent. (We could easily, but we need not, investigate the three other probabilities, as $p_\beta(\alpha^-)$, etc.)

We now turn to a *second state of knowledge*, which is reached by B actually opening one of the drawers of *his* chest and informing A, that there is a gold coin in it. The probabilities referring to this state of knowledge shall be distinguished by a dash.

By analogy with the previous case, B's finding leaves *his* probability of holding No. 2 unchanged at 1/3. The probability that he has *not* got it, is thus 2/3, and if so, there are even chances for A to have it or not to have it. Hence

$$p'_\alpha = \frac{1}{3}.$$

To find p'_β consider that B's probability of holding No. 2 is 1/3, and if he held it, p'_β would obviously be 1/2; B's probability of holding No. 1 is 2/3, and if he held it, p'_β would be only 1/4. Thus

$$p'_\beta = \frac{1}{3} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{4} = \frac{1}{3}.$$

Now clearly

$$p'_\beta(\alpha^+) = \frac{1}{2}.$$

Hence now α and β are no longer independent. From the results of section 5 we easily compute

$$p'_\beta(\alpha^-) = \frac{1}{4} \qquad p'_\alpha(\beta^+) = \frac{1}{2} \qquad p'_\alpha(\beta^-) = \frac{1}{4}.$$

This could be confirmed by direct, but somewhat involved, reasoning.

The ace of spades:

I owe this example to oral communication in 1938 by Henry Whitehead, Balliol College, Oxford.[5] It illustrates how the precise wording of a questionnaire tells on the outcome of a statistical enquiry.

[5] But if there be a mistake in the following computation, it is mine not his.

A hand of whist is dealt. Only one player (A) takes up his hand. He is asked whether he has an ace, and he answers truthfully: yes. What is the probability of his having more than one? (p'_x)

After having answered this question, he is asked: have you the ace of spades? He answers truthfully: yes. What is *now* the probability of his having more than one ace? (p_x)

The point is, that the additional information *seems* to be irrelevant, while it is not. Actually $p_x > p'_x$ and the difference is not slight.

We answer the second question first. The probability that the three other aces are *not* among the remaining 12 cards of A's hand, but among the 39 others is

$$\frac{39 \times 38 \times 37}{51 \times 50 \times 49}.$$

Hence

$$p_x = 1 - \frac{39 \times 38 \times 37}{51 \times 50 \times 49}.$$

Now we turn to the first question. We intend to compute p'_x from the product rule

$$p_2 = p_1 p'_x , \tag{π}$$

where p_2 is A's (general) probability of having at least two aces and p_1 that of his having at least one ace.

His probability of having no ace at all (p_0) is

$$p_0 = \frac{39 \times 38 \times 37 \times 36}{52 \times 51 \times 50 \times 49},$$

and

$$p_1 = 1 - p_0 .$$

Now we compute $1 - p_2$, which is the prob. of A having not more than one ace. The prob. of his having neither the ace of hearts, nor that of clubs nor that of diamonds is

$$\frac{39 \times 38 \times 37}{52 \times 51 \times 50}.$$

There are three other similar events. The prob. that at least one of the four is true, is obviously $1 - p_2$, so we get by the addition rule

$$1 - p_2 = 4\frac{39 \times 38 \times 37}{52 \times 51 \times 50} - 6p_0 + 4p_0 - p_0$$

$$= 4\left(\frac{39 \times 38 \times 37}{52 \times 51 \times 50} - p_0\right) + p_0$$

$$= 1 - p_x + 1 - p_1 .$$

Thus

$$p_2 = p_1 - (1 - p_x) ,$$

and from (π)

$$p'_x = 1 - \frac{1 - p_x}{p_1} \qquad \text{or} \qquad 1 - p'_x = \frac{1 - p_x}{p_1} .$$

It is seen, that p_x is bigger than p'_x. The numerical values are

$$p_x = \cdot 56117 \qquad\qquad p'_x = \cdot 36967 .$$

What significance has the additional information, that A is holding—of all aces—the ace of spades!? It would, of course, have no significance, if we had asked A: tell us the suit of the ace or of one of the aces you are holding, and he had answered: spades. But the fact that among his aces was the one we chose to ask him about, increases the likelihood of his holding more than one. Indeed, the more aces he has, the greater is the likelihood of his answering yes to our second question. If a bet were intended, one might call it a rather cunning question. For A would have to be cute to realize, what the inquirer can benefit from knowing the suit of A's ace.