

BY THE SAME AUTHOR

INDIAN CURRENCY AND FINANCE
8vo. Pp. viii +263.1913.
7s. 6d. net.

THE ECONOMIC CONSEQUENCES
OF THE PEACE
8vo. Pp. vii +279 . 1919.
8s. 6d. net.

## A TREATISE ON PROBABILITY



MACMILLAN AND CO., LIMITED<br>LONDON • BOMBAY - CALCUTTA • MADRAS MELBOURNF.

THE MACMILLAN COMPANY
NEW YORK - BOSTON • CHICAGO DALLAS • SAN FRANCISCO

THE MACMILIAN CO. OF CANADA, LTD.
TORONTO

## A TREATISE ON PROBABILITY

JOHN MAYNARD KEYNES



COPYRIGHT

## PREFACE

The subject matter of this hook was first broached in the brain of Leilniz, who, in the dissertation, written in his twenty-third year, on the mode of electing the kings of Poland, conceived of Probahility as a branch of Logic. A few years before, "un probleme," in the words of Poisson, "proposé ì un anstere jansíniste par nan homme du momde, at ite lorizine du calloul des probabilités." In the intervening centuries the algehraical exercises, in which the Chevalier de la Méré interested P'ascal, have su far preduminated in the learned world over the prufounder enquiries of the philosopher into those processes of human faculty which, by determining reasonable preference, guide our choice, that I'robability is oftener reckoned with Mathematies than with Logic. There is much here, therefore, which is novel, and, being novel, unsifted, inaceurate, or deficient. I propound my systematic conception of this sulject for criticism and enlargement at the hand of others, doubtful whether I myself am likely to get much further, by waiting longer, with a work, which, heginning as a Fellowship Dissertation, and interrupted by the war, has already extended over many years.

It may be pereeived that I have been much influenced by W. E. Johnson, (. E. Moore, and Bertrand Russell, that is to say, by Cambridge, which, with great dehts to the writers of Continental Europe, yet continues in direct succession the English tradition of Locke and Berkeley and Hume, of Mill and Sidgwick, who, in spite of their diverrences of
doctrine, are united in a preference for what is matter of fact, and have conceived their subject as a branch rather of science than of the creative imagination, prose writers, hoping to be understood.

> J. M. KRYNES.

```
Kma's Conmese, Cambrumem,
    1/r!! 1, 1920.
```


## CONTENTS

PART I<br>

## CHAPTER I

$\therefore .1$
:
CHAP'TER II
 ..... (1)
CHAPTEI: III
 ..... $\because 0$
CHAPTER IV
 ..... 11
(:H.1ITEI ।
 ..... (6.)
("IIAD'IEK: V
 ..... 71

## A TREATISE ON PROBABILITY

## CHAPTER VII <br> CHAPTER VII

Historical Retrospect ..... 79
CHAPTER VIII
The Frequency Theory of Probabllity ..... 92
CHAPTER IX
The Constructive Theory of Part I. summarised . ..... 111
PART II
FUNDAMENTAL THEOREMS
CHAPTER X
Introductory ..... 115
CHAPTER XI
The Theory of Groups, witi special mererbnce to Logical C'onstistence, Inferfince, and Logical Priority ..... 123
CHAPTER XII
The Definitions and Axioms of Infermnce and Probabinity ..... 133
CHAPTER XIII
The Fundamintal Theorems of Necessary Inference. ..... 139
CIIAPTER XIV
The Fundampental Theorems of Probable Inference ..... 144

## CHAPTER XV <br> CHAPTER XV

Nomertcal Meanurement and Approximation of Proba- bilitifs ..... 15
CHAPTER XVI
Observations on the Theormms of Chapter Xly., ant their Detelopments, including Testmony. ..... 164
CHAPTER XVII
Gome Problams in Inverse Probability, inclulding Averagas ..... 186
P.Al'l IH

(H.1ITER NVHI
Intaunection ..... $\therefore 17$
CHAPTER XIX
The Nature of Argument by Analogy ..... $22: 1$
CHAPTER XX
The Vifue of Multidication of Instances, or Ptre I midection ..... 233
CHAPTER XXI
T're Natcre of Indictive Aboument (ontinted ..... 242
CHAPTER XXU
The Justification of these: Methons ..... 251
CHAPTEK XXIII
some Historical Notes on Intucorion ..... 265
Noter on P'art Ill. ..... 274
PART IV
SOME PHILOSOPHICAL APMLICATION゙ OF PROBABHITTY
CHAPTER XXIV
The Meanings of Objective Chance, and of Randomness ..... 281
CHAPTER XXV
Some Problems arising out of the Discussion of (hance ..... 293
CHAPTER XXVI
The Application of Probabinty to Conduct ..... 307
PART V
THE FOUNDATIONS OF STATISTICAL INFERENCE
CHAPTER XXVII
The Nature of Statistical Inference ..... 327
CHAPTER XXVIII
The Law of Great Numbers ..... 3.32
CHAPTER XXIX
The Use of it prioni Probabhiftes for the Prediction of  Porsson, anl) Toberrychiser ..... 337
CHAPTER XXX
The Mathematical lise of Spatisticat Frequencies for the
 of Laflact: ..... 367

## CONTENTS

## CHAPTER XXXI

The Inversion of Bernouthi＇s＇Theorem ..... 111.1 ..... 384
CHAPTER XXXII
  of Lexts ..... 391

（）UTline of a（＇Onstrective Theory ..... 406
 ..... $12!!$
1．いいに ..... 1.19

1) AR'I

FUNDAMENTAL IDEAS

## CHAPTER I

## THE MEANING OF PROBABILITY

 trateroit des degres de Probabiliti."- Lémsiz.

1. Part of our knowledge we obtain direct; and part by argument. The Theory of Probability is concerned with that part which we ohtain be argument, and it treats of the different degrees in which the results so obtained are condusibe or inconclusive.

In mast hranches of academic logice such as the theory of the s. Howism or the weometry of ideal space, all the arguments am at demonstrative certainty. They claim to be conclusive. But many ofher arguments are rational and clam somm weight without pretending to be certain. In Metaphesies, in semenee and in Comduce, most of the arguments. upen which we habithally hase our rational beliefs, are admited to be inconelusise in a greater or less degree. Thus for a philosophical treatment of these branches of knowledge, the study of probability is reguired.

The eourse which the history of thenght has led Lamic tw follow has encouraged the view that douhtularguments are mot within its scope. But in the actual exercise of reason we do not wait on certainty, or deem it irrational to depend on a doubtful argument. If logic investigates the general principles of valid themght. the study of arguments. to which it is ratiomal to attarh somer weight, is as much a part of it as the stuly of these which are demonstrative.
2. The terms metmin and probulde deseribe the varions degrees of rational hedief about a propmsition which different amomots of knowledere authorise us to untertain. All propensitions ate true or false, but the knowledge we have of them depends on our circumstances; and while it is often comsemient to speak of
propositions as certain or probable, this expresses strictly a relationship in which they stand to a corpus of knowledge, actual or hypothetical, and not a characteristic of the propositions in themselves. A proposition is capable at the same time of varying degrees of this relationship, depending upon the knowledge to which it is related, so that it is without significance to call a proposition probable unless we specify the knowledge to which we are relating it.

To this extent, therefore, probability may be called subjective. But in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is rational to entertain in given conditions, and not merely with the actual beliefs of particular individuals, which may or may not be rational.

Given the body of direct knowledge which constitutes our ultimate premisses, this theory tells us what further rational beliefs, certain or probable, can be derived by valid argument from our direct knowledge. This involves purely logical relations between the propositions which embody our direct knowledge and the propositions about which we seek indirect knowledge. What particular propositions we select as the premisses of our argument naturally depends on subjective factors peculiar to ourselves; but the relations, in which other propositions stand to these, and which entitle us to probable beliefs, are objective and logical.
3. Let our premisses consist of any set of propositions $h$, and our conclusion consist of any set of propositions $a$, then, if a knowledge of $h_{i}$ justifies a rational belief in $a$ of degree $a$, we say that there is a probability-relation of degree $a$ between $a$ and $h .{ }^{1}$

In ordinary speech we often describe the conclusion as being doubtful, uncertain. or only probable. But, strictly, these terms ought to be applied, either to the degree of our rational belief in the conclusion. or to the relation or argument between two sets of propositions, knowledge of which would afford grounds for a corresponding degree of rational belief. ${ }^{2}$

[^0]4. With the term "event," which has taken hitherto so important a place in the phraseology of the subject, I shall dispense altogether. ${ }^{1}$ Writers on Probability have senerally dealt with what they term the "happening" of "event.." In the problems which they first studied this did not involve much departure from common usage. But these expressions are now used in a way which is vasue and unambiguous: and it will be more than a verbal improvement to discuss the truth and the probability of propositions instead of the occurrence and the probability of events. ${ }^{2}$
5. These general ideas are not likely to provoke much criticism. In the ordinary course of thought and argument, We are constantly assuming that knowledge of one statement, while mot prorimig the truth of a second, vields nevertheless some ground for believing it. We assert that we ought on the evidence to prefer such and such a belief. We claim rational gromuls for assertions which are not conclusively demonstrated. Wic allow, in fact, that statements may be unproved. without, for that reason. being unfounded. And it does not seem on reflection that the information we convey by these expressions is wholly. suljective. When we argue that Darwin gives valid grounds for our accepting his theory of natural selection, we do mot simply. mean that we are psychologically inclined to agree with him: it is certain that we also intend to convey our belief that we are acting rationally in regarding his theory as probable. We believe that there is some real objective relation between Darwin's evidence and his comelusions, which is indeprondent of the mere fact of our belief, and which is just as real and ohjeetive, though of a different degree, as that which would exist if the argumpht were as demonstrative as a syllowism. II. are clamme, in fact, to congise comectly a lowical commedion between one set of propositions which we call our evidence and which we suppose ourselves to know, and another set which we call our conclusions, and to which we attach there or less weight.
${ }^{1}$ Except in thoso chapters (Chap. XVII., for example) where I am dealing chiefly with the work of others.
${ }^{2}$ The first writer I know of to notice this was Ancillon in lloutes sur les
 venir est probable, c'est dire qu'une proposition est probable." The point was nmphasised by Boole, Laws of Thought, pp. 7 and 167. See also Czubor,


according to the grounds supplied by the first. It is this type of objective relation between sets of propositions - the type which we claim to be correctly perceiving when we make such assertions as these - to which the reader's attention must be directed.
6. It is not straining the use of words to speak of this as the relation of probability. It is true that mathematicians have employed the term in a narrower sense; for they have often confined it to the limited class of instances in which the relation is adapted to an algebraical treatment. But in common usage the word has never received this limitation.

Students of probability in the sense which is meant by the authors of typical treatises on Wahrscheinlichleitsrechmmg or Calcul des probabilités, will find that I do eventually reach topics with which they are familiar. But in making a serious attempt to deal with the fundamental difficulties with which all students of mathematical probabilities have met and which are notoriously unsolved, we must begin at the beginning (or almost at the beginning) and treat our subject widely. As soon as mathematical probability ceases to be the merest algebra or pretends to guide our decisions, it immediately meets with problems against which its own weapons are quite powerless. And even if we wish later on to use probability in a narrow sense, it will be well to know first what it means in the widest.
7. Between two sets of propositions, therefore, there exists a relation, in virtue of which, if we know the first, we can attach to the latter some degree of rational belief. This relation is the subject-matter of the logic of probability.

A great deal of confusion and error has arisen out of a failure to take due account of this relational aspect of probability. From the premisses " $a$ implies $b$ " and " $a$ is true," we can conchude something about $b$ namely that $b$ is true which does not involve $a$. But, if $a$ is so related to $b$, that a knowledge of it renders a probable belief in $b$ rational, we camot conclude anything whatever about $b$ which has not reference to a ; and it is not true that every set of self-consistent premisses which includes $a$ has this same relation to $b$. It is as useless, therefore, to say " $b$ is probable " as it would be to say " $b$ is equal," or " $b$ is greater than," and as unwarranted to conclude that, becatise " makes b probable, therefore " and o together make b
probable, as to argue that because a is less than b, therefore a and $c$ together are less than $b$.

Thus, when in ordinary speech we name some opinion as probable without further qualification, the phrase is generally Alliptical. Wie mean that it is probable when certain considerabtions. implicitly or explicitly present to our minds at the moment. are taken into arcount. We use the word for the sake of shortnesos. just ats we speak of a place as being three miles distant. when we mean three miles distant from where we are then sit nated, or from some starting-point to which we tacitly refer. No proposition is in itself either probable or improbable. just as mo phare can bee intrinsically distant; and the probability of the same statement varies with the evidence presented, which is, as it were, its origin of reference. We may fix our attention on our own knowledge and, treating this as our origin. considar the probabilities of all other suppositions,-according to the usual practice which leads to the elliptical form of common speech ; or we may, equally well, fix it on a proposed conclusion and comsider what degree of probability this would deriwe from varions sits of astumptions. Which might constitute the corpmes of knowledge of ourselves or others, or which are merely hypotheses.

Reflection will show that this account harmonises with familiar experience. There is nothing novel in the supposition that the probatility of a theory turns upen the evidence In which it is supported : and it is common to assert that an opinion was probahme on the widence at first to hand. but on further informa tion was untenable. As our knowledge or our hypothesis changes, our conclusions have new probabilities, not in themselves, but relatively to these new premisses. New logical relations have now become important, namely those between the conclusions which we are investigating and our new assumptions; but the old relations between the conclusions and the former assumptions still exist and are just as real as these new ones. It would be as absurd to deny that an opinion was probable, when at a later stage certain objections have come to light, as to deny, when we have reached our destination, that it was ever three miles distant; and the opinion still is probable in relation to the old hypotheses, just as the destimation is still three miles distant from our starting-point.
8. A definition of probability is not possible, unless it contents us to define degrees of the probability-relation by reference to degrees of rational belief. We cannot analyse the probabilityrelation in terms of simpler ideas. As soon as we have passed from the logic of implication and the categories of truth and falsehood to the logic of probability and the categories of knowledge, ignorance, and rational belief, we are paying attention to a new logical relation in which, although it is logical, we were not previously interested, and which cannot be explained or defined in terms of our previous notions.

This opinion is, from the nature of the case, incapable of positive proof. The presumption in its favour must arise partly out of our failure to find a definition, and partly because the notion presents itself to the mind as something new and independent. If the statement that an opinion was probable on the evidence at first to hand, but became untenable on further information, is not solely concerned with psychological belief, I do not know how the element of logical doubt is to be defined, or how its substance is to be stated, in terms of the other indefinables of formal logic. The attempts at definition, which have been made hitherto, will be criticised in later chapters. I do not believe that any of them accurately represent that particular logical relation which we have in our minds when we speak of the probability of an argument.

In the great majority of cases the term " probable" seems to be used consistently by different persons to describe the same concept. Differences of opinion have not been due, I think, to a radical ambiguity of language. In any case a desire to reduce the indefinables of logic can easily be carried too far. Even if a definition is discoverable in the end, there is no harm in postponing it until our enquiry into the object of definition is far advanced. In the case of "probability" the object before the mind is so familiar that the danger of misdescribing its qualities through lack of a definition is less than if it were a highly abstract entity far removed from the normal channels of thought.
9. This chapter has served briefly to indicate, though not to define, the subject matter of the book. Its object has been to emphasise the existence of a logical relation between two sets of propositions in cases where it is not possible to argue demonstratively from one to the other. This is a contention
of a most fundamental character. It is not entirely novel, lut has seldom received due emphasis, is often overlooked, and sommetimes denied. The view, that probability arises ont of the existence of a specific relation between premiss and conclusion, depends for its acceptance upon a reflective judement on the true character of the concept. It will be our object to discuss. under the title of Probability, the principal properties of this relation. First, however, we must digresis in order to comsider briffly what we mean by knouledye. rotionnl buliff. and dryoment.

## CHAPTER II

## PROBABILITY IN RELATION TO THE THEORY OF KNOWLEDGE

1. I no not wish to become involved in questions of epistemology to which I do not know the answer ; and I am anxious to reach as soon as possible the particular part of philosophy or logic which is the subject of this book. But some explanation is necessary if the reader is to be put in a position to understand the point of view from which the author sets out; I will, therefore, expand some part of what has been outlined or assumed in the first chapter.
2. There is, first of all, the distinction between that part of our belief which is rational and that part which is not. If a man believes something for a reason which is preposterous or for no reason at all, and what he believes turns out to be true for some reason not known to him, he cannot be said to believe it ratiomally, although he believes it and it is in fact true. On the other hand, a man may rationally believe a proposition to be mobable, when it is in fact false. The distinction between rational belief and mere belief, therefore, is not the same as the distinction between true beliefs and false beliefs. The highest degree of rational belief, which is termed certain rational belief. corresponds to knowledge. We may be said to know a thing when we have a certain rational belief in it, and vice versa. For reasons which will appear from our account of probable degrees of rational belief in the following paragraph, it is preferable to regard linouledye as fundamental and to define rational belief by reference to it.
3. We come next to the distinction between that part of our rational belief which is certain and that part which is only probable. Belief, whether rational or not, is capable of degree. The highest degree of rational belief, or rational certainty of
belief. and its relation to knowledge have been introduced above. What. however, is the relation to knowledge of probable deqrers of rational belief?

The proposition (seny, 4 ) that we know in this casse is mot the samm as the propmsition (suy, $p$ ) in which we have a probable degree (suy, a) of rational belief. If the evidence upon which we hase our belief is $h$, then what we linous. namely $q$. is that the propesition $p$ bears the probatility-relation of degree a to the sot of propesitions $h$; and thi knowledere of ours justitios us: in a rational belief of degree a in the proposition $\mu$. It will be convenient to call propesitions sucth as $p$, which donot contain assertions about probability-relations, "primary propositions" ; and propositions such as $q$, which assert the existence of a probability-relation, "secondary propositions." '
4. Thus knowleder of a proposition always corresponds to rertainty of rational belief in it and at the same time to actual trulh in the proposition itself. We camont know a proposition unless it is in fact true. A probable degree of rational belief in a pmpmsition. on the other hand, arises out of knowledere of some corresponding secondary proposition. A man may rationally berlieve a propesition to be probable when it is in fact falsed. if the seremdary propensition on which he depends is true and certain; while a man cannot rationally believe a proposition to be probable even when it is in fact true, if the secondary proposition on which he depends is not true. Thus rational belief of whatever degree can only arise out of knowledge, abthough the knowledge may be of a propesition secombary in the above sense, to the proposition in which the rational degree of belief is entertained.
5. At this point it is desirable to colligate the three senses in which the term probubility has been so far employed. In its most fundamental sense, I think, it refers to the logical relation between two sets of propositions, which in §y of Chapter I. I have termed the probability-relation. It is with this that I shall be mainly concerned in the greater part of this Treatise. Derivative from this sense, we have the sense in which, as above, the 1, rom probulde in applied to the degrees of rational belief arising out of knowledge of secondary propositions which assert the

[^1]existence of probability-relations in the fundamental logical sense. Further it is often convenient, and not necessarily misleading, to apply the term probable to the proposition which is the object of the probable degree of rational belief, and which bears the probability-relation in question to the propositions comprising the evidence.
6. I turn now to the distinction between direct and indirect knowledge--between that part of our rational belief which we know directly and that part which we know by argument.

We start from things, of variou classes, with which we have, what I choose to call without reference to other uses of this term, direct acquaintance. Acquaintance with such things does not in itself constitute knowledge, although knowledge arises out of acquaintance with them. The most important classes of things with which we have direct acquaintance are our own sensations, which we may be said to experience, the ideas or meanings, about which we have thoughts and which we may be said to understand, and facts or characteristics or relations of sense-data or meanings, which we may be said to perceive ;--experience, understanding, and perception being three forms of direct acquaintance.

The objects of knowledge and belief-as opposed to the objects of direct acquaintance which I term sensations, meaning*, and perceptions-I shall term propositions.

Now our knowledge of propositions seems to be obtained in two ways: directly, as the result of contemplating the objects of acquaintance ; and indirectly, by argument, through perceiving the probability-relation of the proposition, about which we seek: knowledge, to other propositions. In the second case, at any rate at first, what we know is not the proposition itself but a secondary proposition involving it. When we know a secondary proposition involving the proposition $p$ as subject, we may be said to have indirect knowledge about $p$.

Indirect lnowledge about $p$ may in suitable conditions lead to rational belief in $p$ of an appropriate degree. If this degree is that of certainty, then we have not merely indirect knowledge about $p$, but indirect knowledge of $p$.
7. Let us take examples of direct knowledge. From acquaintance with a sensation of yellow I can pass directly to a knowledge of the proposition "I have a sensation of yellow." From acyuaintance with a sensation of yellow and with the
meanings of "rellow," ". colour," ." existence," I may be ahle (t) pass to a direct knowledge of the propositions " 1 understand the meaning of yellow." " my sensation of yellow exists." " yellow is a colour." Thus, by some mental process of which it is difficult to give an account, we are able to pass from direct adguaintance with things to a knowledge of propositions about the things of which we have sensations or understand the meaning.

Next, by the contmplation of propositions of which we have direct knowledge, we are able to pass indirectly to knowledue of or about other propesitions. The mental process by which we pass from direct knowledqe to indirect knowledge is in some cases and in some degree capable of analysis. We pass from a knowledge of the proposition a to a knowledge about the proposition $b$ by pereefiving a logical relation between them. With this logical relation we have direct acquaintance. The logie of knowledere is mainly occupied with a study of the logical relations, direct acquaintance with which permits direct knowledge of the seremdary proposition asserting the probability-relation, and so to indient knowledse about, and in some cases of. the primary proposition.

It is not always possible, however, to analyse the mental procesis in the case of indirect knowleduc. or to saty the perreption of whet logical relation we have pasied from the knowleden of one propesition to knowledge about amother. But although in some cases we serem to pass direetly from one propesition to another. I am inclined to believe that in all legitimate tramsitions of this kind some logical relation of the proper kind must exist between the propositions, even when we are not explicitly aware of it. In any case, whenever we pass to knowleder about one propesition by the contemphation of it in prlation to another proposition of which we have kmowedere wen when the provess is unanalysable I eall it an neyment. The kmowledge, such as we have in ordinary thomght by passime from one propesition to another without heing able to say what hesical relations. if any, we hate pereeived between them, mas ber termed uncomplated knowladge. And kmonledtre, which results from a distinct apperfemsion of the relevant legrical relations, may be termed knowledge proper.
8. In this way, therefore, I distinguish between direct and
indirect knowledge, between that part of our rational belief which is based on direct knowledge and that part which is based on argument. About what kinds of things we are capable of knowing propositions directly, it is not easy to say. About our own existence, our own sense-data, some logical ideas, and some logical relations, it is usually agreed that we have direct knowledge. Of the law of gravity, of the appearance of the other side of the moon, of the cure for phthisis, of the contents of Bradshaw, it is usually agreed that we do not have direct knowledge. But many questions are in doubt. Of which logical ideas and relations we have direct acquaintance, as to whether we can ever know directly the existence of other people, and as to when we are knowing propositions about sense-data directly and when we are interpreting them-it is not possithle to give a clear answer. Moreover, there is another and peculiar kind of derivative knowledge-by memory.

At a given moment there is a great deal of our knowledge which we know neither directly nor by argument - we remember it. We may remember it as knowledge, but forget how we originally knew it. What we once knew and now consciously remember, can fairly be called knowledge. But it is not easy to draw the line between conscious memory, unconscious memory or habit, and pure instinct or irrational associations of ideas (acquired or inherited)-the last of which cannot fairly be called knowledge, for unlike the first two it did not even arise (in us at least) out of knowledge. Especially in such a case as that of what our eyes tell us, it is difficult to distinguish between the different ways in which our beliefs have arisen. We cannot always tell, therefore, what is remembered knowledge and what is not knowledge at all ; and when knowledge is remembered, we do not always remember at the same time whether, originally, it. was direct or indirect.

Although it is with knowledge by argument that I shall be mainly concerned in this book there is one kind of direct knowledge, namely of secondary propositions, with which I cannot help but be involved. In the case of every argument, it is only directly that we can know the secondary proposition which makes the argument itself valid and rational. When we know something by argument this must be through direct acquaintance with some logical relation between the conclusion and the premiss.

In all knowledge, therefore, there is some direct element: and logic can never be made purely mechanical. Ill it can do is so to arrange the reasoning that the logical relations. which have to be perceived directly, are made explicit and are of a simple kind.
9. It must be added that the term certainty is sometimes used in a merely psychological sense to describe a state of mind without reference to the logical grounds of the belief. With this sense I am not concerned. It is also used to descreribe the highest degree of rational belief; and this is the sense relectant to our present purpose. The peculiarity of certainty is that linowledge of a secondary proposition involving certaints. together with knowledge of what stands in this secondary propesition in the position of exidence, leads to limouledlye of and not merely about, the rorresponding primary proposition. Knowledge. on the other hand, of a secondary proposition involving a degree of probability lower than certainty, tomether with knowledge of the premiss of the secondary proposition. leads only to a rutiomal belief of the appropriate degtee in the primary proposition. The knowledge present in this latter case I have called knowledge about the primary proposition or conclusion of the argument, as distinct from knowledge of it.

Of probatility we can say no more than that it is a lower demere of rational belief than certainty; and we may say, if we like, that it deals with degrees of certainty. ${ }^{1}$ Or we may make probability the more fundamental of the two and recard certainty as a special case of probability, as being, in fact, the muximmm probability. Speaking somewhat loosely we may say that, if our premisses make the conclusion certain, then it follones from the premisses: and if they make it very probable, then it very nearly follows from them.

It is sometimes useful to use the term " impossibility " as the negative correlative of "certainty," although the former sometimes has a different set of associations. If $a$ is certain, then the contradictory of $a$ is impossible. If a knowledere of a makes $b$ certain, then a knowledere of "makes the contradictory

[^2]of $b$ impossible. Thus a proposition is impossible with respect to a given premiss, if it is disproved by the premiss ; and the relation of impossibility is the relation of mininum probability. ${ }^{1}$
10. We have distinguished between rational belief and irrational belief and also between rational beliefs which are certain in degree and those which are only probable. Knowledge has been distinguished according as it is direct or indirect, according as it is of primary or secondary propositions, and according as it is of or merely about its object.

In order that we may have a rational belief in a proposition $p$ of the degree of certainty, it is necessary that one of two conditions should be fulfilled-(i.) that we know $p$ directly ; or (ii.) that we know a set of propositions $h$, and also know some secondary proposition $q$ asserting a certainty-relation between $p$ and $h$. In the latter case $h$ may include secondary as well as primary propositions, but it is a necessary condition that all the propositions $h$ should be known. In order that we may have rational belief in $p$ of a lower degree of probability than certainty, it is necessary that we know a set of propositions $h$, and also know some secondary proposition $q$ asserting a probability-relation between $p$ and $h$.

In the above account one possibility has been ruled out. It is assumed that we cannot have a rational belief in $p$ of a degree less than certainty except through knowing a secondary proposition of the prescribed type. Such belief can only arise, that is to say, by means of the perception of some probability-relation. To employ a common use of terms (though one inconsistent with the use adopted above), I have assumed that all direct knowledge is certain. All knowledge, that is to say, which is obtained in a manner strictly direct by contemplation of the objects of acquaintance and without any admixture whatever of argument and the contemplation of the logical bearing of any other knowledge on this, corresponds to cerlain rational belief and not to a merely probable degree of rational belief. It is true that there do seem to be degrees of knowledge and rational belief, when the source of

[^3]the belief is solelv in acquaintance, as there are when its source is in argument. But I think that this appearance arises partly out of the difticulty of distinguishing direct from indirect kinowledge, and partly out of a confusion between probuble knowledere and cemue knowledge. I camot attempt here to analyse the meaning of vague knowledge. It is certainly not the same thing as kowledge proper, whether certain or probable, and it does not seem likely that it is susceptible of strict logical treatnent. At any rate I do not know how to deal with it. and in spite of its importance I will not complicate a difficult subject by endeaworing to treat adequately the theory of vague knowledge.

I assume then that only true propositions can be known, that the: term " probable knowledge " ought to be replaced by the term " probahle degree of rational belief," and that a probable degree of rational belief canmot arise directly but only as the result of an argument. out of the knowledge, that is to say, of a secomdary proposition asserting some logical probatilityrelation in which the object of the belief stands to some known proposition. With arguments, if they exist, the ultimate premisises of which are known in some other manner than that describeal above, such as might be called "probable knowledge," my theory is mot adequate to deal without modification. ${ }^{\text {b }}$

For the objects of certain belief which is based on direct knowlelter as opposied to certain belief arising indirectly, there is a well-estathlished expression; propositions, in which our rational belief is both certain and direct, are said to be stlferillent.
11. In conclusion, the relativity of knowledge to the individual may be briefly tour hed on. Some part of knowledge knowledter of our own existence or of our own sensations-is clearly relative to imtisidnal experience. We camot speak of knowledge absolutely-only of the knowledge of a particular person. Other parts of knowledre-knowledge of the axioms of logic, for ex-ample- may sem more objective But we must admit, I think, that this tow is rilative to the constitution of the human mind, and that the conatitution of the human mind may vary in some dearre from man to man. What is silf-evident to me and what

[^4]I really know, may be only a probable belief to you, or may form no part of your rational beliefs at all. And this may be true not only of such things as $m y$ existence, but of some logical axioms also. Some men-indeed it is obriously the case-may have a greater power of logical intuition than others. Further, the difference between some kinds of propositions over which human intuition seems to have power, and some over which it has none, may depend wholly upon the constitution of our minds and have no significance for a perfectly objective logic. We can no more assume that all true secondary propositions are or ought to be universally known than that all true primary propositions. are known. The perceptions of some relations of probability may be outside the powers of some or all of us.

What we know and what probability we can attribute to our rational beliefs is, therefore, subjective in the sense of being relative to the individual. But given the body of premisses which our subjective powers and circumstances supply to us, and given the kinds of logical relations, upon which arguments can be based and which we have the capacity to perceive, the conclusions, which it is rational for us to draw, stand to these premisses in an objective and wholly logical relation. Our logic is concerned with drawing conclusions by a series of steps of certain specified kinds from a limited body of premisses.

With these brief indications as to the relation of Probability, as I understand it, to the Theory of Knowledge, I pass from problems of ultimate analysis and definition, which are not the primary subject matter of this book, to the logical theory and superstructure, which occupies an intermediate position between the ultimate problems and the applications of the theory, whether such applications take a generalised mathematical form or a concrete and particular one. For this purpose it would only encumber the exposition, without adding to its clearness or its accuracy, if I were to employ the perfectly exact terminology and minute refinements of language, which are necessary for the avoidance of error in very fundamental enquiries. While taking pains, therefore, to avoid any divergence between the substance of this chapter and of those which succeed it, and to employ only such periphrases as could be translated, if desired, into perfectly exact language, I shall not cut myself off from the convenient, hut hosere, expressions, which have been habitually employed

## by previous writers and have the advantage of being, in a general way at least, immediately intelligible to the reader. ${ }^{1}$

${ }^{1}$ This question, which faces all contemporary writers on logical and philosophical subjects, is in my opinion much more a question of style-and therefore to be settled on the same sort of considerations as other such questions-than is generally supposed. There are occasions for very exact methods of statement, such as are employed in Mr. Russell's Principia Muthematica. But there are advantages also in writing the English of Hume. Mr. Moore has developed in Principia Ethica an intermediate style which in his hands has force and beauty. But those writers, who strain after exaggerated precision without going the whole hog with Mr. Russell, are sometimes merely pedantic. They lose the reader's attention, and the repetitious complication of their phrases cludes his comprehension, without their really attaining, to compensate, a complete precision. Confusion of thought is not always best avoided by technical and unaccustomed expressions, to which the mind has no immediate reaction of understanding; it is possible, under cover of a careful formalism, to make statements, which, if expressed in plain language, the mind would immediately repudiate. There is much to be said, therefore, in favour of understanding the substance of what you are saying all the time, and of never reducing the substantives of your argument to the mental status of an $x$ or $y$.

## CHAPTER III

## THE MEASUREMENT OF PROBABILITIES

1. I have spoken of probability as being concerned with degrees of rational belief. This phrase implies that it is in some sense quantitative and perhaps capable of measurement. The theory of probable arguments must be much occupied, therefore, with comparisons of the respective weights which attach to different arguments. With this question we will now concern ourselves.

It has been assumed hitherto as a matter of course that probability is, in the full and literal sense of the word, measurable. I shall have to limit, not extend, the popular doctrine. But. keeping my own theories in the background for the moment, I will begin by discussing some existing opinions on the subject.
2. It has been sometimes supposed that a numerical comparison between the degrees of any pair of probabilities is not only conceivable but is actually within our power. Bentham, for instance, in his Rationale of Judicial Evidence, ${ }^{1}$ proposed a scale on which witnesses might mark the degree of their certainty ; and others have suggested seriously a 'barometer of probability.' ${ }^{2}$

That such comparison is theoretically possible, whether or not we are actually competent in every case to make the comparison, has been the generally accepted opinion. The following quotation ${ }^{3}$ puts this point of view very well :
. I do not see on what ground it can be doubted that every
${ }^{1}$ Book i. chap vi. (referred to by Venn).
2 The reader may be reminded of (iibbon's proposal that:-" A Theological Barometer might be formed, of which the Cardinal (Baronius) and our countryman, Dr. Middleton, showld constitute the opposite and remote extremities, as the former sumk to the lowest degree of credulity, which was compatible with iearnins, and the latter ruse to the highest pitch of scepticism, in any wise consistent with Religion."
${ }^{3}$ W. F. Donkin, Phil. Mag., 1851. He is replying to an article by J. D. Forbes (Phil. Mag., Aug. 1849) which had cast doubt upon this opinion.
definite state of belief concerninis a proposed hyputhesis is in itself capable of being represented be a numerical expression, however difficuit or impracticatble it may be to ascertain its actual value. It would he very difiernit to estimete in numbers the ris cim of all of the particles of a human boily at any instant; but no one doubts that it is capable of numerical wapession. I mention this because I ann non sure that Professor Forbes has distinguished the difficulty of ascertaiming mumbers in certain cases from a supposed difliculty of expression by means of mumbers. The former difficulty is real, but merely relative to our knowledge and skill: the latter, if real, would he alsolute and inherent in the subject-matter, which I conceive is not the case."

De Morgan held the same opinion: on the gromel that, wherever we have differences of dorrer, numerical comparison must be theoretically posisible. He assumes, that is to say, that all probabilities can be placed in an orion of mamitmbe, and argues from this that they must be measurahle. Philosophers, however, who are mathematicians, would no lonerer auree that, even if the premiss is somud, the conclusion frillows from it. Oljecets can be arranced in an order, which we can reasonably call ome of dearee or magnitude, without its beime possible for conceive a system of measurement of the differences berween the individuals.

This opinion may also have been held by others, if not by De Morcan, in part because of the marrow associations which Probability has had for them. The Caleulus of Probability has rewived far more attention than its leric, and mathmaticians, under no compulsion to deal with the whole of the subject, have naturally confined their attention to these sperial cases, the existence of which will be demonstrated at a later stage, where aleobraical representation is possible. Probatility has become assoreiated, therefore, in the mimds of thenerits with those prohlems in which we are presented with a number of cxelusive and exhaustive altornatives of equal prohability; ambthe principhes, which are readily applicable in such circumstances, have heen supposed, without much further enquiry, to posisess immoral validity.
3. It is also the case that theories of probability haw heren

[^5]propounded and widely accepted, according to which its numerical character is necessarily involved in the definition. It is often said, for instance, that probability is the ratio of the number of "favourable cases" to the total number of "cases." If this definition is accurate, it follows that every probability can be properly represented by a number and in fact is a number; for a ratio is not a quantity at all. In the case also of definitions based upon statistical frequency, there must be hy definition a numerical ratio corresponding to every prolability. These definitions and the theories based on them will be diseussed in Chapter VIII. ; they are connected with fundamental differences of opinion with which it is not necessary to burden the present argument.
4. If we pass from the opinions of theorists to the experience of practical men, it might perhaps be held that a presumption in favour of the numerical valuation of all probabilities can be based on the practice of underwriters and the willingness of Lloyd's to insure against practically any risk. Underwriters are actually willing, it might be urged, to name a numerical measure in every case, and to back their opinion with money. But this practice shows no more than that many probabilities are greater or less than some numerical measure, not that they themselves are numerically definite. It is sufficient for the underwriter if the premium he names exceeds the probable risk. But, apart from this, I doubt whether in extreme cases the process of thought, through which he goes before naming a premium, is wholly rational and determinate ; or that two equally intelligent brokers acting on the same evidence would always arrive at the same result. In the case, for instance, of insurances effected before a Budget. the figures quoted must be partly arbitrary: There in in them an element of caprice, and the broker's state of mind, when he quotes' a figure, is like a bookmaker's when he names odds. Whilst he may be able to make sure of a profit, on the principles of the bookmaker, yet the individual figures that make up the book are, within certain limits, arbitrary. He may be almost certain, that is to say, that there will not be new taxes on more than one of the articles tea, sugar, and whisky; there may be an opinion abroad, reasonable or unreasonable, that the likelihood is in the order whisky, tea, sugar ; and he may. therefore be able to effect insurances for equal amounts in each
at 30 per cent, 40 per cent, and 45 per cent. He has thus made sure of a profit of 1.5 per cent, however absurd and arbitrary his quotations may be. It is not necessary for the success of underwriting on these lines that the probabilities of these new taxes
 that there should be merehants willing to insure at these rates. These merchants, moreover, may be wise to insure even if the quotations are partle arhitrary; for they may run the risk of insolvency unless their possible loss is thus limited. That the transaction is in principle one of bookmaking is shown be the fact, that, if there is a specially large demand for insurance aquinst one of the pessibilities, the rate rises:- the probability has not chamed, but the ${ }^{*}$ book ${ }^{*}$ is in danger of being upset. I Presidential election in the lonited States supplies a more precise examph: On August 23, 1912, 6if per cent was quoted at Lhoyd's to pay a total loss should Dr. Woodrow Wilson be elected, 30 per cent should Mr. Taft be eleceted, and 20 per cent should Mr. Roorsevelt be elfected. A broker, who could effect insurances in mpual amoments arainst the election of each candidate, would be certain at these rates of a profit of 10 per cent. Sulsequent now lifications of these terms would laresely depend upon the number of applicante for cach kind of policy. Is it possible to maintain that these figures in any way represent reasoned numerical estimates of probability?

In sume insurances the arbitrary element seems even grater. Consider. for instance. the reinsurance rates for the II curutuh, a ressinl which disappeared in south African waters. The lapse of time made rates rise: the departure of shipsin seareh of her made them fall : some nameless wreckage is foumd and there rise: it is remembered that in similar circumstances thiety yars amo a ressel flomed, helphess hut mot seriously damased. for twn months, and ther fall. Can it be pretended that the figures which were quoted from day to day is per exnt. e", per cent, 78 per cent-were rationally determinate, or that the actual figure was mot within wide limits arhitrary and due to the caprice of individuals? In fact underwriters themselses distinumish between risks which are properly insurahle, wither becanse their probahility can be estimated betweon comparatively, narrew numerical limits or because it is possible to mak": a "hook" which covers all possibilities, and other risks which canmot bee
dealt with in this way and which cannot form the basis of a regular business of insurance,-although an oceasional gamble may be indulged in. I believe, therefore, that the practice of underwriters weakens rather than supports the contention that all probabilities can be measured and estimated numericalls:
5. Another set of practical men, the lawyers, have heen more subtle in this matter than the philosophers. ${ }^{1}$ A distinction, interesting for our present purpose, between probabilities, which can be estimated within somewhat narrow limits, and those which cannot, has arisen in a series of judicial decisions respecting damages. The following extract ${ }^{2}$ from the Times Law Reports seems to me to deal very clearly in a mixture of popular and legal phraseology, with the logical point at issue :

This was an action brought by a breeder of racehorses to recover damages for breach of a contract. The contract was that Cyllene, a racehorse owned by the defendant, should in the season of the year 1909 serve one of the plaintifi's brood mares. In the summer of 1908 the defendant, without the consent of the plaintiff, sold Cyllene for $£ 30,000$ to go to South America. The plaintiff claimed a sum equal to the average profit he had made through having a mare served by Cyllene during the past four years. During those four years he had had four colts which had sold at $£ 3300$. Upon that basis his loss came to 700 guineas.

Mr. Justice Jelf said that he was desirous, if he properly could, to find some mode of legally making the defendant compensate the plaintiff; but the question of damages presented formidable and, to his mind, insuperable difficulties. The damages, if any, recoverable here must be either the estimated loss of profit or else nominal damages. The estimate could only be based on a succession of contingencies. Thus it was assumed that (inter alia) Cyllene would be alive and well at the time of the intended service ; that the mare sent would be well bred and not barren ; that she would not slip her foal ; and that the foal would be born alive and healthy. In a case of this kind he could only

[^6]rely on the weiching of chances: and the law gencrally recarded damares which depended on the weighing of chances as tor remote, and therefore irrecoverable. It was drawing the line between an estimate of damage based on probabilities, as in "Simpson $\because$ I. and X.II. Railway (o." (1, (Q.B.D), 2ith), where Cockburn, C'.J. said: "To some extent, no doubt, the damare must be a matter of speculation, but that is mo reason for not awardine any damages at all," and a claim for damames of a totally problematical charactor. He (Mr. Justice Juff) thonerht the present case was well over the line. Having referred to " layme on Damages" (8th ed., p. T(1), he pointed out that in "Watson $\boldsymbol{x}$. Ambereah Railway ( 0 ." (15). Jur., 448) I'atteson. J., seemed to think that the chance of a prize might he takem into account in estimating the damages for breach of a contract to sent a machine for loading bares by railway too late for a show: hut Erle, J., appeared to think such damage was too remote. In his Lordship's view the chance of wimnine a prize was not of sufficiently ascertainable value at the time the contract was made to be within the contemplation of the parties. Further, in the present case, the contingencies were far more numerous and uncertain. He would enter judgment for the plaintifl for nominal damares, which were all he was entitled to. They would be asserssel at 1 s .

One other similar case may be quoted in further elucidation of the same point, and because it also illustrates another point the importance of making clear the assumptions relative to which the probability is calculated. This case ${ }^{1}$ arose out of an offer of a Beanty Prize ${ }^{2}$ by the Daty Express. Out of fomon photomaphis submitted a number were to be selected and published in the newspaper in the following manner :

The United Kingdom was to be divided into districts and the photographes of the selected candidates living in each dist rict were to bee submitted to the readers of the paper in the district, when were to select by their votes those whom they considered the most hemutiful, and a Mr. Seymour Hicks was then to make an appointment with the 5o ladies obtaining the greatest number of roters and himself select. 12 of them. The plaintiff, whe came

[^7]out head of one of the districts, submitted that she had not been given a reasonable opportunity of keeping an appointment, that she had thereby lost the value of her chance of one of the 12 prizes, and claimed damages accordingly. The jury found that the defendant had not taken reasonable means to give the plaintiff an opportunity of presenting herself for selection, and assessed the damages, provided they were capable of assessment, at $₫ 100$, the guestion of the possibility of assessment being postponed. This was argued before MIr. Justice Pickford, and subsequently in the Court of Appeal before Lord Justices Yaughan Williams, Fletcher Moulton, and Farwell. Two questions arose --relative to what evidence ought the probahility to be calculated, and was it numerically measurable? Counsel for the defendant contended that, "if the value of the plaintiff"s chance was to be considered, it, must be the value as it stood at the beginning of the competition, not as it stood after she had been selected as one of the 50 . As 6000 photographs had been sent in, and there was also the personal taste of the defendant as final arbiter to be considered, the value of the chance of success was really incalculable." The first contention that she ought to be considered as one of 6000 not as one of 50 was plainly preposterous and did not hoodwink the court. But the other point, the personal taste of the arbiter, presented more difficulty. In estimating the chance, ought the Court to receive and take account of evidence respecting the arbiter's preferences in types of beauty ? Mr. Tustice Pickford, without illuminating the question, held that the damages were capable of estimation. Lord Justice Vaughan Williams in giving judgment in the Court of Appeal argued as follows:

As he understood it, there were some 50 competitors, and there were 12 prizes of equal value, so that the average chance of success was about one in four. It was then said that the questions which might arise in the minds of the persons who had to dive the decisions were so numerous that it was impossible to apply the doctrine of averages. He did not agree. Then it was said that if precision and certainty were impossible in any case it would be right to describe the damages as unassessable. He agreed that there might be damages so unassessable that the doctrine of averages was not pessible of applieation because the figures necessary to be applied were not forthoming. Several
cases were to be found in the reports where it had been su heht. but he denied the proposition that because precision and certainty had not been arrived at, the jury had no function or duty to determine the damares. . . . He (the Lord Justiee) demind that the mere fact that you could not assess with precision and certainty relieved a wromedoer from praving damages for his lireach of duty. He would not lay down that in every case it eonld he lef: to the jury to assess the damazes: there were cases where the loss was sod demodent on the mere unrestricted volition of anotlows person that it was impossible to arriwe at any assessable loss from the breach. It was true that there was no market here ; the rient to compete was personal and could not bee transforped. Ho conld not admit that a competitor who found herself one of Fil could have gone into the market and sold her right to compete. It the same time the jury micht reasonably have asked themsolues the question whether, if there was a rimht to compere, it could have been transferred, and at what price. Vnder these circumstances he thought the matter was one for the jury.

The attitude of the Lered Justioe is clear. The blaintill had widenty subiared damare, and justice requiren that she shoubd be compmensated. But it was empally evident, that. melatise to the completest information availahle and account home taken of the arthiter's persemal taste, the probability could be he no means "stimatend with mumerical precision. Further. it was impossible tor say how much weight ourht to be attached to the fact that the plaintiff had been leow of her district (there were ferer than an diatricts) : yot it was plain that it made her chamee leeter than the chanmes of these of the in left in, whe were not head of their districts. Let rough justice be done, therefore. let the case be simplified by ignoring some part of the evidence. The "doctrine of averages" is then applicable, or, in other words, the phentifl's losis may he assessed at twelve liftiethe of the value of the prize. ${ }^{1}$
6. How does the matter stand, then ? Whether or not such a thime is thememically conerivable, no exercise of the practical julement is passilla, by which a numerical value cath actmalls be given to the probability of every argument. So far from

[^8]our being able to measure them, it is not even clear that we are always able to place them in an order of magnitude. Nor has any theoretical rule for their evaluation ever been suggested.

The doubt, in view of these facts, whether any two probabilities are in every case even theoretically capable of comparison in terms of numbers, has not, however, received serious consiler:tion. There seems to me to be exceedingly strong reasons for entertaining the doubt. Let us examine a few more instances.
7. Consider an inductiou or a generalisation. It is usually held that each additional instance increases the generalisation's probability. A conclusion, which is based on three experiments in which the unessential conditions are varied, is more trustworthy than if it were based on two. But what reason or principle can be adduced for attributing a numerical measure to the increase ? ${ }^{1}$

Or, to take another class of instances, we may sometimes have some reason for supposing that one object belongs to a certain category if it has points of similarity to other known members of the category (e.g. if we are considering whether a certain picture should be ascribed to a certain painter), and the greater the similarity the greater the probability of our conclusion. But we cannot in these cases measure the increase ; we can say that the presence of certain peculiar marks in a picture increases the probability that the artist of whom those marks are known to be characteristic paintel it, but we cannot say that the presence of these marks makes it two or three or any other number of times more probable than it would have been without them. We can say that one thing is more like a second object than it is like a third ; but there will very seldom be any meaning in saying that it is twice as like. Probability is, so far as measurement is concerned, closely analogous to similarity. ${ }^{2}$
${ }^{1}$ It is true that Laplace and others (even amongst contemporary writers) have believed that the probability of an induction is measurable by means of a formula known as the rule of succession, according to which the probability of an induction based on $n$ instances is $\begin{gathered}n+1 \\ n+2\end{gathered}$. Those who have been convinced by the reasoning employed to establish this rule must be asked to postpone judgment until it has been examined in Chapter XXX. But we may point out here the absurdity of supposing that the odds are 2 to 1 in favour of a generalisation based on a single instance-a conclusion which this formula would seem to justify.
${ }^{2}$ There are very few writers on probability who have explicitly admitted that probabilitics, though in some sense quantitative, may be incapable of

Or consider the ordinary circumstances of life. Wer are out for a walk -what is the mobability that we shall reach home aliwe: Hden this always a numerical measure? If at thunderstorm bursts umon us, the probability is less than it was before: but is it chanced by some definite numerical amount? There minht, of course, be data which would make these probabilities mumerically comparable: it might be argued that a knowledge of the statistics of death by lightning would make such a com parison prosible. But if such information is not included within the knowledere to which the probability is referred, this fact is not rolusant io the probability actually in question and cannot antep its value. In some cases, moreover, where general statistics are availabl. the mumerical probability which mioht be derived from themi is inapplicatle because of the presence of additional knowledee with regrard to the particular case. (iibbon calculated his prospects of life from the volumes of vital statistice and the calculations of acturies. But if a doctor had been called to his assistance the nice precision of these calculations would have become usicless; (iibbon's prospects would have been better or wome than before, but he would no longer have been able to calculate to within a day or week the period fur which he then possessed an even chance of survival.

In these instances we can, perhaps, arrange the probabilities in an order of magnitude and assert that the new datum strenathens or wakens the argument, although there is no basis for an estimate heme much stronger or weaker the new aromment is than the old. But in another class of instamees is it wom pessibl. to arrance the probabilities in ann order of magnitude, or to say that one is the greater and the other less ?
8. Comsider three sots of experiments, each directed towards establishine a ereneralisation. The first set is more numbems;
numerical comparison. Edeeworth, " Philosophy of Chance" (Mind, 1884, p. 225). admitted that " there may well be important quantitative, although not
 nuny, 1. 43) may alsi, be cited as holding a somewhat similar opinion. He maintains that a lack of comparability in the grounds often stands in the way of the measurability of the probable in ordinary usage, and that there are not necessarily good reasons for measuring the value of one argument against thit of anuther. On the other hand, a numerical statement for the degree of the probable, although gencrally imposible, is not in itself contradictory to the notion; and of three statements, relating to the same circumstances, we can well say that one is more probable than another, and that one is the most probathe of the three.
in the second set the irrelevant conditions have been more carefully varied; in the third case the generalisation in view is wider in scope than in the others. Which of these generalisations is on such evidence the most probable ? There is, surely, no answer; there is neither equality nor inequality between them. We cannot always weigh the analogy aqainst the induction, or the scope of the generalisation against the bulk of the evidence in support of it. If we have more grounds than before, comparison is possible; but, if the grounds in the two cases are quite different, even a comparison of more and less. let alone numerical measurement, may be impossible.

This leads up to a contention, which I have heard supported, that, although not all measurements and not all comparisons of probability are within our power, yet we can say in the case of every argument whether it is more or less likely than not. Is our expectation of rain, when we start out for a walk, always more likely than not, or less likely than not, or as likely as not? I am preprared to argue that on some occasions none of these alternatives hold, and that it will be an arbitrary matter to decide for or against the umbrella. If the barometer is high, but the clouds are black, it is not always rational that one should prevail over the other in our minds, or even that we should balance them,though it will be rational to allow caprice to determine us and to waste no time on the debate.
9. Some cases, therefore, there certainly are in which no rational basis has been discovered for numerical comparison. It is not the case here that the method of calculation, prescribed by theory, is beyond our powers or too laborious for actual application. No method of calculation, however impracticable, has been suggested. Nor have we any primu facie indications of the existence of a common unit to which the magnitudes of all probabilities are naturally referrible. A degree of probability is not composed of some homogeneous material, and is not apparently divisible into parts of like character with one another. An assertion, that the magnitude of a given prob ability is in a numerical ratio to the magnitude of every other, seems, therefore, unless it is based on one of the current definitions of probability, with which I shall deal separately in later chapters, to be altogether devoid of the kind of support, which can usually be supplied in the case of quantities of which
the mensurability is not open to denial. It will he worth while, however, to pursue the argument a little further.
10. There appear to be four alternatives. Either in some cases there is no probability at all: or probabilities do not all belone to at single set of magnitudes medsurable in terms of a common unit; or these measures always exist, but in many cases are, and must remuin, unknown; or probabilities do belong to sheh as set and thoir measures are contuble of being determined by us, although we are not always able so to determine them in practice.
11. Laplace and his followers excluded the first two alternatives. They argued that every conclusion has its place in the numerical range of probibilities from "to I, if orl!y we heme it, and the developed their theory of whomen probabilites.

In dealing with this contention, we must be clear as to what we moan by saving that a probability is unkoun. Do we mean unknown throush lack of shill in arguing from given evidence, or unknown through lack of evidence? The first is alone admissible, for mow evidence would wive us a new probatility, not a fuller knowlodere of the old one: we have nont discovered the probahility of a statement on wiven evidenee, he determining its probability in relation to quite different evidence. We must anot allow the theory of unk mon prohabilities to sain platnsibility from the second sense. A relation of probability does not yield us, as a rule, information of much value, unless it invests the (6, molush with a probability which lies be tween narrow munerical limits. In ordinary practice, therefore, we do not always regard ourselves as linowing the probability of a conclusion, unless we can estimate it numerically. We are apt, that is to say, to restrict the use of the expression probable to these numerical examples, and to allege in other cases that the probatility is unknown. We might say, for example, that we do not know, when we go on a railway journey, the probability of death in a railway atecilont. unless we are told the statisties of accibmit. in former years ; or that we do not know our chances in a luttery; unless we are told the number of the tickets. But it must be chear neme refle tim that if we use the term in this semee. Which is no doubt a perfectly lexitimate sense, --we ought to say that in the casne of seme aremments a rilation of probatility comes not exist, and not that it is mnknown. For it is not this prohability
that we have discovered, when the accession of new evidence makes it possible to frame a numerical estimate.

Possibly this theory of unknown probabilities may also gain strength from our practice of estimating arguments, which, as I maintain, have no numerical value, by reference to those that have. We frame two ideal arguments, that is to say, in which the general character of the evidence largely resembles what is actually within our knowledge, but which is so constituted as to yield a numerical value, and we judge that the probability of the actual argument lies between these two. Since our standards, therefore, are referred to numerical measures in many cases where actual measurement is impossible, and since the probability lies between two numerical measures, we come to believe that it must also, if only we knew it, possess such a measure itself.
12. To say, then, that a probability is unknown onght to mean that it is unknown to us through our lack of skill in arguing from given evidence. The evidence justifies a certain degree of knowledge, but the weakness of our reasoning power prevents our knowing what this degree is. At the best, in such cases, we only know vagucly with what degree of probability the premisses invest the conclusion. That probabilities can be unknown in this sense or known with less distinctness than the argument justifies, is clearly the case. We can through stupidity fail to make any estimate of a probability at all, just as we may through the same cause estimate a probability wrongly. As soon as we distinguish between the degree of belief which it is rational to entertain and the degree of belief actually entertained, we have in effect admitted that the true probability is not known to everybody.

But this admission must not be allowed io carry us too far. Probability is, cide Chapter II. (§ 12), relative in a sense to the principles of human reason. The degree of probability, which it is rational for us to entertain, does not presume perfect logical insight, and is relative in part to the secondary propositions which we in fact know ; and it is not dependent upon whether more perfect logical insight is or is not conceivable. It is thie degree of probability to which those logical processes lead, of which our minds are capable ; or, in the language of Chapter II., which those secondary propositions justify, which we in fact know. If we do not take this view of probability, if we do not limit it
in this way and make it, to this extent, relative to human powers. we are altogether adrift in the mbenown for we cannot ever know what degree of probability would be justified by the perception of logical relations which we are, and must always be, incapable of comprehending.
13. Those who have maintained that, where we camot assign a numerical probability, this is not because there is nome, but simply hecause we do not know it, have really meant. I feel sure, that with some addition to our knowledge a numerical value would be assignable. that is to say that our conclutions would have a munerical probability relative to slighty different premisses. Enless, therefore, the reader clings to the opinion that, in erery one of the instances I have cited in the earlier paragraphs of this chapter, it is theoretically pessible on that evidence to assign a numerical value to the probability, we are l. ft with the first two of the alternatives of $\S 10$. which were as frollows: either in some cases there is ne probability at all ; or probabilities do not all belong to a single set of magnitudes measarable in terms of a common unit. It would be difficult to maintain that there is no logical relation whatever between our premiss and our conclusion in those cases where we camot assign a numerical value to the probability; and if this is so. it is really a question of whether the logical relation has chatacteristics, other than mensurability, of a kind to justify us in calling it a probability-relation. Which of the two we favour is therefore partly a matter of definition. We might, that is to say, pick out from probatilities (in the widest sense) a set, if there is one, all of which are measurable in terms of a common unit. and call the members of this set, and them only, probabilities (in the narrow sense). Tor restrict the term "probability" in this way would be, I think, very inconvenient. For it is possible, as I shall show, to find sereral sets, the members of each of which are measurable in terms of a unit common to all the members of that set: so that it would be in some devree arbitrary ${ }^{1}$ which we chose. Further, the distinction between probabilities, which would be thus measurable and those which would not, is not fundamental.

At any rate: I aim here at dealine with probability in its

[^9]widest sense, and am averse to confining its scope to a limited type of argument. If the opinion that not all probabilities can be measured seems paradoxical, it may be due to this divergence from a usage which the reader may expect. Common usage, even if it involves, as a rule, a flavour of numerical measurement, does not consistently exclude those probabilities which are incapable of it. The confused attempts, which have been made, to deal with numerically indeterminate probabilities under the title of unknown probabilities, show how difficult it is to confine the discussion within the intended limits, if the original definition is too narrow.
14. I maintain, then, in what follows, that there are some pairs of probabilities between the members of which no comparison of magnitude is possible ; that we can say, nevertheless, of some pairs of relations of probability that the one is greater and the other less, although it is not possible to measure the difference between them; and that in a very special type of case, to be dealt with later, a meaning can be given to a numerical comparison of magnitude. I think that the results of observation, of which examples have been given earlier in this chapter, are consistent with this account.

By saying that not all probabilities are measurable, I mean that it is not possible to say of every pair of conclusions, about which we have some knowledge, that the degree of our rational belief in one bears any numerical relation to the degree of our rational belief in the other ; and by saying that not all probabilities are comparable in respect of more and less, I mean that it is not always possible to say that the degree of our rational belief in one conclusion is either equal to, greater than, or less than the degree of our belief in another.

We must now examine a philosophical theory of the quantitative properties of probability, which would explain and justify the conclusions, which reflection discovers, if the preceding discussion is correct, in the practice of ordinary argument. We must bear in mind that our theory must apply to all probabilities and not to a limited class only, and that, as we do not adopt a definition of probability which presupposes its numerical mensurability, we cannot directly argue from differences in degree to a numerical measurement of these differences. The problem is subtle and difficult, and the following solution is, therefore,
proposed with hesitation; but I am strongly convinced that something resembling the conclusion here set forth is true.
15. The so-called magnitudes or degrees of knowledge or probability, in virtue of which one is greater and another less, really arise out of an order in which it is possible to place them. Certainty, impossibility, and a probability, which has an intermediate value, for example, constitute an ordered series in which the probability lies between certainty and impossibility. In the same way there may exist a second probability which lies betueen certainty and the first probability. When, therefore, we say that one probability is greater than another, this precisely means that the degree of our rational belief in the first case lies betueen certainty and the degree of our rational beliof in the second case.

On this theory it is easy to see why comparisons of more and less are not always possible. They exist hetween two probabilities, only when they and certainty all lie on the same ordered serius. But if more than one distinct series of probabilitiess exist, then it is clear that only those, which belong to the same series, can be compared. If the attribute 'greater' as applied to one of two terms arises solely out of the relative order of the terms in a series, then comparisons of greater and less must always be possible between terms which are members of the same series, and can never be possible between two terms which are not members of the same series. Some probabilities are not comparable in respect of more and less, because there exists more than one path, so to speak, between proof and disproof, between certainty and impossibility ; and neither of two probabilities, which lie on independent paths. bears to the other and to certainty the relation of 'between' which is necessary for quantitative comparison.

If we are comparing the probabilitios of two arguments, where the conclusion is the same in both and the evidence of one excereds the evidence of the other by the inclusion of some fact which is favourably relevant, in such a case a relation seems clearly to exist between the two in virtue of which one lies nearer to certainty than the other. Several types of aroument can be instanced in which the existence of such a relation is equally apparent. But we cannot assume its presence in every case or in comparing in respect of more and less the probabilities of every pair of arguments.
16. Analogous instances are by no means rare, in which, by a convenient looseness, the phraseology of quantity is misapplied in the same manner as in the case of probability. The simplest example is that of colour. When we describe the colour of one object as bluer than that of another, or say that it has more green in it, we do not mean that there are quantities blue and green of which the object's colour possesses more or less; we mean that the colour has a certain position in an order of colours and that it is nearer some standard colour than is the colour with which we compare it.

Another example is afforded by the cardinal numbers. We say that the number three is greater than the number two, but we do not mean that these numbers are quantities one of which possesses a greater magnitude than the other. The one is greater than the other by reason of its position in the order of numbers ; it is further distant from the origin zero. One number is greater than another if the second number lies between zero and the first.

But the closest analogy is that of similarity. When we say of three objects $A, B$, and $C$ that $B$ is more like $A$ than $C$ is, we mean, not that there is any respect in which $B$ is in itself quantitatively greater than C, but that, if the three objects are placed in an order of similarity, $B$ is nearer to $A$ than $C$ is. There are also, as in the case of probability, different orders of similarity. For instance, a book bound in blue morocco is more like a book bound in red morocco than if it were bound in blue calf ; and a book bound in red calf is more like the book in red morocco than if it were in biue calf. But there may be no comparison between the degree of similarity which exists between books bound in red morocco and blue morocco, and that which exists between books bound in red morocco and red calf. This illustration deserves special attention, as the analogy between orders of similarity and probability is so great that its apprehension will greatly assist that of the ideas I wish to convey. We say that one argument is more probable than another (i.e. nearer to certainty) in the same kind of way as we can describe one object as more like than another to a standard object of comparison.
17. Nothing has been said up to this point which bears on the question whether probabilities are ever capable of numerical comparison. It is true of some types of ordered series that
there are measurable relations of distance between their members as well as order, and that the relation of one of its members to an 'origin' can be numerically compared with the relation of another member to the same origin. But the legitimacy of such comparisons must be matter for special enquiry in each case.

It will not be possible to explain in detail how and in what sense a meaning can sometimes be given to the numerical measurement of probabilities until Part II. is reached. But this chapter will be more complete if I indicate brietly the conclusions at which we shall arrive later. It will be shown that a process of compounding probabilities can be defined with such properties that it can be conveniently called a process of addition. It will sometimes be the case, therefore, that we can say that one probability $\mathbb{C}$ is equal to the sum of two other probabilities A and B , i.e. $(\mathrm{C}-\mathrm{A}: \mathrm{B}$. If in such a case A and B are equal, then we may write this $\mathrm{C}-2 \mathrm{~A}$ and say that C is double A . Similarly if $\mathrm{D}=\mathrm{C}+\mathrm{A}$, we may write $\mathrm{D}=3 \mathrm{~A}$, and so on. We can attach a meaning, therefore, to the equation $\mathrm{P}^{\prime}=n . \mathrm{A}$, where P ' and A are relations of probability, and $n$ is a number. The relation of certainty has been commonly taken as the unit of such comventional measurements. Hence if P represents certainty, we should say, in ordinary languawe, that the magnitude of the probahility A is . It will be shown also that we can define a process, applicable to probabilities, which has the properties of arithmetical multiplication. Where numerical measurement is possible, we can in consequence perform algebraical operations of considerahle complexity. The attention, out of propertion to their real importance, which has been paid, on account of the opportunities of mathematical manipulation which they afford. to the limited class of numerical probabilities, seems to bee a part explanation of the belief, which it is the principal object of this chapter to prove crroneous, that all prohabilities must belong to it.
18. We must look, then, at the quantitative characteristies of probability in the following way. Some sets of probabilities we can place in an ordered series, in which we can say of any pair that one is nearer than the other to certainty, that the argument in one case is nearer proof than in the other, and that there is more reason for one conclusion than for the other. But
we can only build up these ordered series in special cases. If we are given two distinct arguments, there is no general presumption that their two probabilities and certainty can be placed in an order. The burden of establishing the existence of such an order lies on us in each separate case. An endeavour will be made later to explain in a systematic way how and in what circumstances such orders can be established. The argument for the theory here proposed will then be strengthened. For the present it has been shown to be agreeable to common sense to suppose that an order exists in some cases and not in others.
19. Some of the principal properties of ordered series of probabilities are as follows:
(i.) Every probability lies on a path between impossibility and certainty ; it is always true to say of a degree of probability, which is not identical either with impossibility or with certainty, that it lies between them. Thus certainty, impossibility and any other degree of probability form an ordered series. This is the same thing as to say that every argument amounts to proof, or disproof, or occupies an intermediate position.
(ii.) A path or series, composed of degrees of probability, is not in general compact. It is not necessarily true, that is to say, that any pair of probabilities in the same series have a probability between them.
(iii.) The same degree of probability can lie on more than one path (i.e. can belong to more than one series). Hence, if B lies between A and C , and also lies between $\mathrm{A}^{\prime}$ and $\mathrm{C}^{\prime}$, it does not follow that of A and $\mathrm{A}^{\prime}$ either lies between the other and certainty. The fact, that the same probability can belong to more than one distinct series, has its analogy in the case of similarity.
(iv.) If ABC forms an ordered series, B lying between A and C, and BCD forms an ordered series, C lying between $B$ and D , then ABCD forms an ordered series, B lying between A and D.
20. The different series of probabilities and their mutual relations can be most easily pictured by means of a diagram. Let us represent an ordered series by points lying upon a path, all the
points on a given path belonging to the same series. It follows from (i.) that the points 0 and 1 , representiny the relations of imposibibility and certainty, lie on every path, and that all faths lie wholly between these peints. It follows from (iv:) that the same point can lie on more than one path. It is possible, therefore, for pathis to intersect and cross. It follows from (iv.) that the probability represented by a given point is greater than that represented by any other point which can be reached by passing along a path with a motion constantly towards the point of impossibility, and less than that represented by any point which can be reached by moving along a path towards the point of certainty. As there are independent paths there will be some pairs of points representing relations of probability such that we cannot reach one by moving from the other along a path always in the same direction.

Thest properties are illustrated in the annexed diagram. 0 represents impossibility, I certainty, and A a numerically measurable probability intermediate between O and I ; U, $\mathrm{V}, \mathrm{W}, \mathrm{X}, \mathrm{Y}, \mathrm{Z}$ are non-numerical probabilities, of which, however, V is less than the numerical prohability A , and is also less than $W, \mathrm{X}$, and $\mathrm{Y} . \mathrm{X}$ and Y
 are both greater than W , and greater than V , but are not comparable with one another, or with A. V and Z are both less than $\mathrm{W}, \mathrm{X}$, and Y , but are not comparable with one another; U is not quantitatively comparable with any of the prohabilities V. W, X, Y, Z. Probabilities which are numerically conparable will all belong to one series. and the path of this series, which we may call the numerical path or strand, will be represental by o.dI.
21. The chief results which have been reached so far are collected together below, and expressed with precision :-
(i.) There are amongst degrees of probability or rational belief various sets, each set composing an ordered series. These series are ordered by virtue of a relation of 'between.' If $B$ is 'between' $A$ and $C, A B C$ ' form a s.ris.
(ii.) There are two degrees of probability $O$ and I between
which all other probabilities lie. If, that is to say, A is a probability, OAI form a series. O represents impossibility and I certainty.
(iii.) If A lies between O and B , we may write this $\widehat{\mathrm{AB}}$, so that $\widehat{O A}$ and $\widehat{\mathrm{AI}}$ are true for all probabilities.
(iv.) If $A B$, the probability $B$ is said to be greater than the probability A , and this can be expressed by $\mathrm{B}>\mathrm{A}$. (v.) If the conclusion $a$ bears the relation of probability P to the premiss $h$, or if, in other words, the hypothesis $h$ invests the conclusion $a$ with probability P , this may be written $a \mathrm{P} h$. It may also be written $a / h=\mathrm{P}$.
This latter expression, which proves to be the more useful of the two for most purposes, is of fundamental importance. If aPh and $a^{\prime} \mathrm{P} h^{\prime}$, i.e. if the probability of $a$ relative to $h$ is the same as the probability of $a^{\prime}$ relative to $h^{\prime}$, this may be written $a / h=a^{\prime} / h^{\prime}$. The value of the symbol $a / h$, which represents what is called by other writers 'the probability of ",' lies in the fact that it contains explicit reference to the dutu to which the probability relates the conclusion, and avoids the numerous errors which have arisen out of the omission of this reference.

## CHAPTER IV

## THE PRINCCIPLE OF INDIFFERENCE

Absolute. 'Sure, Sir, this is not very reasonable, to summon my affection for a lady I know nothing of.'
Sir Anthony. 'I am sure, Sir, 'tis more unreasonable in you to object to a lady you know nothing of.' ${ }^{1}$

1. In the last chapter it was assumed that in some cases the probabilities of two arguments may be equal. It was also argued that there are other cases in which one probability is, in some sense. Irreater than another. But so far there has been nothing to show how we are to know when two probabilities are equal or unequal. The recosenition of equality, when it exists, will be dealt with in this chapter, and the recorgnition of inequality in the next. In historical account of the various theories about this prohlem, which have been held from time to time, will be given in Chapter VII.
2. The determination of equality between probabilities has recerised hitherto much more attention than the determination of inerguality. This has been due to the stress which has been laid on the mathematical side of the subject. In order that num rical measurement may be possible, we must be given a number of equally probable alternatives. The discovery of a rule, We which equiprobahility could be established, was, therefore, "nsmitial. I rule, adequate to the purpose, introduced by James Bermoulli. who was the real founder of mathematical protatility: ${ }^{2}$ has been widely adopted, generally under the title of The P'rimeiple of Rom-shafficient Reason. down to the present time. This description is clumsy and unsatisfactory, and, if it is justifiathe to brak away from tradition, I prefer to call it The Principle of Indifference.
[^10]The Principle of Indifference asserts that if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability. Thus equal probabilities must be assigned to each of several arguments, if there is an absence of positive ground for assigning unequal ones.

This rule, as it stands, may lead to paradoxical and even contradictory conclusions. I propose to criticise it in detail, and then to consider whether any valid modification of it is discoverable. For several of the criticisms which follow I am much indebted to Von Kries's Die Principien der Wahrscheinlichkeit. ${ }^{1}$
3. If every probability was necessarily either greater than, equal to, or less than any other, the Principle of Indifference would be plausible. For if the evidence affords no ground for attributing unequal probabilities to the alternative predications, it seems to follow that they must be equal. If, on the other hand, there need be neither equality nor inequality between probabilities, this method of reasoning fails. Apart, however, from this objection, which is based on the arguments of Chapter III., the plausibility of the principle will be most easily shaken by an exhibition of the contradictions which it involves. These fall under three or four distinct heads. In §§ 4-9 my criticism will be purely destructive, and I shall not attempt in these paragraphs to indicate my own way out of the difficulties.
4. Consider a proposition, about the subject of which we know only the meaning, and about the truth of which, as applied to this subject, we possess no external relevant evidence. It has been held that there are here two exhaustive and exclusive alternatives - the truth of the proposition and the truth of its contradictory--while our knowledge of the subject affords no ground for preferring one to the other. Thus if $a$ and $\bar{a}$ are contradictories, about the subject of which we have no outside knowledge, it is inferred that the probability of each is $\frac{1}{2} \cdot{ }^{2}$ In

[^11]the same way the prohabilities of two other propositions, $b$ and $c$, having the same subject as $u$, may be cach $\frac{1}{2}$. But without having any evidence bearing on the subject of these propositions we may know that the predicates are contraries amonyst themselves, and, therefore, exclusive alternatives - a supposition which leads by means of the same principle to values inconsistent with those just obtained. If, for instance, having no evidence relevant to the colour of this book, we could conclude that $\frac{1}{2}$ is the probiability of 'This book is red,' we could conclude equally that the probability of each of the propositions "This book is black' and 'This book is blue ' is also $\frac{1}{2}$. So that we are faced with the impossible case of three exclusive alternatives all as likely as not. A defender of the Principle of Indifference might rejoin that we are assuming knowledge of the proposition: 'Two different colours cannot be predicated of the same subject at the same time'; and that, if we know this, it constitutes relevant outside evidence. But such evidence is about the predicate, not about the subject. Thus the defender of the Principle will be driven on, either to confine it to cases where we know nothing about either the subject or the predicate, which would be to emasculate it for all practical purposes, or else to revise and amplify it, which is what we propose to do ourselves.

The difficulty cannot be met by saying that we must know and take account of the number of possible contraries. For the number of contraries to any propesition on any evidence is always infinite ; $\bar{a}^{\prime}$, is contrary to $a$ for all values of $b$. The same point can he put in a form which does not involve contraries or contradictories. For example, $a_{i} h=\frac{1}{2}$ and $a b / h=\frac{1}{2}$, if $h$ is

[^12]irrelevant both to $a$ and to $b$, in the sense required by the crude Principle of Indifference. ${ }^{1}$ It follows from this that, if $a$ is true, $b$ must be true also. If it follows from the absence of positive data that 'A is a red book' has a probability of $\frac{1}{2}$, and that the probability of 'A is red ' is also $\frac{1}{2}$, then we may deduce that, if A is red, it must certainly be a book.

We may take it, then, that the probability of a proposition, about the subject of which we have no extraneous evidence, is not necessarily $\frac{1}{2}$. Whether or not this conclusion discredits the Principle of Indifference, it is important on its own account, and will help later on to confute some famous conclusions of Laplace's school.
5. Objection can now be made in a somewhat different shape. Let us suppose as before that there is no positive evidence relating to the subjects of the propositions under examination which would lead us to discriminate in any way between certain alternative predicates. If, to take an example, we have no information whatever as to the area or population of the countries of the world, a man is as likely to be an inhabitant of Great Britain as of France, there being no reason to prefer one alternative to the other. ${ }^{2} \mathrm{He}$ is also as likely to be an inhabitant of Ireland as of France. And on the same principle he is as likely to be an inhabitant of the British Isles as of France. And yet these conclusions are plainly inconsistent. For our first two propositions together yield the conclusion that he is twice as likely to be an inhabitant of the British Isles as of France.

Unless we argue, as I do not think we can, that the knowledge that the British Isles are composed of Great Britain and Ireland is a ground for supposing that a man is more likely to inhabit them than France, there is no way out of the contradiction. It is not plausible to maintain, when we are considering the relative populations of different areas, that the number of names of subdivisions which are within our knowledge, is, in the absence of any evidence as to their size, a piece of relevant evidence.

At any rate, many other similar examples could be invented,

## ${ }^{1} a / h$ stands for ' the probability of $a$ on hypothesis $h$.'

${ }^{2}$ This example raises a difficulty similar to that raised by Von Kries's example of the meteor. Stumpf has propounded an invalid solution of Von Kries's difficulty. Against the example proposed here, Stumpf's solution has less plausibility than against Von Kries's.
which would require a special explanation in each case ; for the above is an instance of a perfectly general difficulty. The possible alternatives may be $a, b, c$, and $d$, and there may be no means of diseriminating between them; but equally there may be no means of discriminating between ( $a$ or $b$ ), $c$, and $d$. This difficulty could be made striking in a variety of ways, but it will be better to criticise the principle further from a somewhat different side.
6. Consider the sperific volume of a given substance. ${ }^{1}$ Leet us suppose that we know the sperific volume to lie between 1 and 3 , but that we have no information as to whereabouts in this interval its exact value is to be found. The Principle of Indifference would allow us to assume that it is as likely to lie between 1 and 2 as between 2 and 3: for there is no reason for supposing that it lies in one interval rather than in the other. But now consider the specific density. The specific density is the reciprocal of the specific volume, so that if the latter is $v$ the former is Our data remaining as before, we know that the specific density must lie between I and $\frac{1}{3}$, and, by the same use of the Principle of Indifference as before, that it is as likely to lie between 1 and $\frac{2}{3}$ as between $\frac{2}{3}$ and $\frac{1}{3}$. But the specifie volume being a determinate function of the specific density, if the latter lies between 1 and $\frac{2}{3}$, the former lies between 1 and $\frac{1}{2}$, and if the latter lies between $\frac{2}{3}$ and $\frac{1}{3}$, the former lies between $1 \frac{1}{2}$ and 3 . It follows, therefore, that the specific volume is as likely to lie between 1 and $1 \frac{1}{2}$ as between $1 \frac{1}{2}$ and 3 ; whereas we have already proved, relatively to precisely the same data, that it is as likely to lie between 1 and 2 as between 2 and 3. Horeover, any other function of the specific volume would have suited our purpose equally well, and by a suitable choice of this function we might have proved in a similar manner that any division whatever of the interval 1 to 3 yields sub-intervals of equal probability. Specific volume and specific density are simply alternative methods of measuring the same objective quantity; and there are many methods which might be adopted, cach yielding on the application of the Principle of Indifference a different probability for a given objective variation in the quantity. ${ }^{2}$

[^13]The arbitrary nature of particular methods of measurement of this and of many other physical quantities is easily explained. The objective quality measured may not, strictly speaking, possess numerical quantitativeness, although it has the properties necessary for measurement by means of correlation with numbers. The values which it can assume may be capable of being ranged in an order, and it will sometimes happen that the series which is thus formed is continuous, so that a value can always be found whose order in the series is between any two selected values ; but it does not follow from this that there is any meaning in the assertion that one value is twice another value. The relations of continuous order can exist between the terms of a series of values, without the relations of numerical quantitativeness necessarily existing also, and in such cases we can adopt a largely arbitrary measure of the successive terms, which yields results which may be satisfactory for many purposes, those, for instance, of mathematical physics, though not for those of probability. This method is to select some other series of quantities or numbers, each of the terms of which corresponds in order to one and only one of the terms of the series which we wish to measure, For instance, the series of characteristics, differing in degree, which are measured by specific volume, have this relation to the series of numerical ratios between the volumes of equal masses of the substances, the specific volumes of which are in question, and of water. They have it also to the corresponding ratios which give rise to the measure of specific density. But these only yield conventional measurements, and the numbers with which we correlate the

[^14]terms which we wish to measure can be selected in a variety of ways. It follows that equal intervals between the numbers which represent the ratios do not necessarily correspond to equal intervals between the qualities under measurement; for these numerical differences depend upon which convention of measurement we have selected.
7. A somewhat analoquas difficulty arises in connection with the problems of what is known as 'geometrical' or 'local' prohatility. ${ }^{1}$ In these problems we are concerned with the position of a peint or infinitesimal area or volume within a continumm." The number of cases here is indefinite, but the Principle of Indifference has been held to justify the supposition that equal lengths or areas or volumes of the continum are in the absence of discriminating evidence, equally likely to contain the point. It has long been known that this assumption leads in numerous cases to contradictory conclusions. If, for instance, two points A and $\mathrm{A}^{\prime}$ are taken at random on the surface of a sphere, and we seek the probability that the lesser of the two ares of the great circle $A \Lambda^{\prime}$ is lesis than $n$. we get one resilt by assuming that the probability of a points lying on a given portion of the sphere's surface is proportional to the area of that pertion, and another result hy assuming that, if a point lies on a given great circle. the probability of its lying on a given are of that circle is propertional to the length of the are, each of these assumptions being equally justified by the Principle of Indifference.

Or consider the following problem : if a chord in a circle is drawn at random, what is the probability that it will be less than the side of the inscribed equilateral triangle. One can argue :-
(a) It is indifferent at what point one end of the chord lies. If we suppose this end fixed, the direction is then

[^15]chosen at random. In this case the answer is easily shown to be $\frac{2}{3}$.
(b) It is indifferent in what direction we suppose the chord to lie. Beginning with this apparently not less justifiable assumption, we find that the answer is $\frac{1}{2}$.
(c) To choose a chord at random, one must choose its middle point at random. If the chord is to be less than the side of the inscribed equilateral triangle, the middle point must be at a greater distance from the centre than half the radius. But the area at a greater distance than this is $\frac{3}{4}$ of the whole. Hence our answer is $\frac{3}{4}$. ${ }^{1}$

In general, if $x$ and $f(x)$ are both continuous variables, varying always in the same or in the opposite sense, and $x$ must lie between $a$ and $b$, then the probability that $x$ lies between $c$ and $d$, where $a<c<d<b$, seems to be $\frac{d-c}{b-a}$, and the probability that $f(x)$ lies between $f(c)$ and $f(d)$ to be $\begin{aligned} & f(d)-f(c) \\ & f(b)-f(a)\end{aligned}$. These expressions, which represent the probabilities of necessarily concordant conclusions, are not, as they ought to be, equal. ${ }^{2}$
8. More than one attempt has been made to separate the cases in which the Principle of Indifference can be legitimately applied to examples of geometrical probability from those in which it cannot. M. Borel argues that the mathematician can define the geometrical probability that a point MI lies on a certain segment PQ of AD as proportional to the length of the segment, but that this definition is conventional until its consequences have been confirmed à posteriori by their conformity with the results of empirical observation. He points out that in actual cases there are generally some considerations present which lead us to prefer one of the possible assumptions to the others. Whether or not this is so, the proposed procedure amounts to an abandonment of the Principle of Indifference as a valid criterion, and leaves our choice undetermined when further evidence is not forthcoming.
M. Poincaré, who also held that judgments of equiprobability in such cases depend upon a 'convention,' endeavoured to mini-

$$
{ }^{1} \text { Bertrand, Calcul des probabilités, p. } 5 .
$$

${ }^{2}$ See (e.g.) Borel, Eléments de la théorie des probabilités, p. 85.
mise the importance of the arbitrary element by showing that, under certain conditions, the result is independent of the particular convention which is chosen. Instead of assuming that the point is equally likely to lie in every infinitesimal interval d.c we may represent the probability of its lying in this interval by the function $\phi(x) d x$. II. Poincaré showed that, in the game of rouge et noir, for instance, where we have a number of compartments arranged in a circle coloured alternately black and white, if we can assume that $\phi(x)$ is a regular function, continuous and with continuous differential coefficients, then, whatever the particular form of the function, the probability of black is approximately equal to that of white. ${ }^{1}$

Whether or not investigations on these lines prove to have a prametical value, they have not, I think, any theoretical importance. If, as I maintain, the probability $\phi(x)$ is not necessarily numerical, it is not a generally justifiable assumption to take its continuity for granted. We have, in the particular example quoted, a number of alternatives, half of which lead to black and half to white ; the assumption of continuity amounts to the assumption that for every white alternative there is a black alternative whose probability is very nearly equal to that of the white. Naturally in such a case we can get an approximately equal probability for the whites as a whole and for the blackis as a whole, without assuming equal probability for each alternative individually. But this fact has no bearing on the theoretical difficulties which we are discussing.
II. Bertrand is so much impressed by the contradictions of geometrical probability that he wishes to exclude all examples in which the number of alternatives is infinite. ${ }^{2}$ It will be argued in the serquel that something resembling this is true. The discussion of this question will be resumed in §§ 21-25.
9. There is yet another group of cases, distinct in character from thase considered so far, in which the principle does not seem to provide us with unambiguous guidance. The typical example is that of an urn contanining black and white badls in an

[^16]unknown proportion. ${ }^{1}$ The Principle of Indifference can be claimed to support the most usual hypothesis, namely, that all possible numerical ratios of black and white are equally probable. But we might equally well assume that all possible constitutions ${ }^{2}$ of the system of balls are equally probable, so that each individual ball is assumed equally likely to be black or white. It would follow from this that an approximately equal number of black and white balls is more probable than a large excess of one colour. On this hypothesis, moreover, the drawing of one ball and the resulting knowledge of its colour leaves unaltered the probabilities of the various possible constitutions of the rest of the bag ; whereas on the first hypothesis knowledge of the colour of one ball, drawn and not replaced, manifestly alters the probability of the colour of the next ball to be drawn. Either of these hypotheses seems to satisfy the Principle of Indifference, and a believer in the absolute validity of the principle will doubtless adopt that one which enters his mind first. ${ }^{3}$

The same point is very clearly illustrated by an example which I take from Von Kries. Two cards, chosen from different packs, are placed face downwards on the table; one is taken up and found to be of a black suit: what is the chance that the other is black also ? One would naturally reply that the chance is even. But this is based on the supposition, relatively unpopular with writers on the subject, that every 'constitution' is equally probable, i.e. that each individual card is as likely to be black as red. If we preler this assumption, we must relin-

[^17]quish the text-book theory that the drawing of a black ball from an urn, containing black and white balls in unknown proportions, affects our knowledge as to the proportion of black and white amongst the remaining balls.

The alternative-or text-book-theory assumes that there are three equal posisibilities- one of each colour, both black, both red. If both cards are black, we are twice as likely to turn up a back card than if only one is black. After we have turned up a black, the probability that the other is black is, therefore, twice as great as the probability that it is red. The chance of the second's being black is therefore $\frac{2}{3} .{ }^{1}$ The Principle of Indifference has nothing to say against either solution. Cntil some further criterion has been proposed we seem compelled to agree with Poincare that a preference for either hypothesis is wholly arhitrary.
10. Such, then, are the kinds of result to which an unguarded use of the Principle of Indifference may lead us. The difficulties, to which attention has been drawn, have been noticed before: but the discredit has not been emphatically thrown on the original soures of error. Yet the principle certainly remains as a negutive criterion; two propositions camot be equally probable, so long as there is any ground for discriminating between them. The principle is a necessary, but not, as it seems, a sufficient condition.

The enunciation of some sufficient rule is certainly essential if we are to make any progress in the subject. But the difficulty of discovering a correct principle is considerable. This difficulty is partly responsible, I think, for the doubts which philosophers and many others have often felt regardine any practical application of the Calculus. Many candid persons, when confronted with the results of Probability, feel a strong sense of the uncertainty of the logical basis upon which it seems to rest. It is difficult to find an intelligible account of the meaning of " probability,' or of how we are ever to determine the probability of any particular propesition ; and yet treatises on the subject profess to arrive at complicated results of the greatest precision and the most profound practical importance.

The incautions methods and exagoerated claims of the school of Laplace have undoubtedly contributed towards the existence of these sentiments. But the general seepticism, which I believe

[^18]to be much more widely spread than the literature of the subject admits, is more fundamental. In this matter Hume need not have felt " affrighted and confounded with that forelorn solitude, in which I am placed in my philosophy," or have fancied himself "some strange uncouth monster, who not being able to mingle and unite in society, has been expell'd all human commerce, and left utterly abandon'd and disconsolate." In his views on probability, he stands for the plain man against the sophisms and ingenuities of " metaphysicians, logicians, mathematicians, and even theologians."

Yet such scepticism goes too far. The judgments of probability, upon which we depend for almost all our beliefs in matters of experience, undoubtedly depend on a strong psychological propensity in us to consider objects in a particular light. But this is no ground for supposing that they are nothing more than " lively imaginations." The same is true of the judgments in virtue of which we assent to other logical arguments ; and yet in such cases we believe that there may be present some element of objective validity, transcending the psychological impulsion, with which primarily we are presented. So also in the case of probability, we may believe that our judgments can penetrate into the real world, even though their credentials are subjective.
11. We must now inquire how far it is possible to rebabilitate the Principle of Indifference or find a substitute for it. There are several distinct difficulties which need attention in a discussion of the problems raised in the preceding paragraphs. Our first object must be to make the Principle itself more precise by disclosing how far its application is mechanical and how far it involves an appeal to logical intuition.
12. Without compromising the objective character of relations of probability, we must nevertheless admit that there is little likelihood of our discovering a method of recognising particular probabilities, without any assistance whatever from intuition or direct judgment. Inasmuch as it is always assumed that we can sometimes judge directly that a conclusion follows from a premiss, it is no great extension of this assumption to suppose that we can sometimes recognise that a conclusion partially follows from, or stands in a relation of probability to, a premiss. Moreover, the failure to explain or define 'probability' in terms of other logical notions, creates a presumption that particular relations
of probability must bee in the first instance directly recognised as such, and camot bee evolved be rule out of data which themselves contain no statements of probability.

On the other hand, alt hough we cannot exclude every element of direct judement, these judements may be limited and controlled, perhaps, hy logical rules and principles which possess a general application. While we may possess a faculty of direct recornition of many relations of probalility, as in the case of many other logical relations, yet some may be much more easily recognisable than others. The object of a logical system of probability is to enable us to know the relations, which cannot be easily perceived, he means of other relations which we can recognise more distinctly-to convert, in fact, vague knowledge into more distinct knowledge. ${ }^{1}$
13. Let us seek to distinguish between the element of direct judgment and the element of mechanical rule in the Principle of Indifference. The enunciation of this principle, as it is ordinarily expressed, cloaks, but does not avoid, the former element. It is in part a formula and in part an appeal to direct inspection: hut in addition to the obscurity and ambinuity of the formula, the appeal to intuition is not as explicit as it should be. The principle states that 'there must be no known reason for preferring one of a set of alternatives to any other.' What does this mean? What are 'reasons,' and how are we to know whether they do or do not justify us in preferring one alternative to another ? I do not know any discussion of Probability in which this question has been so much as asked. If, for example, we are considering the probability of drawing a black ball from an urn containing halls which are

[^19]black and white, we assume that the difference of colour between the balls is not a reason for preferring either alternative. But how do we know this, unless by a judgment that, on the evidence in hand, our knowledge of the colours of the balls is irrelevant to the probability in question? We know of some respects in which the alternatives differ ; but we judge that a knowledge of these differences is not relevant. If, on the other hand, we were taking the balls out of the urn with a magnet, and knew that the black balls were of iron and the white of tin, we might regard the fact, that a ball was iron and not tin, as very important in determining the probability of its being drawn. Before, then, we can begin to apply the Principle of Indifference, we must have made a number of direct judgments to the effect that the probabilities under consideration are unaffected by the inclusion in the evidence of certain particular details. We have no right to say of any known difference between the two alternatives that it is ' no reason ' for preferring one of them, unless we have judged that a knowledge of this difference is irrelevant to the probability in question.
14. A bricf digression is now necessary, in order to introduce some new terms. There are in general two principal types of probabilities, the magnitudes of which we seek to compare,those in which the evidence is the same and the conclusions different, and those in which the evidence is different but the conclusion the same. Other types of comparison may be required, but these two are by far the commonest. In the first we compare the likelihood of two conclusions on given evidence ; in the second we consider what difference a change of evidence makes to the likelihood of a given conclusion. In symbolic language we may wish to compare $x / h$ with $y / h$, or $x / h$ with $x / h_{1} h$. We may call the first type judgments of preference, or, when there is equality between $x / h$ and $y / h$, of indifference; and the second type we may call judgments of relevance, or, when there is equality between $x / h$ and $x / h_{1} h$, of irrelecance. In the first we consider whether or not $x$ is to be preferred to $y$ on evidence $h$; in the second we consider whether the addition of $h_{1}$ to evidence $h$ is relevant to $x$.

The Principle of Indifference endeavours to formulate a rule which will justify judgments of indifference. But the rule that there must be no ground for preferring one alternative to another,
involves. if it is to be a griding rule at all, and mot a petition principii, an appeal to judgments of irrelerance.

The simplest definition of Irrelevance is as follows: $h_{1}$ is irrelevant to $x$ on evidence $h$. if the prohability of $y$ on evidence $h h_{1}$ is the same as its probalility on evidence h. ${ }^{1}$ But for a reason which will appear in Chapter VI., a stricter and more complicated definition, as follows is theoretically preferable: $h_{1}$ is irrelecant to $r$ on evidence $l$, if there is no propesition, inferrible from $h_{1} h$ but not from $h$, such that its addition to evidence $h$ affects the probability of $x^{2}$ Any proposition which is irrelevant in the strict semse is, of course, also irrelevant in the simpler semse: hut if we were to adopt the simpler definition it would sometimes oecur that is part of evidence would be relevant, which taken as a whole was irrelerant. The more elaborate definition by avoiding this proves in the sequel more convenient. If the condition $x_{i} h_{1} h=r \cdot h_{h}$ alone is satistied, we may say that the evidence $h_{1}$ is "irrelevant as a whole.' ${ }^{3}$

It will be convenient to define also two other phrases. $h_{1}$ and $h_{2}$ are independent and complementary parts of the evidence. if between them they make up $h$ and neither can be inferred from the wher. If $r$ is the conclusion, and $h_{1}$ and $h_{2}$ are independent and complementary parts of the evidence, then $h_{1}$ is relevant if the addition of it to $h_{2}$ affects the probability of $x .^{4}$

Some propositions regarding irrelevance will be proved in Part II. If $h_{1}$ is the contradictory of $h_{1}$ and $s_{i} / h_{1} h^{\prime} s_{i} h$, then shik or h. Thus the contradictory of irrelevant evidence is also irrelevant. Also, if $s^{\prime} / \mathrm{gh}^{\prime} x_{i} h$, it follows that $y_{1}^{\prime}$ ah $y^{\prime} h$. H.nne if, on initial evidence $h, y$ is irrelevant to $r$, then, on the same intial evidenee, $x$ is irrelevant to $y$, ie. if in a given state of kmwleden one oecurrence has no bearing on another, then equally the second has no bearing on the first.
15. This distinction enables us to formulate the Principle of Indifference at any rate more precisely. There must be no relerent evidence relating to one alternative, unless there is correspemdong evidonce relating to the other; our relevant

[^20]evidence, that is to say, must be symmetrical with regard to the alternatives, and must be applicable to each in the same manner. This is the rule at which the Principle of Indifference somewhat obscurely aims. We must first determine what parts of our evidence are relevant on the whole by a series of judgments of relevance, not easily reduced to rule, of the type described above. If this relevant evidence is of the same form for both alternatives, then the Principle authorises a judgment of indifference.
16. This rule can be expressed more precisely in symbolic language. Let us assume, to begin with, that the alternative conclusions are expressible in the forms $\phi(a)$ and $\phi(b)$, where $\phi(x)$ is a propositional function. ${ }^{1}$ The difference between them, that is to say, can be represented in terms of a single variable.

The Principle of Indifference is applicable to the alternatives $\phi(a)$ and $\phi(b)$, when the evidence $h$ is so constituted that, if $f(a)$ is an independent part of $h$ (see § 14) which is relevant to $\phi(a)$, and does not contain any independent parts which are irrelevant to $\phi(a)$, then $h$ includes $f(b)$ also.

The rule can be extended by successive steps to cases in which we have more than one variable. We can, if the necessary conditions are fulfilled, successively compare the probabilities of $\phi\left(a_{1} a_{2}\right)$ and $\phi\left(b_{1} a_{2}\right)$, and of $\phi\left(b_{1} a_{2}\right)$ and $\phi\left(b_{1} b_{2}\right)$, and establish equality between $\phi\left(a_{1} a_{2}\right)$ and $\phi\left(b_{1} b_{2}\right)$.

This elucidation is suited to most of the cases to which the Principle of Indifference is ordinarily applied. Thus in the favourite examples in which balls are drawn from urns, we can infer from our evidence no relevant proposition about white balls, such that we cannot infer a corresponding proposition about black balls. Most of the examples, to which the mathematical theory of chances has been applied, and which depend upon the Principle of Indifference, can be arranged, I think, in the forms which the rule requires as formulated above.
17. We can now clear up the difficulties which arose over the group of cases dealt with in $\S 9$, the typical example of which was the problem of the urn containing black and white balls in an unknown proportion. This more precise enunciation of the Principle enables us to show that of the two solutions the equiprobability of each 'constitution' is alone legitimate, and the

[^21]equiprobahility of each numerical ratio erromeous. Let us write the alternative 'The proportion of black balls is $x$ " $\equiv \phi(x)$, and the datum' There are $n$ balls in the hag, with regard to mone of which it is known whether thev are black or white $\equiv h$. On the 'ratio' hypothesis it is argued that the Principhe of Indifference justifies the judgment of indifference, $\phi(. r)^{\prime} / h$ $\phi(y) / h$. In order that this may be valid, it must be possible to state the relovant evidence in the form $f\left(r^{\prime}\right) f(y)$. But this is not the case. If $x=\frac{1}{2}$ and $y=\frac{1}{4}$, we have relevant knowledge about the way in which a proportion of black balls of one half can arise, which is not identical with our knowledge of the way in which a proportion of one quarter can arise. If there are four halls, $\mathrm{A}, \mathrm{B},(\mathrm{C}, \mathrm{l})$, one half are black, if $\mathrm{A}, \mathrm{B}$ or $\mathrm{A}, \mathrm{C}$ or $\mathrm{A}, \mathrm{D}$ or B, ( or $\mathrm{B}, \mathrm{D}$ ) or ( ${ }^{\prime}$, D are black; and one quarter are black, if A or B or C or I ) are black. These propositions are not identical in form. and only by a false judgment of irrelevance can we ignore them. On the 'constitution' hypothesis, however, where A, B black and A, ( black are treated as distinct alternatives, this want of symmetry in our relevant evidence cannot arise.
18. We can also deal with the point which was illustrated by the difficulty raised in §4. We considered there the probabilities of $a$ and its contradictory $\bar{a}$ when there is no external evidence relevant to either. What exactly do we mean by saying that there is no relevant evidence? Is the addition of the word esternul simificant? If " represents a particular proposition, we must know something about it, namely, its meaning. May not the apprehension of its meaning afford us some relevant evidenee! If so, such evidence must not be excluded. If, then, we say that there is no relevant evidence, we must mean no evidence beyond what arises from the mere apprehension of the meaning of the symbol $a$. If we attach no meaning to the symbel, it is uselesis to discuss the value of the probatility; for the probability, which belongs to a propesition as an oljeect of knowledxt, mot ats a form of words, canot in such a case exist.

What exactly does the symbel a stand for in the ahove? Does it stand for any proposition of which we know no more than that it is a proposition? Or doees it stand for is particular proposition which we understand but of which we know no more than is involved in understanding it? In the former casse we:
cannot extend our result to a proposition of which we know even the meaning ; for we should then know more than that it is a proposition; and in the latter case we cannot say what the probability of $a$ is as compared with that of its contradictory, until we know what particular proposition it stands for ; for, as we have seen, the proposition itself may supply relevant evidence.

This suggests that a source of much confusion may lie in the use of symbols and the notion of variables in probability. In the logic of implication, which deals not with probability but with truth, what is true of a variable must be equally true of all instances of the variable. In Probability, on the other hand, we must be on our guard wherever a variable occurs. In Implication we may conclude that $\psi$ is true of anything of which $\phi$ is true. In Probability we may conclude no more than that $\psi$ is probable of anything of which we only know that $\phi$ is true of it. If $x$ stands for anything of which $\phi(x)$ is true, as soon as we substitute in probability any particular value, whose meaning we know, for $x$, the value of the probability may be affected; for knowledge, which was irrelevant before, may now become relevant. Take the following example: Does $\phi(1) / \psi(a)=$ $\phi(b) / \psi(b)$ ? That is to say, is the probability of $\phi$ 's being true of $a$, given only that $\psi$ is true of $a$, equal to the probability of $\phi$ 's being true of $b$, given only that $\psi$ is true of $b$ ? If this simply means that the probability of an object's satisfying $\phi$ about, which nothing is known except that it satisfies $\psi$ is equal to ditto ditto, the equation is an identity. For in this case $\phi(1) / \psi(a)$ means the same as $\phi(b) / \psi(b)$, i.e. we know nothing about $x$ and y except that they satisfy $\psi$, and there is nothing whatever by which we can distinguish $a$ from $b$. But if $a$ and $b$ represent specific entities, which we can distinguish, then the equality does not necessarily hold. If, for instance, $\boldsymbol{\phi}(x)$ stands for ' $x$ is Socrates,' then it is plainly false that $\phi(a) / \psi(a)=\phi(b) / \psi(b)$, where $a$ stands for Socrates and $b$ does not.
19. Bearing this danger in mind, we can now give further precision to the enunciation of the Principle of Indifference given in §16. Our knowledge of the meaning of a must be taken account of so fur us it is relcorent; and the Principle is only satisfied if we have corresponding knowledge about the meaning of $b$. Thus $\phi(a) / h=\phi(b) / h$ may be true for one pair of values $a, b$, and not true for another pair of values $a^{\prime}, b^{\prime}$.

This makes it possible to explain in part the contradiction discussed in \$4. Even if it were true that the probability of $a$ is $\frac{1}{2}$, when we know nothing except that a is a proposition, it does not follow that the probability of 'This book is red ' is $\frac{1}{2}$, when we know the meanings of 'book 'and 'red,' even if we know no more than this. Knowledqe arising directly out of acquaintance with the meaning of red 'may be sufficient to enable us to infer that 'red' and 'not-red' are not satisfactory alternatives to which to apply the Principle of Indifference. How this may come about will be discussed in $\S \S 20,21$.

But the contradictions are not yet really solved: for some of the difficulties discussed in § 4 can arise even when we know no more of 'o and $b$ than that they are difficom propositions. In fact, although we have now stated more clearly than before how the Principle should be enunciated. it is not yet posisible to explain or to avoid all the contradictions to which it hed us in $\$ \$ 4$ to 7. For this purpose we must proceed to a further qualification.
20. The examples, in which the Principle of Indifference broke down, had a great deal in common. We broke up the field of possibility, as we may term it, into a number of areas by a series of disjunctive judgments. But the alternative areas were mot ultimate. They were capable of further subdivision into other areas similur in limel to the former. The paradoxus and contradictions arose, in each case, when the alternatives. which the Principle of Indifference treated as equivalent, act nally. contained or might contain a different or an indefinite number of more elementary units.

In the type of cases in which the Principle of Indifference seemed to permit the assertion that, in the absence of relevant evidence, a proposition is as likely as its contradictory, its comtradictory is mot an ultimate and indivisible alternatiee (in the sense to be explained in § 21 below), wem if the proposition itself satisfies this condition. For its contradictory can be disjunctisele resolved into an indefinite number of sets of contraries to the proposition. It was out of this that our difliculties first arense. "This book is not red' includes amonest others the altermatives 'This book is black' and 'This book is blue.' It is not, therefore, an ultimate alternative.

In the same way the contradietion of $\$ 5$ arose out of the pessibility of splittine the alternatives 'He inhabits the Pritish

Isles' into the sub-alternatives ' He inhabits Ireland or he inhabits Great Britain.' And in the third type of case, to which the example of specific volume and density belongs, the alternative ' $v$ lies in the interval 1 to 2 ' can be broken up into the sub-alternatives ' $v$ lies in the interval 1 to $1 \frac{1}{2}$ or $1 \frac{1}{2}$ to 2 .'
21. This, then, seems to point the way to the qualification of which we are in search. We must enunciate some formal rule which will exclude those cases, in which one of the alternatives involved is itself a disjunction of sub-alternatives of the same form. For this purpose the following condition is proposed.

Let the alternatives, the equiprobability of which we seek to establish by means of the Principle of Indifference, be $\phi\left(a_{1}\right)$, $\phi\left(a_{2}\right) \cdots \phi\left(a_{r}\right),{ }^{1}$ and let the evidence be $h$. Then it is a necessary condition for the application of the principle, that these should be, relatively to the evidence, indivisible alternatives of the form $\phi(x)$. We may define a divisible alternative in the following manner :

An alternative $\phi\left(a_{r}\right)$ is divisible if

$$
\begin{aligned}
& \text { (i.) }\left[\phi\left(a_{r}\right) \equiv \phi\left(a_{r^{\prime}}\right)+\phi\left(a_{r^{\prime \prime}}\right)\right] / h=1, \\
& \text { (ii.) } \phi\left(a_{r^{\prime}}\right) . \phi\left(a_{r^{\prime \prime}}^{\prime \prime} / h=0,\right. \\
& \text { (iii.) } \phi\left(a_{r^{\prime}}\right)^{\prime} / h \neq 0 \text { and } \phi\left(\mu_{r^{\prime \prime}}\right) / h \neq 0
\end{aligned}
$$

The condition that the sub-alternatives must be of the same form as the original alternatives, i.e. expressible by means of the same propositional function $\phi(x)$, deserves attention. It might be the case that the original alternatives had nothing substantial in common; i.e. $\phi(x) \equiv x$ is the only propositional function common to all of them, the alternatives being $a_{1}, a_{2}, \ldots, a_{r}$. In these circumstances the condition in question cannot be satisfied. For the proposition $a_{r}$ can always be resolved into the disjunction $a_{i} b+a_{r} \bar{b}$, where $b$ is any proposition and $\bar{b}$ its contradictory. If, on the other hand, the alternatives which we are comparing can be expressed in the forms $\phi\left(a_{1}\right)$ and $\phi\left(a_{2}\right)$, where the function $\phi(x)$ is distinct from $x$, it is not necessarily the case that either of these can be resolved into a disjunctive combination of terms which can be expressed in their turn in the same form.

Dispensing with symbolism, we can express these conditions as follows: Our knowledge must not enable us to split up the

[^22]alternative $\phi\left(\tau_{r}\right)$ into a disjunction of two sub-alternatives, (i.) which are themselves expressible in the same form $\phi$, (ii.) which are mutually exclusive, and (iii.) which, on the evidence, are possible.

In short, the Principle of Indifterence is not applicable to a pair of alternatives, if we know that either of them is caprable of being further split up into a pair of posisible but incompatible alternatives of the same form as the original pair.

22 . This rule commends itself to common sense. If we know that the $t w o$ alternatives are compounded of a diflerent number or of an indefinite number of sub-alternatives which are in other respects similar, so far as our evidence goes to the original alternatives, then this is a relevant fact of which we must take account. And as it affects the two alternatives in differing and unsymmetrical ways, it breaks down the fundamental condition for the valid application of the Irinciple of Indifference.

Neither this consideration nor that discussed in $\$ \S 18$ and 19 substantially modify the Principle of Indifference as enunciated in §16. They have only served to make explicit what was always implicit in the Principle, by explaining the mamer in which our knowledge of the form ame meaning of the alternatives may be a relevant part of the cvidence. The apparent contradictions arnse from paying attention to what we may term the extruneons. evidence only, to the neglect of such part of the evidence as bore upon the form and meaning of the alternatives.
23. The application of this result to the examples cited in $\S 18$ is not difficult. It exchudes the class of cases in which a proposition and its contradictory constitute the alumatives. For if $b$ is the proposition and $\dot{\delta}$ its contradictory, we cannot find a propesitional function $\phi(x)$ which will satisfy the necessary conditions. It deals also with the type of comtradiction which arose in considwring the probability that an individual taken at random was an inhabitant of a given region. If, on the other hand, the term 'country' is so defined that one count ry c:anmot include two countries, then an individual is, relatively to suitable hypotheses, as likely to bee an inhabitant of one as of anmether. For the function $\phi(r)$, where $\phi(r)$... 'the individual is an inhabitant of country $r$ satisties the conditions. And it deabls with the example of ranges of spectific volume and speecitic density,
because there is no range which does not contain within itself two similar ranges. As there are in this case no definite units by which we can define equal ranges, the device, which will be referred to in $\S 25$ for dealing with geometrical probabilities, is not available.
24. It is worth while to add that the qualification of § 21 is fatal to the practical utility of the Principle of Indifference in those cases only in which it is possible to find no ultimate alternatives which satisfy the conditions. For if the original alternatives each comprise a definite number of indivisible and indifferent sub-alternatives, we can compute their probabilities. It is often the case, however, that we cannot by any process of finite subdivision arrive at indivisible sub-alternatives, or that, if we can, they are not on the evidence indifferent. In the examples given above, for instance, where $\phi(x) \equiv x$, or where $x$ is a part of unspecified magnitude in a continuum, there are no indivisible sub-alternatives. The first type comprises all cases, amongst others, in which we weigh the probabilities of a proposition and its contradictory; and the second includes a great number of cases in which physical or geometrical quantities are involved.
25. We can now return to the numerous paradoxes which arise in the study of geometrical probability (see $\S \S 7,8$ ). The qualification of $\S 21$ enables us, I think, to discover the source of the confusion. Our alternatives in these problems relate to certain areas or segments or arcs, and however small the elements are which we adopt as our alternatives, they are made up of yet smaller elements which would also serve as alternatives. Our rule, therefore, is not satisfied, and, as long as we enunciate them in this shape, we cannot employ the Principle of Indifference. But it is easy in most cases to discover another set of alternatives which do satisfy the condition, and which will often serve our purpose equally well. Suppose, for instance, that a point lies on a line of length $m . l$., we may write the alternative ' the interval of length $l$ on which the point lies is the $x$ th interval of that length as we move along the line from left to right' $\equiv \phi(x)$; and the Principle of Indifference can then be applied safely to the $m$ alternatives $\phi(1), \phi(2) \ldots \phi(m)$, the number $m$ increasing as the length $l$ of the intervals is diminished. There is no reason why $l$ should not be of any definite length however small.

If we deal with the problems of geometrical probability in
this way, we shall aroid the contradictory conclusions, which arise from confusing towe ther distinct elementary areas. In the prohlem. for instance. of the chord drawn at ramdem in a cirche. which is discussed in § 7, the chord is regarded, not as a onedimensional line. hut as the limit of an area, the shape of which is different in each of the variant solutions. In the first solution it is the limit of a triangle. the length of the base of which tends (1) zoro: in the secomd solution it is the limit of a quadrilateral, two of the sides of which are parallel and at a distance apart which tends to zuro: and in the third solution the area is defined he the limiting position of a central section of undefined shape. These distinct hyportheses lead inevitably to difterent results. If we were dealing with a strictly linear chord, the Principle of Indifference would yield us no result, as we could not enunciate the alternatives in the required form ; and if the chord is an elementary area, we must know the shape of the area of which it is the limit. So long as we are careful to enunciate the alternatives in a form to which the Principle of Indifference can bee applied unambiguously. We shall be presented from confusing 1:ate heer distinet problems, and shall be able to reach conclusions in geometrical probability which are unambiguously valid.

The substance of this explanation can be put in a slightly difieront way by saying that it is mot a matter of mdifference in these cases in what manner we proceed to the limit. We must assigh the probabilities be fore proceeding to the limit, which we can do unambiguously. But if the problem in hand does not stop at small finite lengths, areas, or volumes, and we hase to prowed to the limit, then the fimal result depends upen the shape in which the body approaches the limit. Hathemati rims will reemenise an analogy between this cass and the determination of potential at points within a conductor. Its value dupmen upen the shape of the area which in the limit represents the point.
26. The positive contributions of this chapter to the determination of valid judgments of equiprobability are two. In the first phawe we. have stated the Principle of Indifterence in a mere: accurate form, by dinplaying its necessary dependence upon judgments of relevance and so bringing out the hidden element of direct judgment or intuition, which it has always involved. It hat been shown that the I'rineiphe lays dewn a rule bey which
direct judgments of relevance and irrelevance can lead on to judgments of preference and indifference. In the second place, some types of consideration, which are in fact relevant, but which are in danger of being overlooked, have been brought into prominence. By this means it has been possible to avoid the various types of doubtful and contradictory conclusions to which the Principle seemed to lead, so long as we applied it without due qualification.

## CHAPTER V

OTHER METHODS OF DETERMINING PROBABILITIES

1. The recognition of the fact, that not all probabilities are numerical, limits the scope of the Principle of Indifference. It has always been agreed that a numerical measure can actually be obtained in those cases only in which a reduction to a set of exclusine and whams ise oquipublule altematives is practicable. Our previous conclusion that numerical measurement is often imposible armes very well, therefore, with the argument of the precedting chaperer that the rules, in virtue of which we can assisert "rquipmobibily, are somewhat limited in their fied of application.

But the recognition of this same fact makes it more necessary to discuss the principts which will justify comparisons of more and less between probabilities, where numerical measurement is theoretically, as well as practically, impossible. We must, for the reasons given in the preceding chapter, rely in the last resort on direct judgment. The object of the following rules and principles is to reduce the judgments of preference and relevance, which we are comperdido make, to a few relatione simple types. ${ }^{1}$
2. We will enquire first in what circumstances we can expect a comparison of more and less to be theoretically possible. I am inclined to think that this is a matter about which, rather unexpectedly perhaps, we are able to lay down definite rules. We are able, I think, always to compare a pair of probabilities which are
or
(i.) of the type $a b / h$ and $a / h$,
(ii.) of the type $a / h h_{1}$ and $a / h$,
provided the additional evidence $h_{1}$ contains only one independent piece of relevant information.
${ }^{1}$ Parts of Chap. XV. are closely connected with the topics of the following paragraphs, and the discussion which is commenced here is concluded there.
(i.) The propositions of Part II. will enable us to prove that

$$
a b / h<a / h \text { unless } b / a h=1 ;
$$

that is to say, the probability of our conclusion is diminished by the addition to it of something, which on the hypothesis of our argument cannot be inferred from it. This proposition will be self-evident to the reader. The rule, that the probability of two propositions jointly is, in general, less than that of either of them separately, includes the rule that the attribution of a more specialised concept is less probable than the attribution of a less specialised concept.
(ii.) This condition requires a little more explanation. It states that the probability $a / h h_{1}$ is always greater than, equal to, or less than the probability $a / h$, if $h_{1}$ contains no pair of complementary and independent parts ${ }^{1}$ both relevant to $a / h$. If $h_{1}$ is favourable, $a / h h_{1}>a / h$. Similarly, if $h_{2}$ is favourable to $a / h h_{1}$, $a / h h_{1} h_{2}>u / h h_{1}$. The reverse holds if $h_{1}$ and $h_{2}$ are unfavourable. Thus we can compare $a / h h^{\prime}$ and $a / h$, in every case in which the relevant independent parts of the additional evidence $h^{\prime}$ are either all favourable, or all unfavourable. In cases in which our additional evidence is equivocal, part taken by itself being favourable and part unfavourable, comparison is not necessarily possible. In ordinary language we may assert that, according to our rule, the addition to our evidence of a single fact always has a definite bearing on our conclusion. It either leaves its probability unaffected and is irrelevant, or it has a definitely favourable or unfavourable bearing, being favourably or unfavourably relevant. It cannot affect the conclusion in an indefinite way, which allows no comparison between the two probabilities. But if the addition of one fact is favourable, and the addition of a second is unfavourable, it is not necessarily possible to compare the probability of our original argument with its probability when it has been modified by the addition of both the new facts.

Other comparisons are possible by a combination of these two principles with the Principle of Indifference. We may find, for instance, that $a / h h_{1}>a / h$, that $a / h=b / h$, that $b / h>b / h h_{2}$, and that, therefore, $a / h h_{1}>b / h h_{2}$. We have thus obtained a comparison between a pair of probabilities, which are not of the types discussed above, but without the introduction

[^23]of any fresh principle. Wie may denote comparisons of this type by (iii.).
3. Whether any comparisons are possible which do not fall within any of the catmontes (i.), (ii.), or (iii.), I do not feel certain. We undoubtedly make a number of direet comparisons which do not seem to be cosered by them. Wie julqe it more probable. for instance, that (aesar invaded Britain than that Romulus founded Rome. But erem in such cases as this, where a reduction into the resular form is not obvious, it might prove possible if we could clearly analyse the real grounds of our judement. We might argue in this instance that, whereas Romulus's foundine of Rome rests suldy on tradtion, we have in cuddition evidence of another kind for 'aesar's inrasion of Britain, and that, in so far as our belief in Caesar's invasion rests on tradition, we have reasons of a pmecisely similar kind as for our belief in Romulus without the additional doutn inwolsed in the mainmance of a fradition betwern the times of Romulus and Caesar. By some such analwis as this our judgment of comparison might be brought within the above categories.

The process of reaching a judgment of comparison in this way may be called 'schmatisation.' ${ }^{1}$ We take initially an ideal scheme which falls within the categories of comparison. Let us represent 'the historical tradition $x$ has been handed down from a date many sears previous to the time of (aesar' by $\psi_{1}(x)$; the historical tradition $x$ has been handed down from the time of Caesar' by $\psi_{2}(x)$; 'the historical tradition $x$ has extra-traditional support' by $\left.\psi_{3}(\cdot)^{\prime}\right)$ : and the two traditions. the Romulus tradition and the Caesar tradition respectively: by $a$ and $b$. Then if our relevant evidence $h$ were of the form $\left.\left.\psi_{1}^{(1)} \psi_{2}(h)\right)_{s_{s}}(l)\right)$. it is easily seen that the comparison a'h<b, $/ 1$ rould be justified on the lines laid down atoove. ${ }^{2}$ I further juder ment, that our actual evidence prosented no relevant divergence from this schematic form. Wrould then establish the practical conclusion. As I ann not aware of any plausible judgment of comparison which we make in common practioe. but which is clearly incapable of reduction to some schematic form, and as I see no locical hasio for surh a comparison. I feel justified in

[^24]doubting the possibitity of comparing the probabilities of arguments dissimilar in form and incapable of schematic reduction. But the point must remain very doubtful until this part of the subject has received a more prolonged consideration.
4. Category (ii.) is very wide, and evidently covers a great variety of cases. If we are to establish general principles of argument and so avoid excessive dependence on direct individual judgments of relevance, we must discover some new and nore particular principles included within it. Two of these-those of Analogy and of Induction-are excessively important, and will be the subject of Part III. of this book. In addition to these a few criteria will be examined and established in Chapter XIV., $\S \S 4$ and 8 (49.1). We must be content here (pending the symbolic developments of Part II.) with the two observations following :
(1) The addition of new ${ }^{1}$ evidence $h_{1}$ to a doubtful ${ }^{2}$ argument a/h is favourably relevant, if either of the following conditions is fulfilled :- $(a)$ if $a / h \hbar_{1}=0 ;(b)$ if $a / h h_{1}=1$. Divested of symbolism, this merely amounts to a statement that a piece of evidence is favourable if, in conjunction with the previous evidence, it is either a necessary or a sufficient condition for the truth of our conclusion.
(2) It might plausibly be supposed that evidence would be favourable to our conclusion which is favourable to favourable evidence- i.e. that, if $h_{1}$ is favourable to $x / h$ and $x$ is favourable to $a / h, h_{1}$ is favourable to a! h. Whilst, however, this argument is frequently emploved under conditions, which, if explicitly stated, would justify it, there are also conditions in which this is not so, so that it is not necessarily valid. For the very deceptive fallacy involved in the above supposition, Mr. Johnson has suggested to me the name of the Fallucy of the Middle Term. The general question-If $h_{1}$ is favourable to $x / h$ and $x$ is favourable to $a / h$, in what conditions is $h_{1}$ favourable to $a / h$ ? $\cdots$ will be examined in Chapter XIV. $\$ \$ 4$ and 8 (49.1). In the meantime, the intuition of the reader towards the fallacy may be assisted by the following observations, which are due to Mr. Johnson:

Let $x, x^{\prime}, x^{\prime \prime} \ldots$ be exclusive and exhaustive alternatives under datum $h$. Let. $h_{1}$ and $a$ be concordunt in regard to each of

[^25]${ }^{2}$ The argument is doublful so long as $a / h$ is neither certain nor impossible.
these alternatives: i.e. any hyputheis which is strmothened by $h_{1}$ will strengthen $n^{2}$, and any hypothenis which is weakened by $h_{1}$ will weaken $a$. It is obvious that, if $h_{1}$ strengthens some of the hypotheses $x, x^{\prime}, x^{\prime \prime}$. ., it will weaken others. This fact helps us to see why we cannot consider the concordauce of $h_{1}$ and a in regard to one simgle altemative, but must be able to assert their concordance with regard to erity one of the exchasive and exhaustive alternatives, including the particular one taken. But a further condition is needed, which (as we shath show) is obviously satistied in two typical problems at least. This further condition is that, for each hypothesis $x, x^{\prime}, x^{\prime \prime}$. . ., it shall hold that, were this hypotiesis known to be true, the knowledge of $h_{1}$ would not weaken the probability of $a$.

These two conditions are sufficient to ensure that $h_{1}$ shall strengthen $a$ (independently of knowledge of $x, x^{\prime}, x^{\prime \prime} \ldots$ ); and, in a sense, they appear to be mexsontiy; for, untess they are satisfied, the dependence of $h_{1}$ upon " would be (so to speak) accidental as regards the "middle terms,' $\left(x, x^{\prime}, x^{\prime \prime}\right.$. . .).

The necessity for reference to all the alternatives $x, x^{\prime}, x^{\prime \prime} \ldots$ is analogous to the requirement of distribution of the middle term in ordinary syllogism. Thus, from premises " All P ' is $x$, all $s$ is $s$," the conclusion that $"$ is are $\mathrm{l}^{\prime \prime}$ does not formally follow ; but given " all P is $x$ and all S is $x^{\prime}$ " it does follow that " no S are P ", where $x$ ' is any contrary to $x$. The two conditions taken together would be anahognos w the argument: all $x \mathrm{~S}$ is P ; all $x^{\prime} \mathrm{S}$ is P ; all $x^{\prime \prime} \mathrm{S}$ is P ; . . . therefore all S is P .

First Typical Problem. - An urn contains an unknown proportion wh different! coloured balls. I ball is drawn and replaced. Then $x, x^{\prime}, x^{\prime \prime} \ldots$ stand for the various possible proportions. Let $h_{1}$ mean " a white ball has been drawn"; and let a mean "a white ball will be again drawn." Then any hyputhersis which is strengthened by $h_{1}$ will strengthen $a$; and any hypothesis which is weakened by $h_{1}$ will weaken $a$. Moreover, were any one of these hypotheses known to be true, the knowledge of $h_{1}$ would not waken the probability of 1 . Hance. in the absomes of definite knowledge as regards $x, x^{\prime}, x^{\prime \prime}$. . ., the knowledge of $h_{1}$ would strengthen the probability of $a$.

S'cond T'ypical I'roblemb.- Let a certan rvent have: taken phace; which may have been $x, x^{\prime}, x^{\prime \prime}$ or . . . Let $h_{1}$ mean that A reports so and so ; and let a mean that 13 reports similarly or
identically. The phrase similarly merely indicates that any hypothesis as to the actual fact, which would be strengthened by A's report, would be strengthened by B's report. Of course. even if the reports were verbally identical, A's evidence would not necessarily strengthen the hypothesis in an equal degree with B 's ; because A and IS may be unequally expert or intelligent. Now, in such cases, we may further aftirm (in general), that, were the actual nature of the event known, the knowledge of A's report on it would not weaken (though it also need not strengthen) the probability that B would give a similar report. Hence, in the absence of such knowledge, the knowledge of $h_{1}$ would strengthen the probability of $a$.
5. Before leaving this part of the argument we must emphasise the part played by direct judgment in the theory here presented. The rules for the determination of equality and inequality between probabilities all depend upon it at some point. This seems to me quite unaroidable. But I do not feel that we should regard it as a weakness. For we have seen that most, and perhaps all, cases can be determined by the application of general principles to one simple type of direct judgment. No more is asked of the intuitive power applied to particular cases than to deternine whether a new piece of evidence tells, on the whole, for or against a given conclusion. The application of the rules involves no wider assumptions than those of other branches of logic.

While it is important, in establishing a control of direct judgment by general principles, not to conceal its presence, yet the fact that we ultimately depend upon an intuition need not lead us to suppose that our conclusions have, therefore, no basis in reason, or that they are as subjective in validity as they are in origin. It is reasonable to maintain with the logicians of the Port Royal that we may draw a conclusion which is truly probable by paying attention to all the circunstances which accompany the case, and we must admit with as little concern as possible Hume's taunt that " when we give the preference to one set of arguments above another, we do nothing but decide from our feeling concerning the superiority of their influence."


## (HAPTEI: リI

## THE WEIGHT OF ARGCMENTS

1. The gitestion to he ratised in this chapter is somewhat nowel: after much consideration I remain uncertain as to how much importance to attach to it. The magnitude of the probability of an ampument, in the sense disconsiad in (hapter III.. depents upen a halance betwern what may be termed the favonrable and the unfa: ourable aidence : a new piece of evidence which lease this balance unchanged, also leaves the protability of the aronment unchanged. But it seems that there may be another respert in which somm lind of quantitation (omparison hetwem arguments is possible. This comparison turns upon a balance, not between the favourable and the unfarourable evidence, but between the absolute amounts of relevant knowledge and of relevant ignorance respectively.

As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according as the new knowledge strengthens the unfavourable or the favourable evidence; but something seems to have increased in cither case,-we have a more substantial basis upon which to rest our conclusion. I exprese this by saying that an accession of new evidence increases the weight of an argument. New evidence will sometimes decrease the probability of an argument, but it will always increase its 'weight.'
2. The measurement of evidential weight presents similar difficulties to those with which we met in the measurement of probability. Only in a restricted class of cases can we compare the weights of two arguments in respecet of more and less. But this must always be possible where the conclusion of the two arguments is the same, and the relevant evidence in the one includes and exceeds the evidence in the other. If the new evidence
is 'irrelevant,' in the inore precise of the two senses defined in § 14 of Chapter IV., the weight is left unchanged. If any part of the new evidence is relevant, then the value is increased.

The reason for our stricter definition of 'relevance' is now apparent. If we are to be able to treat 'weight ' and 'relevance' as correlative terms, we must reqard evidence as relevant, part of which is favourable and part unfavourable, eren if, taken as a whole, it leaves the probability unchanger. With this definition, to say that a new piece of evilence is 'relerant " is the same thing as to say that it increases the 'weight' of the argument.

A proposition cannot be the subject of an argument, unless we at least attach snme meaning to it, and this meaning, even if it only relates to the form of the proposition, may be relevant in some arguments relating to it. But there may be no other relevant eridence; and it is sometimes convenient to term the probability of such an argument an à priori probability. In this case the weight of the argument is at its lowest. Starting, therefore, with minimmun weight, corresponding to à prioni probability, the evidential weight of an argument rises, though its probability may either rise or fall, with every accession of relevant evidence.
3. Where the conclusions of two arguments are different, or where the evidence for the one does not overlap the evidence for the other, it will often be impossille to compare their weights, just as it may be impossible to compare their probabilities. Some rules of cumparison, however, exist, and there secms to be a close. though not a complete, correspondence between the conditions under which pairs of arguments are comparable in respect of probability and of weight respectively. We iound that there were three principal types in which comparison of probability was possible, other comparisons being based on a combination of these :-
(i.) Those based on the Principle of Indifference, subject to certain conditions, and of the form $\phi a / \psi a \cdot h_{1}=\phi b / \psi b . h_{2}$, where $h_{1}$ and $h_{2}$ are irrelevant to the arguments.
(ii.) $a / h h_{1} \fallingdotseq a / h$, where $h_{1}$ is a single unit of information, containing no independent parts which are relevant.

Let us represent the evidential weight of the argument, whose probability is a/h, by $V(a, k)$. Then, (onresponding to
the abowe, we find that the following comparisons of weight atw possible
(i.) $V\left(\phi^{\prime \prime} \psi^{\prime} h_{1}\right) \cdots\left(\phi^{\prime} / \psi^{\prime}, i\right)$, where $h_{1}$ and $h$ are irrelevant in the strict sense. Arguments, that is to say, to which the Principle of Indifference is applicable, have equal evidential weights.
(ii.) $\nabla\left(u / h h_{1}\right)>\mathrm{V}(u / h)$, unless $h_{1}$ is irrelevant, in which case $\boldsymbol{V}\left(a / h h_{1}\right)=\mathrm{V}(a / h)$. The restriction on the composition of $h_{1}$, which is necessary in the case of comparisons of magnitude, is not necessary in the case of weight.

There is, however, no rule for comparisons of weight correspontingto (iii.) abowe. It might be thought that $V\left(\ldots b_{i} / h\right)<V(u / h)$. on the ground that the more complicated an argument is, relatioe 10) given premisses, the less is its evidential weight. But this is invalid. The argunent $a b_{\text {, }} h$ is further off proof than was the argument $a / h$; but it is nearer disproof. For example, if $a b / h=0$ and $a / h>0$, then $V(a b / h)>V(a / h)$. In fact it would seem to be the case that the weight of the arcmume $a / h$ is always equal to that of $\bar{a} / h$, where $\bar{a}$ is the contradictory of $a$; i.e.. $V\left({ }^{\prime} h\right)=V(\bar{a} h)$. For an argument is always as near proving or disproving a jroposition, as it is to disproving or proving its contradictory.
4. It may be pointed out that if $a / h=b / h$, it does not necessarily follow that $\left.V^{\prime} n^{\prime} h\right)$ V $(b, h)$. It has herom asserted already that if the first "puality follows directly from a single application of the Princigle of Indifi-renee. the second equality also holds. But the first equality can exist in other cases also. If, for instance, $a$ and $b$ are members revpectively of diffechetsets of three equally. probable exclusive and exhastive alternatives thenu'h $b, h$; hut these argumonts may have very different weights. If, however. $a$ and $b$ coun eath, relatisely to $h$, be inferred from the other, ie. if $a / b h=1$ and $b / a h=1$, then $\mathrm{V}(a / h)=\mathrm{V}(b / h)$. For in proving or disproving one, we are necessarily proving or disproving the other

Further principles could, no doubt, be arrived at. The above can be combined to reach results in cases upon which unaided common-sense might feel itself unable to pronounce with confidence. Suppose, for instance, that we have three exclusive and exhaustive alternatives, $a, b$, and $c$, and that $a / h=b / h$ in virtue of the Principle of Indifference, then we have $\boldsymbol{V}(a, h) \quad V(b, h)$ and $V(n h) \quad V(\bar{a} h)$, a that $V(h, h) \quad V(\bar{a}, h)$. It is
also true. since $\bar{a} /(b)(c) / h \quad 1$ and $(b:(c) / \bar{a} h \quad$, that $\mathrm{V}(\bar{a} / h)$ $\mathrm{V}((b+c) / h)$. Hence $\mathrm{V}(b / h)=\mathrm{V}((b+c) / h)$.
5. The preceding paragraphs will have made it clear that the weighing of the amount of evidence is quite a separate process from the balancing of the evidence for and against. In so far, however, as the question of weight has been discussed at all, attempts have been made, as a rule, to explain the former in terms of the latter. If $x / h_{1} h_{2}=\frac{2}{3}$ and $x / h_{1}=\frac{3}{4}$, it has sometimes been supposed that it is more probable that $x / h_{1} h_{2}$ really is $\frac{2}{\overline{3}}$ than that $x / h_{1}$ really is $\frac{3}{4}$. According to this view, an increase in the amount of evidence strengthens the probability of the probability, or, as De Morgan would say, the presumption of the probability: A little reflection will show that such a theory is untenable. For the probability of $x$ on hypothesis $h_{1}$ is independent of whether as a matter of fact $x$ is or is not true, and if we find out subsequently that $x$ is true, this does not make it false to say that on hypothesis $h_{1}$ the probability of $x$ is $\frac{3}{4}$. Simi. larly the fact that $x / h_{1} h_{2}$ is $\frac{2}{3}$ does not impugn the conclusion that $x / h_{1}$ is ${ }_{4}^{3}$, and unless we have made a mistake in our judgment or our calculation on the evidence, the two probabilities are $\ddot{y}_{3}$ and $\frac{3}{4}$ respectively.
6. A second method, by which it might be thought, perhaps, that the question of weight has been treated, is the method of probable error. But while probable error is sometimes connected with weight, it is primarily concerned with quite a different question. 'Probable error,' it should be explained, is the name given, rather inconveniently perhaps, to an experssion which arises when we consider the probability that a given quantity is measured by one of a number of different macnitudes. Our datu may tell us that one of these magnitudes is the most probable measure of the quantity ; but in some cases it will also tell us how probable each of the other possible magnitudes of the quantity is. In such cases we can determine the probability that the quantity will have a maqnit de: which does not differ from the most probable by more than a specified amount. The amount, which the difference between the actual value of the quantity and its most probable value is as likely as not to exceed, is the 'probable error.' In many practical questions the existence of a small probable error is of the greatest importance, if our conclusions are to prove valuable. The probability that
the quantity has any pamienlar masmitude may be very small: hut this may matter sery litth, if there is a hiog protrability that it lies within a certain range.

Now it is obvious that the determination of probable error is intrinsically a different poohlem from the determination of weight. The methent of probable error is simply a summation of a number of alternatioe and exclusive probabilities. If we say that the most probable magnitude is $r$ and the probable wror $\%$. this is a way, convenient for many purposes, of smmmine up a number of probable conclusions rewarding a varioty of magnitudes other than $x$ which, on the evidence, the quantity may pesiss. The commetion between probable error and weight, such as it is, is due to the fact that in scientific prohboms a large probable error is not uncommomly due to a great lack of evidenee, and that as the avalable evidenere increases there is a temdency for the probable error to diminish. In these cases the probable error may concervably be a goon practical measure of the weight.

It is necessary, however, in a theoretical discussion, to point out that the comection is casual, and only exists in a limited class of cases. This is easily shown by an example. We may have data on which the probability of $x=5$ is $\frac{1}{3}$, of $x=6$ is $\frac{1}{4}$, of $x=7$ is $\frac{1}{5}$, of $x=8$ is $\frac{1}{6}$, and of $x=9$ is $\frac{1}{2}$. Additional evidence might show that of must either ho is or a or 9, the pormabilities of
 of the latter argument is areater than that of the former, but the probable error, so far from beine dimminhed, has hem inereased. There is, in fact, mo reasm whatever for supposing that the probable error must mecessarily dimminsh, as the weight of the argument is increased.

The typical case, in which there may he a practical connection between weight and probable error, may be illustrated by the two cases following of balls drawn from an urn. In each case we require the probability of drawing a white ball ; in the first case we know that the urn contains black and white in equal proportions; in the second case the proportion of each colour is unknown, and each ball is as likely to be black as white. It is evident that in either case the probability of drawing a white ball is $\frac{1}{2}$, but that the weight of the aremment in favour of this conelusion is greater in the first case. When we consider the most probable proportion in which balls will be drawn in the lome rum, if atter
each withdrawal they are replaced, the question of probable error enters in, and we find that the greater evidential weight of the argument on the first hypothesis is accompanied by the smaller probable error.

This conventionalised example is typical of many scientific problems. The more we know about any phenomenon, the less likely, as a rule, is our opinion to be modified by each additional item of experience. In such problems, therefore, an argument of high weight concerning some phenomenon is likely to be accompanied by a low probable error, when the character of a series of similar phenomena is under consideration.
7. Weight cannot, then. be explained in terms of probability. An argument of high weight is not 'more likely to be right ' than one of low weight; for the probabilities of these arguments only state relations between premiss and conclusion, and these relations are stated with equal accuracy in cither case. Nor is an argument of high weight one in which the probable error is small ; for a small probable error only means that magnitudes in the neighbourhood of the most probable magnitude have a relatively high probability, and an increase of evidence does not necessarily involve an increase in these probabilities.

The conclusion, that the 'weight' and the 'probability ' of an argument are independent properties, may possibly introduce a difficulty into the discussion of the application of probability to practice. ${ }^{1}$ For in deciding on a course of action, it seems plausible to suppose that we ought to take account of the weight as well as the probability of different expectations. But it is difficult to think of any clear example of this, and I do not feel sure that the theory of 'evidential weight' has much practical significance.

Bernoulli's second maxim, that we must take into account all the information we have, amounts to an injunction that we should be guided by the probability of that argument, amongst those of which we know the premisses, of which the evidential weight is the greatest. But should not this be re-enforced by a further maxim, that we ought to make the weight of our arguments as great as possible by getting all the information we can ? ${ }^{2}$ It is

[^26]difficult to see, however, to what point the strengthening of an argument's weight by increasing the evidence ought to bue pushed. We may argue that. when our knowledre is slight but capable of increase, the course of action, which will. relative to such knowledge, probably produce the aroatest amount of grod, will often consist in the acquisition of more knowledere. But there clearly romes a proint when it is no lonuer worth whil. to spend trouble. before artinge in the acquisition of further information, and there is no evident principle by which to determine how far we ought to carry our maxim of strengtheming the weight of our argument. A little reflection will probably convince the reader that this is a very confusing problem.
8. The fumdamental distinction of this chapter may be hriefly repeated. One argument has more weight than another if it is based upon a greater amount of relevant evidence; but it is not always, or even generally, possible to say of two sets of propositions that one set embodies more evidence than the other. It has a greater probelitily than another if the balanee in its favour, of what evidence there is, is greater than the balance in favour of the argument with which we compare it : hut it is mot always or even generally, possible to say that the balance in the one cat... is greater than the balance in the ofthe. The weight, to speak metaphorically: masures the sum of the fatourable and unfavourable evidence, the probability measures the difference.
9. The phenomenon of 'weight' can be described from the point of view of other theories of probability than that which is adopted here. If we follow certain (ierman logicians in regarding probability as being based on the disjunctive judgment, we may say that the weight is increased when the number of alternatives is reduced, although the ratio of the number of favourable to the number of unfavourable alternatives may not have been disturbed; or, to adopt the phraseology of another (ierman school, we may say that the weight of the probability is increased, as the field of possibility is contracted.

The same distinction may be explained in the language of the frequency theory. ${ }^{1}$ We should then say that the weight is increased if we are able to employ as the class of reforence a class which is contained in the original class of reference.
10. The subjeent of this chaptor has men usualle bemp diserussed
by writers on probability, and I know of only two by whom the question has been explicitly raised : ${ }^{1}$ Meinong, who threw out a suggestion at the conclusion of his review of Von Kries' "Principien," published in the Göttingische gelehrte Anzeigen for 18:00 (see especially pp. 70-74), and A. Nitsche, who took up Meinong's suggestion in an article in the Vierteljahrsschrift für wissenschaftliche Philosophie, 1892, vol. xvi. pp. 20-35, entitled "Die Dimensionen der Wahrscheinlichkeit und die Evidenz der Ungewissheit."

Meinong, who does not develop the point in any detail, distinguishes probability and weight as 'Intensität ' and 'Qualität,' and is inclined to regard them as two independent dimensions in which the judgment is free to move - they are the two dimensions of the 'Urteils-Continuum.' Nitsche regards the weight as being the measure of the reliability (Sicherheit) of the probability, and holds that the probability continually approximates to its true magnitude (reale Geltung) as the weight increases. His treatment is too brief for it to be possible to understand very clearly what he ineans, but his view seems to resemble the theory already discussed that an argument of high weight is ' more likely to be right ' than one of low weight.

[^27]
## ('IIAPTER VII

## HISTORICAL RETROSPECT

1. The charactoristic features of our Philosophy of Probability must be determined by the solutions which we offer to the problons attacked in (hapters III. and IV. Whilst a great part of the legrical calculus. which will be developed in Part II., would be applicable with slight modification to several distinet theories of the onhjeet, the ultimate prohlems of establishing the premisisces of the calculus hrine into the light every fundamental difference of opinion.

These problems are oiten, for this reason perhaps, luft on one side he writers whene inturest chiofly lies in the more formal parts of the subject. But Probability is not yet on so sound a basis that the formal or mathematical side of it ran be safoly developerd in isolation, and some attempts have naturally been made to solve the prohben which Bishop Buther sets to the logician in the concluding words of the brief discussion on probability with which he prefaces the Analogy. ${ }^{1}$

In this chapter, therefore, we will review in their historical order the answers of Philosophy to the questions, how we know relations of pmobaility. what eround we have for our juldenent. and by what method we can advance our knowledge.
2. The natural man is disposied to the opinion that probability is essentially connected with the inductions of experience and. if he is a little. mere sophisticated, with the Latws of C'ansation

[^28]and of the Uniformity of Nature. As Aristotle says, "the probable is that which usually happens." Events do not always occur in accordance with the expectations of experience; but the laws of experience afford us a good ground for supposing that they usually will. The occasional disappointment of these expectations prevents our predictions from being more than probable ; but the ground of their probability must be sought in this experience, and in this experience only.

This is, in substance, the argument of the authors of the Port Royal Logic (1662), who were the first to deal with the logic of probability in the modern manner: "In order for me to judge of the truth of an event, and to be determined to believe it or not believe it, it is not necessary to consider it abstractly, and in itself, as we should consider a proposition in geometry ; but it is necessary to pay attention to all the circumstances which accompany it, internal as well as external. I call internal circumstances those which belong to the fact itself, and external those which belong to the persons by whose testimony we are led to believe it. This being done, if all the circumstances are such that it never or rarely happens that the like circumstances are the concomitants of falsehood, our mind is led, naturally, to believe that it is true." ${ }^{1}$ Locke follows the Port Royal Logicians very closely : "Probability is likeliness to be true. . . . The grounds of it are, in short, these two following. First, the conformity of anything with our own knowledge, observation, and experience. Secondly, the testimony of others, vouching their observation and experience" $;^{2}$ and essentially the same opinion is maintained by Bishop Butler: "When we determine a thing to be probably true, suppose that an event has or will come to pass, it is from the mind's remarking in it a likeness to some other event, which we have observed has come to pass. And this olservation forms, in numberless instances, a presumption, opinion, or full conviction that such event has or will come to pass." ${ }^{3}$

Against this view of the subject the criticisms of Hume were directed: "The idea of cause and effect is derived from experience, which informs us, that such prarticular objects, in all past

[^29]instances, have been constantly conjoined with each other. . . . According to this account of things . . . probability is founded on the presumption of a resomblance betwist these objecte, of which we have hat experienee, and these, of which we have had nome ; and therefore tis imposible this presumption can arise from probability." 1 "When we are accustomed to see two impressions conjoined together, the appearance or idea of the one immediately carries us to the idea of the other. . . . Thus all probable reasoning is nothing but a species of sensation. 'Tis not solely in poetry and music, we must follow our taste and sentiment, but likewise in philosophy. When I am convinced of any principle, 'tis only an idea, which strikes more strongly upon me. When 1 wise the preference to one set of aremments above anot her, I do nothing but decide from my feeling concerning the superiority of their influence." ${ }^{2}$ Hume, in fact, points out that, while it is trm that past experimen gives rise to a peycholocical anticipation of some events rather than of others, no ground has been given for the validity of this superior anticipation.
3. But in the meantime the subject had fathon imo the hands of the mathematicians, and an entirely new method of approach was in course of development. It had become obvious that many of the judgments of probability which we in fact make do not depend upon past experience in a way which satisfies the canons laid down by the Port Royal Logicians or by Locke. In particular, alternatives are judged equally probable, without there being necessarily any actual experience of their approximately equal frequency of occurrence in the past. And, apart from this, it is evident that judgments based on a somewhat indefinite experience of the past do not easily lend themselves to precise numerical appraisement. Accordingly James
 probability, while not repudiating the old test of experience, had based many of his conclusions on a quite different criterion- the rule which I have named the Principle of Indifference. The traditional method of the mathematical school essentially depemds upon reatucing all the possible conclusions to a mumber of 'equi-probable cases.' And, according to the Principle of

[^30]${ }^{2}$ Op. cit. p. 403.
 p. 178.

Indifference, 'cases' are held to be equi-probable when there is no reason for preferring any one to any other, when there is nothing, as with Buridan's ass, to determine the mind in any one of the several possible directions. To take Czuber's example of dice, ${ }^{1}$ this principle permits us to assume that each face is equally likely to fall, if there is no reason to suppose any particular irregularity, and it does not require that we should know that the construction is regular, or that each face has, as a matter of fact, fallen equally often in the past.

On this Principle, extended by Bernoulli beyond those problems of gaming in which by its tacit assumption Pascal and Huyghens had worked out a few simple exercises, the whole fabric of mathematical probability was soon allowed to rest. The older criterion of experience, never repudiated, was soon subsumed under the new doctrine. First, in virtue of Bernoulli's famous Law of Great Numbers, the fractions representing the probabilities of events were thought to represent also the actual proportion of their occurrences, so that experience, if it were considerable, could be translated into the cyphers of arithmetic. And next, by the aid of the Principle of Indifference, Laplace established his Law of Succession by which the influence of any experience, however limited, could be numerically measured, and which purported to prove that, if $B$ has been seen to accompany A twice, it is two to one that B will again accompany A on A 's next appearance. No other formula in the alchemy of logic has exerted more astonishing powers. For it has established the existence of God from the premiss of total ignorance ; and it has measured with numerical precision the probability that the sun will rise to-morrow.

Yet the new principles did not win acceptance without opposition. D'Alembert, ${ }^{2}$ Hume, and Ancillon ${ }^{3}$ stand out as the sceptical critics of probability, against the credulity of
${ }^{1}$ Whahrscheinlichkeitsrechnung, p. 9.
2 D'Alembert's scepticism was directed towards the current mathematical theory only, and was not, like Hume's, fundamental and far-reaching. His opposition to the received opinions was, perhaps, more splendid than discriminating.
${ }^{3}$ Ancillon's communication to the Berlin Academy in 1794, entitled Doutes sur les buses du calcul des probabilités, is not as well known as it deserves to be. He writes as a follower of Hume, but adds much that is original and interesting. An historian, who also wrote on a variety of philosophical subjects, Ancillon was, at one time, the Prussian Minister of Foreign Affairs.
mighteenth-century philosophers who were ready to swallow without too many questions the conclusions of a science which clamed and seemed to bring an entire now field within the dominion of Reason. ${ }^{1}$

The first effective eriticism came from Hume, who was also the first to dist inguish the method of Locke and the philosophers from the methed of Bermoulli and the mathematicians. " P'roh, ability," he says, "or teasoning from conjecture, may be divided into two ki:.nls, viz. that which is foumded on chunce and that which arises from causes." ${ }^{2}$ By these two kinds he evidently means the mathematical method of counting the eqnal chances based on Indifference, and the inductive method based on the experience of uniformity. He argues that 'chance' alone can be the fourdation of nothing, and " that there must always be a mixture of canses among the chaneses, in order to be the foundation of any reasoning." ${ }^{3}$ His previous argument agranst probabilities, which were based on an assumption of cause, is thus extended to the mathematical method also.

But the great prestige of Laplace and the 'verifications' of his principhes which his more famons mesults wope supposed to supply had, by the berimmine of the minetemth centurs. established the science on the Principle of Indifterence in an almost unquestioned position. It may be moted, howerer, that De Morwan, the primeipal student of the subjeet in Emelamt seems to have regarded the method of actual experiment and the methond of counting cases, which were "rqually probable on grounds of Indifference, as alternative methods of equal validity.
4. The reaction aqainst the traditional teaching during the past humbed yoars has mot phssessend sufficient force 10 displame

[^31]the established doctrine, and the Principle of Indifference is still very widely accepted in an unqualified form. (riticism has proceeded along two distinct lines; the one, originated by Leslie Ellis, and developed by Dr. Venn, Professor Edgeworth, and Professor Karl Pearson, has been almost entirely confined in its influence to England ; the other, of which the beginnings are to be seen in Boole's Laws of Thought, has been developed in Germany, where its ablest exponent has been Von Kries. France has remained uninfluenced by either, and faithful, on the whole, to the tradition of Laplace. Even Henri Poincaré, who had his doubts, and described the Principle of Indifference as "very vague and very elastic," regarded it as our only guide in the choice of that convention, "which has always something arbitrary about it," but upon which calculation in probability invariably rests. ${ }^{1}$
5. Before following up in detail these two lines of development, I will summarise again the earlier doctrine with which the leaders of the new schools found themselves confronted.

The earlier philosophers had in mind in dealing with probability the application to the future of the inductions of experience, to the almost complete exclusion of other problems. For the data of probability, therefore, they looked only to their own experience and to the recorded experiences of others; their principal refinement was to distinguish these two grounds, and they did not attempt to make a numerical estimate of the chances. The mathematicians, on the other hand, setting out from the simple problems presented by dice and playing cards, and

[^32]requiring for the application of their methods a basis of numerical measurement, wwilt on the mesative rather than the positive side of their evidence, and found it easier to measure equal denerees of ignorance than emputalent yuantities of experience.
 or, as it was then termed, the Principle of Non-Sufficient Reason. The great adhemment of the eightenth century was. in the eyes of the catly nineterenth, the reconcilation of the $t$ wo points of view and th" momarement of probabilities, which were grounded on experience. Sy a method whone logical basis was the Principle of Xon-Sutiocint liasom. This would indeed have been a very astonishing discovery, and would, as its authors declared, have gradually hrousht ahmot "wory phase of human activity within the power of the most refined mathematical analysis.

But it was not long before more sceptical persons began to suspect that this theory proved too much. Its calculations, it is true, were constructed from the duta of experience, but the more simple athd the liss comples. the experience the better satisried was the theory. What was required was mot a wide experience or detailed information, but a completeness of symmeter in the little information there might be. It seemed to follow from the Laphatian doctrine that the primary qualification for ome who wouk bee woll informed was an equally balaneed igmorance.
6. The obvious reaction from a teaching, which seemed to derise from abstractions results relevant to experience, was into the arms of mupiricism ; and in the state of philesephy at that time: Emoland was the natural home of this reaction. The first protest, of which I am aware, came from Leslie Ellis in 1842. ${ }^{1}$ At the conclusion of his liomuriss on an ulleged promf of the . Methond of least squares, " Mere ignorance," he says, " is no ground for any inference whatever. Ex nitilo mhil." In Venn's Lengie of Chume lillis's sumerestions are developed into a complew. theory: ${ }^{3}$ "Experience is our sole quide. If we want to discover what is in reality a series of things, not a series of our own conceptions, we must appeal to the things themselves to obtain it, for
 was an early disciple of the same school: "The probability," he

[^33]says, " of head occurring $n$ times if the coin is of the ordinary make is approximately at least $\left(\begin{array}{l}\frac{1}{2}\end{array}\right)^{n}$. This value is rigidly deducible from positive experience, the observations made by gamesters, the experiments recorded by Jevons and De Morgan."

The doctrines of the empirical school will be examined in Chapter VIII., and I postpone my detailed criticism to that chapter. Venn rejects the applications of Bernoulli's theorem, which he describes as " one of the last remaining relics of Realism," as well as the later Laplacian Law of Succession, thus destroying the link between the empirical and the à priori methods. But, apart from this, his view that statements of probability are simply a particular class of statements about the actual world of phenomena, would have led him to a closer dependence on actual experience. He holds that the probability of an event's having a certain attribute is simply the fraction expressing the proportion of cases in which, as a matter of actual fact, this attribute is present. Our knowledge, however, of this proportion is often reached inductively, and shares the uncertainty to which all inductions are liable. And, besides, in referring an event to a series we do not postulate that all the members of the series should be identical, but only that they should not be known to differ in a relevant manner. Even on this theory, therefore, we are not solely deternined by positive knowledge and the direct data of experience.
7. The Empirical School in their reaction against the pretentious results, which the Laplacian theory affected to develop out of nothing, have gone too far in the opposite direction. If our experience and our knowledge were complete, we should be beyond the need of the Calculus of Probability. And where our experience is incomplete, we cannot hope to derive from it judgments of probability without the aid either of intuition or of some further à priori principle. Experience, as opposed to intuition, camnot possibly afford us a criterion by which to judge whether on given evidence the probabilities of two propositions are or are not equal.

However essential the data of experience may be, they cannot by themselves, it seems, supply us with what we want. Czuber, ${ }^{1}$ who prefers what he calls the Principle of Compelling Reason (das P'rinzip des zwingemden (irundes), and holds that Probability
has an objective and not merely formal interpretation only when it is grounded on: definite knowledge, is rightly comperled to admit that we canmet on oltowether without the Principhe of Non-Sufficient leasen. On the grounds both of its own intuitive plausibility and of that of some of the conclusions for which it is necessary, we are inevitably led towards this principle as a necessary basis for judgments of probability. In some sense, judgments of prohability dh s.em to low based on equally batanced degrees of ignorance.
8. It is from this starting-point that the German logicians have set nut. They have pereecived that there are few judements of peobability which are alowe ther independent of some principle resembline that of Non-Sufficient Reasom. But they also apprehend, with Boole, that this may be a very arbitrary method of procedure.

It was pointed out in § 18 of Chapter IV. that the cases, in which the Primeiple of Indifierence (or Non-Sufticient Reason) breaks down, have a great deal in common, and that we break up the fird of prositsility into a number of areas, act ually unequal, but indistinguishable on the evidence. Several German logicians, therefore have modeavoured to determine some rule by which it minht bee presibie to perstulate actual equality of atea for the fields of the various possibilities.

By far the most complete and closely reasoned solution on these lines is that of Von Kries. ${ }^{1}$ He is primarily anxious to discover a proper basis for the numerical measurement of probabilities, and he is thus led to examine with care the grounds of valid judgments of equiprobability. His criticisms of the Principle of Non-Sufficient Reason are searching, and, to meet them, he elaborates a mumber of qualifying conditions which are, he argues, meerssaty amb sufficient. The value of his book, hemever. lies, in the opinion of the present writer, in the critical rather than in the const ructive parts. The mamer in which his qualifying conditions are expressed is offon, to an Fimglish reater at any rate, somewhat obscure, and he seems sometimes to cover up difficulties, rather than solve them, by the invention of new technical terms. These characteristics render it difficult to expomul him :ndequat ly in a smmary, and the reander must he

[^34]referred to the original for a proper exposition of the Doctrine of Spielräume. Briefly, but not very intelligibly perhaps, he may be said to hold that the hypotheses for the probabilities of which we wish to obtain a numerical comparison, must refier to 'fields' (Spielräume) which are 'indifferent,' 'comparable ' in magnitude, and 'original' (ursprünglich). Two fields are 'indifferent' if they are equal before the Principle of Non-Sufficient Reason ; they are 'comparable' if it is true that the fields are actually of equal extent ; and they are 'original' or ultimate if they are not derived from some other field. The last condition is exceedingly obscure, but it seems to mean that the objects with which we are ultimately dealing must be directly represented by the 'fields ' of our hypotheses, and there must not be merely correlation between these objects and the objects of the fields. The qualification of comparability is intended to deal with difficulties such as that connected with the population of different areas of unknown extent ; and the qualification of originality with those arising from indirect measurement, as in the case of specific density.

Von Kries's solution is highly sugrestive, but it does not seem, so far as I understand it, to supply an unambiguous criterion for all cases. His discussion of the philnsophical character of probability is brief and inadequate, and the fundamental error in his treatment of the subject is the physical, rather than logical, bias which seems to direct the formulation of his conditions. The condition of Ursprünglichleit, for instance, seems to depend upon physical rather than logical criteria, and is, as a result, much more restricted in its applicability than a condition, which will really meet the difficulties of the case, ought to be. But, although I differ from him in his philosophical conception of probability, the treatment of the Principle of Indifference, which fills the greater part of his book, is, I think, along fruitful lines, and I have been deeply indebted to it in formulating my own conditions in Chapter IV.

Of less closely reasoned and less detailed treatments, which aim at the same kind of result, those of Sigwart and Lotze are worth noticing. Sigwart's ${ }^{1}$ position is sufficiently explained by the following extract: "The possibility of a mathematical treatment lies primarily in the fact that in the disjunctive judgment

[^35]the number of terms in the disjunction phays a decisive part. Inasmuch as a limited momber of mutaally exclusive possibilities is presented, of which one alone is actual, the element of number forms an essential part of our knowledge. . . . Our kmondenter mas mathe ne to assmme that the particular terms of the disjunction are so far equivalent that they express an equal degree of specialisation of a general concept, or that they cover equal parts of the whole extension of the concept. . . . This equivalence is most intuitable where we are dealing with equal parts of a spatial area, or equal parts of a period of time.
But even where this obvious quality is not fortheoming, we may ground our expectations upon a hypothetical equivalence, where we see no reason for considering the extent of one possibility to be greater than that of the others. . . ."

In the begiming of this passage Sigwart seems to be aware of the fundamental difficulty, although exception may be taken to the vagueness of the phrase " equal degree of specialisation of a general concept." But in the last sentence quoted he surrenders the advantages he has gained in the earlier part of his explanation, and, instead of insisting on a knowledge of an equal degree of specialisation, he is satisfied with an absence of any knowledge to the contrary. Hence, in spite of his initial qualifications, he ends unrestrainedly in the arms of Non-Sufficient Reason. ${ }^{1}$

Lotze, ${ }^{2}$ in a brief discussion of the subject, throws out some remarks well worth quoting: "We disclaim all knowledge of the circumstances which condition the real issue, so that when we talk of equally possible cases we can only mean coördinated as equiralent speries in the compass of an universal case; that is to say, if we conmerate the special forms, which the senus can assume, we whe a disjunctive judement of the form: if the condition $B$ is fulfilled, one of the kinds $f_{1} f_{2} f_{\ldots}$. . of the universal consequent F will occur to the exclusion of the rest. Which of all those different consequents will, in fact, occur, depends in all cases on the special form $b_{1} b_{2} b_{3} \ldots$ in which that universal condition is fulfillerl. . . A coördinuted case is a case which answers to one and only one of the mutually exclusive values $b_{1} b_{2} \ldots$ of the condition $B$, and these rival values may occur in

[^36]
reality; it does not answer to a more general form B , of this condition, which can never exist in reality, because it embraces several of the particular values $b_{1} b_{2}$. . ."

This certainly meets some of the difficulties, and its resemblance to the conditions formulated in Chapter IV. will be evident to the careful reader. But it is not very precise, and not easily applicable to all cases, to those, for instance, of the measurement of continuous quantity. By combining the suggestions of Von Kries, Sigwart, and Lotze, we might, perhaps, patch up a fairly comprehensive rule. We might say, for instance, that if $b_{1}$ and $b_{\text {, are }}$ are classes, their members must be finite in number and enumerable or they must compose stretches ; that, if they are finite in number, they must be equal in number ; and that, if their members compose stretches, the stretches must be equal stretches ; and that if $b_{1}$ and $b_{2}$ are concepts, they must represent concepts of an equal degree of specialisation. But qualifications so worded would raise almost as many difficulties as they solved. How, for instance, are we to know when concepts are of an equal degree of specialisation?
9. That probability is a relation has often received incidental recognition from logicians, in spite of the general failure to place proper emphasis on it. The earliest writer, with whom I am acquainted, explicitly to notice this, is Kahle in his Elementa logicae Probabilium methodo mathematica in usum. Scientiarum et Vitue adornata published at Halle in 1735. ${ }^{1}$ Amongst minre recent writers casual statements are common to the effect that the probability of a conclusion is relative to the grounds upon which it is based. Take Boole ${ }^{2}$ for instance: " It is implied in the definition that probability is always relative to our actual
${ }^{1}$ This work, which seems to have soon fallen into complete neglect and is now extremely rare, is full of interest and original thought. The following quotations will show the fundamental position taken up: "Est cognitio probabilis, si desunt quaedam requisita ad veritatem demonstrativam (p. 15). Propositio probabilis esse potest falsa, et improbabilis esse potest vera; ergo cognitio hodie possibilis, crastina luce mutari potest improbabilem, si accedunt reliqua requisita omnia, in certitudinem (p. 26). . . . Certitudo est terminus relativus: considerare potest ratione representationum in intellectu nostro. . . . Incerta nobis dependent a defectu cognitionis (p.35). . . . Actionem imprudenter et contra regulas probabilitatis susceptam eventus felix sequi potest. Ergo prudentia actionum ex successu solo non est aestimanda (p. 62). . . . Logica probabilium est scientia dijudicandi gradum certitudinis corum, quibus desunt requisita ad veritatem demonstrativam (p. 94)."

2 "On a General Method in the Theory of Probabilities," Phil. Mray., 4th Series, viii., 1854. See also, "On the Appliration of the Theory of Prohabilities
state of information and varies with that state of information." Or Bradle: : ${ }^{1}$ ". Probmbility tells us what we ought to believe. what we ought to believe on certain data . . . Probability is no more 'relative' and 'subjective' than is any other act of logical inference from hypothetical premises. It is relative to the data with which it has to deal, and is not relative in any other sense." Or even Laplace, when he is explaining the diversity of human opinions: "Dans les choses qui ne sont que vraisemhables, la diftírome dus dommes que chaque homme a sur clles. est une des causes principales de la diversité des opinions que l'on voit régner sur les mêmes objets . . . c'est ainsi que le même fait. récité devant une nombreuse assmblée, ohtient divers degrés de croyance. suivant lótemdue des commaissances des auditmurs." ${ }^{2}$
10. Here we may leave this account of the various directions in which prowris has semed posisible, with the hope that it may assist the reader, who is dissatisfied with the solution proposed in Chapter IV., to determine the line of argument alone which he is likeliest to discover the solution of a difficult problem.
to the Question of the Combination of Testimonies or Judgments" (Edin. Phil. Trans. xxi. p, 600): "Our estimate of the probability of an event varies not absolutely with the circumstances which actually affect its oceurence, but with our knowledge of those ciroumstances."

${ }^{2}$ Essai philosophiqu!, p. 7.

## CHAPTER VIII

## THE FREQUENCY THEORY OF PROBABILITY

1. The theory of probability, outlined in the preceding chapters, has serious difficulties to overcome. There is a theoretical, as well as a practical, difficulty in measuring or comparing degrees of probability, and a further difficulty in determining them à priori. We must now examine an alternative theory which is much freer from these troubles, and is widely held at the present time.
2. The theory is in its essence a very old one. Aristotle foreshadowed it when he held that "the probable is that which for the most part happens " ; ${ }^{1}$ and, as we have seen in Chapter VII., an opinion not unlike this was entertained by those philosophers of the seventeenth and eighteenth centuries who approached the problems of probability uninfluenced by the work of mathematicians. But the underlying conception of earlier writers received at the hands of some English logicians during the latter half of the nineteenth century a new and much more complicated form.

The theory in question, which I shall call the Frequency Theory of Probability, first appears ${ }^{2}$ as the basis of a proposed logical scheme in a brief essay by Leslie Ellis On the Forndutions of the Theory of Probubilities, and is somewhat further developed in his Remalis on the Fundumental Principles of the Theory of

## 1 Rhet. i. 2, 1357 a 34.

${ }^{2}$ I give killis the priority because his paper, published in 1843, was read on Feb. 14, 1842. The same conception, however, is to be found in Cournot's Exposition, also published in 1843: "La théorie des probabilités a pour objet certains rapports numériques qui prendraient des valeurs fixes et complétement déterminées, si Jon pouvait répéter a lomfini les épreuves des mêmes hasards. et qui, pour un nombre fini d'épreuves, oscillent entre des limites d'autant plus resserrées, d'autant plus voisines des valeurs finalcs, que le nombre des épreuves est plus grand."

Probubilities." "If the probability of a given event bew correctly detemmed," he sals, "the went will on a komer rum of trials temed to recar with frepuence proportional to their prohathity: This is generally proved mathematically. It seems to me to be true à priori. . . . I have been unable to sever the judgment that one went is more likely to hafnen than another from the belint that in the long run it will occur more frequently." Ellis explicitly introduces the conception that probability is essentially. concerned with a group or series.

Although the priority of incention must be allowed to Leslie Ellis, the theory is commomly assoneiated with the name of Vemu. In his Logic of Chance ${ }^{2}$ it first received elaborate and systematic treatment, and, in spite of his having attracted a number of followers, there has been no other comprehemsive attempt to meet the theory's special difficulties or the criticisms directed against it. I shall begin, therefore, be examining it in the form in which Venn has expounded it. Vemn's exposition is much coloured by an empirical view of logic, which is not perhaps as necessary to the essential part of his doctrine as he himself implies, and is not shared by all of those who must be classed as in general agreement with him about probability. It will be necessary, therefore, to supplement a criticism of Venn by an account of a more general frequency theory of probability, divested of the empiricism with which he has clothed it.
3. The following quotations from Venn's Logic of Chance will show the general drift of his argument: The fumdamental comception is that of a series (p. 4). The series is of events which have a certain number of features of attributes in common ( $p$. 10 ). The characteristic distinctive of probability is this,-- the oceasional attributes, asdistimenishod from the permanment, are foumb on an examination to tend to exist in a certain definite proportion of the whole number of conses (p). 11). We meguire that there should be in mature larew classes of objects, throughout all the individual members of which a general resemblance extends. For this

[^37]purpose the existence of natural kinds or groups is necessary (p. 55). The distinctive characteristics of probability prevail principally in the properties of natural kinds, both in the ultimate and in the derivative or accidental properties (p.63). The same peculiarity prevails again in the force and frequency of most natural agencies (p. 64). There seems reason to believe that it is in such things only, as distinguished from things artificial, that the property in question is to be found (p. (65). How, in any particular case, are we to establish the existence of a probability series? Experience is our sole guide. If we want to discover what is in reality a series of things, not a series of our own conceptions, we must appeal to the things themselves to obtain it, for we cannot find much help elsewhere (p). 174). When probability is divorced from direct reference to objects, as it substantially is by not being founded upon experience, it simply resolves itself into the common algebraical doctrine of Permutations and Combinations (p. 87). By assigning an expectation in reference to the individual, we mean nothing more than to make a statement about the average of his class (p. 151). When we say of a conclusion within the strict province of probability, that it is not certain, all that we mean is that in some proportion of cases only will such conclusion be right, in the other cases it will be wrong (p. 210).

The essence of this theory can be expressed in a few words. To say, that the probability of an event's having a certain characteristic is ${ }_{y}^{x}$, is to mean that the event is one of a number of events, a proportion ${ }_{y}^{x}$ of which have the characteristic in question; and the fact, that there is such a series of events possessing this frequency in respect of the characteristic, is purely a matter of experience to be determined in the same manner as any other question of fact. That such series do exist happens to be a characteristic of the real world as we know it, and from this the practical importance of the calculation of probabilities is derived.

Such a theory possesses manifest advantages. There is no mystery about it-no new indefinables, no appeals to intuition. Measurement leads to no difficulties ; our probabilities or frequencies are ordinary numbers, upon which the arithmetical apparatus can be safely brought to bear. And at the same time it
sums to crystallise in a clear, "xplicit shape the flomine opinion of common sense that an event is or is not probable in certain supposed circumstances according as it is or is not usual as a matter of fact and experience.

The two principal tenets, then, of Venn's system are these,--
 and that all the requisite fants must be deemmined ampirically, a statement in probability morely sammine up, in a conveniont way a group of experiences. Aggregate regularity combined with individual differenme happens, the says, th the charact ristib of many events in the real world. It will often be the case, therefore, that we can make statements regarding the average of a certain class, or regarding its characteristics in the long run, which we cannot make about any of its individual members without great risk of error. As our knowleden remarting the class as a whole may give us valuable gridance in dealing with an individual instance, we require a convenient way of saying that an individual belomes to a class in which certain characteristics appear on the average with a known frequency ; and this the conventional language of probability gives us. 'The importance of probability depemts solely upon the actual existome of such groups or real kinds in the world of experience, and a judement of probability must meressarily depend for its valudity upon our empirical knowledge of them.
4. It is the obvious, as well as the correct, criticism of such a theory, that the identification of prohability with statistical frequency is a very grave departure from the established use of words; for it clearly excludes a ereat number of judgment: which are generally believed to deal with probatility. Vemn himself was well aware of this, and cannot be accused of supposing that all b, miefs, which are commonly called probahle, are really concerned with statistimal frequency. But sume of his followers. to judge from their published work, have mot always seen, so clearly as he did, that his theory is not conermed with the satme subject as that with which other writers have dealt umder the same title. Vimon justifics his prowedure by arguim that mother meaning, of which it is pmosithe to take strict lomic... congisance, can reasomably be given to the term, and that the other meaninge, with which it has been used, hase mot emough in common to permit their reduction to a singhe legital scheme. It is useless,
therefore, for a critic of Venn to point, out that many supposed judgments of probability are not concerned with statistical frequency; for, as I understand the Logic of C'hunce, he admits it ; and the critic must show that the sense different from Venn's in which the term probability is often employed has an important logical interpretation about which we can generalise. This position I seek to establish. It is, in my opinion, this other sense alone which has importance ; Venn's theory by itself has few practical applications, and if we allow it to hold the field, we must admit that probability is not the guide of life, and that in following it we are not acting according to reason.
5. Part of the plausibility of Venn's theory is derived, I think, from a failure to recognise the narrow limits of its applicability, or to notice his own admissions regarding this. "In every case," he says (p. 124), " in which we extend our inferences by Induction or Analogy, or depend upon the witness of others, or trust to our own memory of the past, or come to a conclusion through conflicting arguments, or even make a long and complicated deduction by mathematics or logic, we have a result of which we can scarcely feel as certain as of the premisses from which it was obtained. In all these cases, then, we are conscions: of varying quantities of belief, but are the laws according to which the belicf is produced and varied the same? If they cannot be reduced to one harmonious scheme, if, in fact, they can at best bee brought to nothing but a number of different schemes, each with its own body of laws and rules, then it is vain to endeavour to force them into one science." All these cases, therefore, in which we are 'not certain,' Venn explicitly excludes from what he chooses to call the science of probability, and he pays no further attention to them. The science of probability is, according to him, no more than a method which enables us to experesis in a convenient form statistical statements of frequence. "ithe province of probability," he says again on paye 160, " is not so extensive as that over which variation of belief might be observed. Probability only considers the case in which this variation is brought about in a certain definite statistical way." ${ }^{1}$ He points

[^38]out on p. 194 that for the purposes of probability we must take the statistical frequency from which we start rendy mode and ask no questions about the process or completeness of its manufacture: "It may be obtained by any of the numerous rules furnished by Induction, or it may be inferred deductively, or given by our own observation; its value may be diminished by its depending upon the testimony of witnesses, or its being recalled by our own memory. Its real value may be influenced by these causes or any combinations of them ; but all these are preliminary questions with which we have nothing directly to do. We assumb our statistical proposition to be true, neglecting the diminution of its value by the processes of attainment."

It must be reeomenised, therefore, that Venn has deliberately excluded from his surver almost all the cases in which we regard our judements as "omly probable `; and, whatever the value or consistency of his own seheme, he has left untouched a wide field of study for others.
6. The main erounds, which have induced Vemn to resard judernents based on statistical frequency as the only cases of probability which possess logical importance, seem to be two: (i.) that wher cases are mainly subjective, and (iii) that they are incapable of accurate measurement.

With regard to the first it must be admitted that there are many instances in which variation of belief is oceasioned by purely psechological causes, and that his argument is valid against those who have defined probability as measuring the degree of sub)jective bolief. But this has mot been the usual way of looking at the subject. Probability is the study of the grounds which lead us to entertain a rational preference for one belief over another. That there are rational eromels other than statistical frequency, for such preferences, Vemn dows not deny: lue ahnits in the quotation given above that the 'real rolue' of our conclusion is influenced by many other con-
on a particular standard. That standard is the phenomenon of statistical
 surh that the mumber of indivimals in wew speries bears an apposimmely constant ratio to the number of individuals in the genus." This use of terms is legitimate, though it is not easy to follow it consistently. But, like Venn's, it leases a-ide the most impertant questions. Then Caleulus of Probabuli. ties, thas interpreted, is no puide hy itarlf as to which opinimu wo ought to follow, and is not a measure of the weight we should attach to conflicting arguments.
siderations than that of statistical frequency. Venn's theory, therefore, cannot be fairly propounded by his disciples as alternative to such a theory as is propounded here. For my Treatise is concerned with the general theory of arguments from premisses leading to conclusions which are reasonable but not certain ; and this is a subject which Venn has, deliberately, not treated in the Logic of Chance.
7. Apart from two circumstances, it would scarcely be necessary to say anything further ; but in the first place some writers have believed that Venn has propounded a complete theory of probability, failing to realise that he is not at all concerned with the sense in which we may say that one induction or analogy, or testimony, or memory, or train of argument is more probable than another ; and in the second place he himself has not always kept within the narrow limits, which he has himself laid down as proper to his theory.

For he has not remained content with defining a probability as identical with a statistical frequency, but has often spoken as if his theory told us which alternatives it is reasonable to prefer. When he states, for instance, that modality ought to be banished from Logic and relegated to Probability (p. 296), he forgets his own dictum that of premisses, the distinctive characteristic of which is their lack of certainty, Probability takes account of one class only, Induction concerning itself with another class, and so forth (p. 321). He forgets also that, when he comes to consider the practical use of statistical frequencies, he has to admit that an event may possess more than one frequency, and that we must decide which of these to prefer on extraneous grounds (p. 213). The device, he says, must be to a great extent arbitrary, and there are no logical grounds of decision ; but would he deny that it is often reasonable to found our probability on one statistical frequency rather than on another? And if our grounds are reasonable, are they not in an important sense logical?

Even in those cases, therefore, in which we derive our preference for one alternative over another from a knowledge of statistical frequencies, a statistical frequency by itself is insufficient to determine us. We may call a statistical frequency a probability, if we choose ; but the fundamental problem of determining which of several alternatives is logically preferable still awaits solution. We cannot be content with the only counsel Venn
can offer, that we should choose a frequency which is derived from a series neither too large nor too small.

The same difficulty, that a probability in limn's sense is insufficient to determine which alternative is lowically preferahle. arises in another comection. In most cases the statistical frequency is not given in experience for certain, but is arriwad at by a process of imburliom, and inductions, he admits, are not certain. If, in the past, three infants out of every ten have died in their first four years, induction may base on this the doubtful assertion, All infants die in that propertion. But we camnot assert on this rround, as Vemn wishes to do, that the prob)ability of the death of an infant in its first four years is in thes. We can say no mome than that it is probable (in my sense) that there is such a probability (in his semser). For the purposie of coming to a decision we cannot compare the value of this conclusion with that of others until we know the probatility (in my sense) that the statistical frequenes really is onths. The cases in which we can determine the logical value of a conclusion entiry on srounds of statistical frequency would seem to be extremely few in number.
8. The second main reason which led Venn to develop his thener is to ber foum in his thelief that probabilities which are based on statistical frequmetes are alom (aprable of accurate measurement. The term probabilities, her arsues, is properly confined to the case of whancos which can be calculated, and all calculable chances can be made tor depend upen statistical frequency. In attempting to estahlish this latter contention he is imworm in som" paradoxical opinims. "In many cases," heradmits, "it is undoubededy true that we do not resort to diesect experience at all. If I want to know what is my chance of holding ton trumps in a game of whist, I do not moquire how often such a thing hats ocecurred before. . . . In practice, à priori determination: is often easy, whilst is pesterimi appeal to experience would be not merely tedious but utterly impracticable." But these cases which are nemally based on the Principle of Indifference can, he maintains, he justified on statistical grounds. In the case of cein tossing there is a considerable experience of the equally frequent oceurrence of heads and tails : the experience gained in this simple case is to be extended to the complex cases by "Induction and Analogy." In one simple cate the
result to which the Principle of Indifference would lead is that which experience recommends. Therefore in complex cases, where there is no basis of experiment at all, we may assume that Experience, if experience there was, would speak with the same voice as Indifference. This is to assert that, because in one case, where there is no known reason to the contrary, there actually is none, therefore in other cases incapable of verification the absence of known reason to the contrary proves that actually there is none.

The attempt to justify the rules of inverse probability on statistical grounds I have failed to understand ; and after a careful reading, I am unable to produce an intelligible account of the argument involved in the latter part of chapter vii. of the Logic of Chance. ${ }^{1}$ I am doubtful whether Venn should not have excluded à posteriori arguments in probability from his scheme as well as inductive arguments. The attempt to include them may have been induced by a desire to deal with all cases in which numerical calculation has been commonly thought possible.
9. The argument so far has been solely concerned with the case for the frequency theory developed in the Logic of Chance. The criticisms which follow will be directed against a more general form of the same theory which may conceivably have recommended itself to some readers. It is unfortunate that no adherent of the doctrine, with the exception of Venn, has attempted to present the theory of it in detail. Professor Karl Pearson, for instance, probably agrees with Venn in a general way only, and it is very likely that many of the foregoing remarks do not apply to his view of probability ; but while I generally disagree with the fundamental premisses upon which his work in probability and statistics seems to rest, I am not clearly aware of the nature of the philosophical theory from which he thinks that he derives them and which makes them appear to him to be satisfactory. A careful exposition of his logical presuppositions would greatly add to the completeness of his work. In the meantime it is only possible to raise general objections to

[^39]any theory of probability which seeks to found itself upon the conception of statistical frequency.

The generalised frequency theory which I proposie to put forward. as perhaps representative of what adherents of this doctrine have in mind, differs from Vemn's in several important respects. ${ }^{1}$ In the first place, it does not regard probability as being identical with statistical frequency, although it holds that all probabilities must be based on statements of frerpuency, and can be defined in terms of them. It accepts the theory that propositions rather than events should the taken as the subjectmatter of probability ; and it adopts the comprehensive view of the subject according to which it includes induction and all other cases in which we beliwe that there are logical grounds for preferring one alternative out of a set none of which are certain. Nor does it follow Vemn in supposing any special connection to exist between a frequency theory of probability and logical empiricism.
10. A proposition can be a member of many distinct classes of propositions, the classes being merely constituted by the existence of particular resemblances between their members or in some such way. We may know of a given proposition that it is one of a particular class of propositions, and we masy also know, precisely or within defined limits, what proportion of this class are true, without our being aware whether or not the given proposition is true. Let us, therefore, call the actual proportion of true propositions in a class the truth-fremumey ${ }^{2}$ of the class, and define the measure of the probability of a proposition relative to a class, of which it is a member, as being equal to the truthfrequency of the class.

The fundamental tenet of a frequency theory of probability is, then that the probatility of a proposition always depends upon referring it to some class whose truth-frequeney is known within wide or narrow limits.

Such a theory possesses most of the adrantages of Vimn's, but escapes his narrowness. There is nothing in it so far which could not be casily expressed with complete precision in the terms of ordinary logic. Nor is it necessariiy confined to proh-

[^40]abilities which are numerical. In some cases we may know the exact number which expresses the truth-frequency of our class ; but a less precise knowledge is not without value, and we may say that one probability is greater than another, without knowing how much greater, and that it is large or small or negligible, if we have knowledge of corresponding accuracy about the truthfrequencies of the classes to which the probabilities refer. The magnitudes of some pairs of probabilities we shall be able to compare numerically, others in respect of more and less only, and others not at all. A great deal, therefore, of what has been said in Chapter III. would apply equally to the present theory, with this difference that the probabilities would, as a matter of fact, have numerical values in all cases, and the less complete comparisons would only hold the field in cases where the real probabilities were partially unknown. On the frequency theory, therefore, there is an important sense in which probabilities can be unknown, and the relative vagueness of the probabilities employed in ordinary reasoning is explained as belonging not to the probabilities themselves but only to our knowledge of them. For the probabilities are relative, not to our knowledge, but to some objective class, possessing a perfectly definite truthfrequency, to which we have chosen to refer them.

The frequency theory expounded in this manner cannot easily avoid mention of the relativity of probabilities which is implicit here, as it is in Venn's. Whether or not the probability of a proposition is relative to given data, it is clearly relative to the particular class or series to which we choose to refer it. A given proposition has a great variety of different probabilities corresponding to each of the various distinct classes of which it is a member ; and before an intelligible meaning can be given to a statement that the probability of a proposition is so-and-so, the class must be specified to which the proposition is being referred. Most adherents of the frequency theory would probably go further, and agree that the class of reference must be determined in any particular case by the duta at our disposal. Here, then, is another point on which it is not necessary for the frequency theory to diverge from the theory of this Treatise. It should, I think, be generally agreed by every school of thought that the probability of a conclusion is in an important sense relative to given premisses. On this issue and also on the point that our
knowledge of many prohabilities is not numerically definite, there might well be for the future an end of disagreement, and disputation might be reserved for the philosophical interpertation of these settled fact:, which it is unreasonable to deniy, however we may explain them.
11. I now proceed to thase contentions upon which my fundamental criticism of the frequency theory is founded. The first of these relates to the method by which the class of reference is to be determined. The magnitude of a probability is always to be measured by the truth-frequency of some class: and this class, it is allowed, must bee determined by reference to the premises, on which the probability of the conclusion is to be determined. Bat, an a given proposition belongs to innumerable different classes, how are we to know which clasis the premisses indicate as apmepriate? What substitute has the frequency theory to offer for judements of relevance and indifierence? And without somethinge of this kind, what principle is there for uniquely determining the class, the truth-frequency of which is to measure the probability of the argument? Indeed the difliculties of showing how riven premisses determine the class of reference, by means of rules expressed in terms of previous ideas, and without the introduction of any notion, which is new and peculiar to probability, appear to me insurmountable.

Whilst no two alternative classes meither includes the wther. it might be thought that where one does include the other, the obvious course would be to take the narrowest and most ipuccialised class. This procedure was examined and rejected by Vimm: thourh the ,hbection to it is due, not, as he supposed, to the lack of suflicient statistics in such cases upen which to foumd is cencralisation, but to the inclusion in the class-concept of marks characteristic of the proposition in question, but nevertheless not relevant to the matter in hand. If the process of narrowing the class were to be carricd to its furthest perint, we should remerally be left with a class whase only member is the propmation in question, for we generally know something about it which is true of no other propesition. Wie cammot, therefore, derine the class of reference as twing the class of propesitions of which everything is true which is known to be true of the proposition whose probability wer seth to determine. And, imdeed, in those examples
for which the frequency theory possesses the greatest prima facie plausibility, the class of reference is selected by taking account of some only of the known characteristics of the quaesitum, those characteristics, namely, which are relevant in the circumstances. In those cases in which one can admit that the probability can be measured by reference to a known truth-frequency, the class of reference is formed of propositions about which our relerant knowledge is the same as about the proposition under consideration. In these special cases we get the same result from the frequency theory as from the Principle of Indifference. But this does not serve to rehabilitate the frequency theory as a general explanation of probability, and goes rather to show that the theory of this Treatise is the gencralised theory, comprehending within it such applications of the idea of statistical truthfrequency as have validity.
'Relevance' is an important term in probability, of which the meaning is readily intelligible. I have given my own definition of it already. But I do not know how it is to be explained in terms of the frequency theory. Whether supporters of this theory have fully appreciated the difficulty I much doubt. It is a fundamental issue involving the essence of the peculiarity of probability, which prevents its being explained away in terms of statistical frequency or anything else.
12. Yet perhaps a modified view of the frequency theory could be evolved which would avoid this difficulty, and I proceed, therefore, to some further criticisms. It might be agreed that a novel element must be admitted at this point, and that relevancy must be determined in some such manner as has been explained in earlier chapters. With this admission, it might be argued, the theory would still stand, divested, it is true, of some of its original simplicity, but nevertheless a substantial theory differing in important respects, although not quite so fundamentally as before, from alternative schemes.

The next important objection, then, is concerned with the manner in which the principal theorems of probability are to be established on a theory of frequency. This will involve an anticipation in some part of later arguments ; and the reader may be well advised to return to the following paragraph after he has finished Part II.
13. Let us begin by a consideration of the 'Addition Theorem.'

If $a^{\prime} h$ denomes the probability of a on hypothesis $h$, this theorem may be written $(a+b) \cdot h=a, h-b / h-a b_{i}, h$, and may be read "On hepothesis $h$ the probability of " $a$ or $b$ " is equal to the probalifity of $a$ the probability of $b$ - the probability of " both $a$ and $b$." " This theorem, interpreted in some way or other, is universally assumed; and we must, therefore, inquirn What prow of it the frequency theory can afford. A little symbolism will assist the argument: Let a represent the truthfrequency of any class $a$, and let $a_{4} / h$ stand for 'the probability of on hopothesis $h$, a buing the class of reference determined by this hyunthosis.' We then have $a_{a j} / h=a_{j}$, and we require to prose a proposition, for values of $y$ and $\delta$ not yet determined, which will be of the form :

Now if , is the class of propositions ( $n: b$ ) such that $a$ is an a ambl a $\beta$, it is easily shown he the ordinary arithmetie of classes
 are members of both $a$ and $\beta$. In the case, therefore, where $\hat{\delta}=\hat{c}^{\prime}$ and $\gamma \quad$ " $\beta$, an addition theorem of the required kind has been established.

But it dons not follow by any reasonable rule that, if $h$ determines $a$ and 3 as the appropriate classes of reference for $a$ and $b$. $h$ must ne cessarily determine $e^{\circ}$ and $\left.a,\right\}^{3}$ as the appropriate classes of reffernee for $(n: h)$ and $1 b$ : it may, for instance, be the case that $h$, while it rendors a and $s$ determinate, vields no information whatever resarding a,r, and points to some quite different chass $\mu$ as the suitable class of reference for ab. On the frequency theory, therefore, we canont maintain that the addition theorem is true in inmeral, but only in these sperial cases where it happens that $\delta=\delta^{\prime}$ and $\gamma=\alpha \beta$.

The following is a good example: We are given that the proportion of black-haired men in the population is $H_{1}{ }_{1}$ and the proportion of colour-blind men ${ }_{q} I_{2}$, and there is no known connection between black - hair and colour - blindness: what is the probability that a man, about whom nothing sperial

[^41]is known, is ${ }^{1}$ either black-haired or colour-h, lind ? If we represent the hypotheses by $h$ and the alternatives by $a$ and $b$, it would usually be held that, colour-blindness and black hair being indepentent for linowledge ${ }^{2}$ relative to the given data, $a l_{1} / h_{1}=\begin{aligned} & P_{1} P_{2}, \\ & y_{2}^{2}\end{aligned}$, and that, therefore, by the addition theorem, $(n+b) / h=\frac{r_{1}}{q}+$ $\mu_{2}$
$q$$p_{1} p_{2}$. But, on the frequency theory, this result might be invalid; for $a \beta_{f}=\frac{p_{1} p_{2}}{q^{2}}$, only if this is the actual proportion in fact of persons who are both colour-blind and black-haired, and that this is the actual proportion camot possibly be inferred from the inclependence for knowledge of the characters in question. ${ }^{3}$

Precisely the same difficulty arises in connection with the multiplication theorem $a b / h=a / b h . b / h .{ }^{4} \quad$ In the frequency notation, which is proposed above, the corresponding theorem will be of the form $a b_{\delta} / h=a_{\gamma} / b h . b_{\beta} / h$. For this equation to be satisfied it is easily seen that $\delta$ must be the class of propositions $x y$ such that $x$ is a member of $a$ and $y$ of $\beta$, and $\gamma$ the class of propositions $x b$ such that $x$ is a member of $a$; and, as in the case of the addition theorem, we have no guarantee that these classes $\gamma$ and $\delta$ will be those which the hypotheses $b h$ and $h$ will respectively determine as the appropriate classes of reference for $a$ and $a b$.

In the case of the theorem of inverse probability ${ }^{5}$

$$
\begin{aligned}
& \text { li/ah alble li/h } \\
& \text { c/ath "ulch ejp }
\end{aligned}
$$

the same difficulcy again arises, with an additional one when practical applications are considered. For the relative probabilities of our $a ̀$ priori hypotheses, $b$ and $c$, will scarcely ever be capable of determination by means of known frequencies, and in the most legitimate instances of the inversi principle's operation

[^42]we depend either upon an inductive argument or upon the Principle of Indifference. It is hard to think of an example in which the frequency conditions are even approximately satisfied.

Thus an important class of case, in which arguments in probatbility, generally accepted as satisfactory, do not satisfly the frequency conditions given above, are those in which the notion is introduced of two propositions being, on certain duta, independent for knowledge. The meaning and definition of this expression is discussed more fully in Part II. ; but I do not see what interpretation the frequency theory can put upon it. Vet if the conception of 'independence for knowledge' is discarded, we shall be brought to a standstill in the vast majority of problems. which are ordinarily considered to be problems in probability, simply from the lack of sufficiently detailed data. Thus the frequency theory is not adequate to explain the processes of reasoning which it sets out to explain. If the theory restricts its operation, as would seem necessary, to those cases in which we linow precisely how far the true members of a and 3 overlap, the vast majority of arguments in which probability has been employed must be rejected.
14. An appeal to some further principle is, therefore, remuirel before the ordinary apparatus of probable inference can be catahlished on considerations of statistical frequency; and it may have occurred to some readers that assistance may be ohtainat from the principles of induction. Here also it will be necessary to anticipate a subsequent discussion. If the argument of l'art III. is correct, nothing is more fatal than Induction to the themry now under criticism. For, so far from Induction's lending support to the fundamental rules of prohability: it is itself dependent on them. In any case, it is gemerally agreed that an inductive conclusion is only probable, and that its probability increases with the number of instances upon whith it is fommbent. Accordine to the frequency theory, this belief is only justified if the majority of imductive conclusions actually are true, and it will be false, exen on our existing data, that any of them are exen probable, if the arknowledend prossibility that a majority are false is an actuality. Yet what possible reason can the frequency theory offer, which does not beg the question, for supposing that a majority are true? And failing this, what ground have we for believing the imductive process to be reasonahlu! Yet we.
invariably assume that with our existing knowledge it is logically reasonable to attach some weight to the inductive method, even if future experience shows that not one of its conclusions is verified in fact. The frequency theory, therefore, in its present form at any rate, entirely fails to explain or justify the most important source of the most usual arguments in the field of probable inference.
15. The failure of the frequency theory to explain or justify arguments from induction or analogy suggests some remarks of a more general kind. While it is undoubtedly the case that many valuable judgments in probability are partly based on a knowledge of statistical frequencies, and that many more can be held, with some plausibility, to be indirectly derived from them, there remains a great mass of probable argument which it would be paradoxical to justify in the same manner. It is not sufficient, therefore, even if it is possible, to show that the theory can be developed in a self-consistent manner ; it must also be shown how the body of probable argument, upon which the greater part of our generally accepted knowledge seems to rest, can be explained in terms of it; for it is certain that much of it does not appear to be derived from premisses of statistical frequency.

Take, for instance, the intricate network of arguments upon which the conclusions of The Origin of Species are founded: how impossible it would be to transform them into a shape in which they would be seen to rest upon statistical frequency! Many individual arguments, of course, are explicitly founded upon such considerations; but this only serves to differentiate them more clearly from those which are not. Darwin's own account of the nature of the argument may be quoted: "The belief in Natural Selection must at present be grounded entirely on general considerations: (1) on its being a vera causa, from the struggle for existence and the certain geological fact that species do somehow change; (2) from the analogy of change under domestication by man's selection ; (3) and chiefly from this view connecting under an intelligible point of view a host of facts. When we descend to details . . . we cannot prove that a single species has changed ; nor can we prove that the supposed changes are beneficial, which is the groundwork of the theory ; nor can we explain why some species have changed and others
have not." ${ }^{1}$ Not only in the main argument, but in many of the subsidiary discussions, ${ }^{2}$ an claborate combination of induction and analocy is superimposed upon a narrow and limited knowledge of statistical frequency. And this is equally the case in almost all everyday argum onts of any degree of complexity. The class of judgments, which a theory of statistical frequency can comprehend, is too narrow to justify its claim to present a complete theory of probability.
16. Before concluding this chapter, we should not overlook the element of truth which the frequency theory embodies and which provides its plausibility. In the first place, it gives a true account, so long as it does not argue that probahility and frequency are idntical, of a laree number of the most precise arguments in probability, and of these (t) which mathematical treatment is easily applicable. It is this characteristic which has recommended it to statisticians, and explains the large measure of its acceptance in England at the present time: for the popularity in this country of an opinion, which has, so far as I know, no thorough supporters abroad, may reasonably be attributed th the chance which has led most of the Figlish writers, who have paid much attention to probability in reeent years, to approach the subject from the statistical side.

In the second place, the statement that the probability of an event is measured by its actual frequency of occurrence "in the long run' has a very close connection with a valid conclusion which can be derived, in certain cases, from Bernoulli's theorem. This theorem and its commection with the theory of frequemer will be the subject of Chapter XXIX.
17. The absence of a remit exposition of the logical basis of the frequency theory hy any of its adherents has been a areat disadvantage to me in criticising it. It is possible that some of the opmions. which I haw examined at length, are now hedd by no one: nor am I absidut-ly cortain, at the present stage of the inquiry, that a partial rehabilitation of the theory may not be prosible. But I am sure that the objections which I have raised cammot he met without a great complication of the theory and without rohbine it of the simpliciey which is its greatest

[^43]preliminary recommendation. Until the theory has been given new foundations, its logical basis is not so secure as to permit controversial applications of it in practice. A good deal of modern statistical work may be based, I think, upon an inconsistent logical scheme, which, avowedly founded upon a theory of frequency, introduces principles which this theory has no power to justify.

## CHAPTER IX

## THE CONSTRUCTIVE THEORY OF PART I. SCTMMARISED

1. That part of our knowledge which we obtain directly, supplies the premissers of that part wheln we ohtan hy atrument. From these premisses we seek to justify some degree of rational belief about all sorts of conclusions. We do this by perceiving certain logical relations between the premisses and the conclusions. The kind of rational belief which we infer in this manner is termed probable (or in the limit certain), and the logical relations, by the pereeption of which it is obtained, we term relations of probubility.

The probability of a conclusion a derived from premisses $h$ we write $a / h$; and this symbol is of fundamental importance.
2. The object of the Theory or Logic of Probability is to systematise such processes of inference. In particular it aims at elucidating rules by means of which the probabilities of different arguments can be compared. It is of great practical importance to determine whith of two conclusions is on the evidence the more probable.

The most important of these rules is the Principle of Indifference. According to this Principle we must rely upon direct judgment for discriminating between the relevant and the irrelevant parts of the evidence. We can only diseard those parts of the evidence which are irrelevant by seeing that they have no logical bearing on the conclusion. The irrelevant evidence being thus discarded, the Principle lays it down that if the evidence for wither conclusion is the same (i.e. symmetrical). then their probabilities also are the same (i.e. equal).

If, on the other hand, there is additional evidence (i.e. in addition th the symmet rical evidenee) for on wf the comelusions. an! ! his exidnow is furourably relerant, then that conclusion is
the more probable. Certain rules have been given by which to judge whether or not evidence is favourably relevant. And by combinations of these judgments of preference with the judgments of indifference warranted by the Principle of Indifference more complicated comparisons are possible.
3. There are, however, many cases in which these rules furnish no means of comparison ; and in which it is certain that it is not actually within our power to make the comparison. It has been argued that in these cases the probabilities are, in fact, not comparable. As in the example of similarity, where there are different orders of increasing and diminishing similarity, but where it is not possible to say of every pair of objects which of them is on the whole the more like a third object, so there are different orders of probability, and probabilities, which are not of the same order, cannot be compared.
4. It is sometimes of practical importance, when, for example, we wish to evaluate a chance or to determine the amount of our expectation, to say not only that one probability is greater than another, but by how much it is greater. We wish, that is to say, to have a numerical measure of degrees of probability.

This is only occasionally possible. A rule can be given for numerical measurement when the conclusion is one of a number of equiprobable, exclusive, and exhaustive alternatives, but not otherwise.
5. In Part II. I proceed to a symbolic treatment of the subject, and to the greater systematisation, by symbolic methods on the basis of certain axioms, of the rules of probable argument.

In Parts III., IV., and V. the nature of certain very important types of probable argument of a complex kind will be treated in detail ; in Part III. the methods of Induction and Analogy, in Part IV. certain semi-philosophical problems, and in Part V. the logical foundations of the methods of inference now commonly known as statistical.

# PAIT II <br> FUNDAMENTAL THEOREMS 

## CHAPTER X

## INTRODUCTORY

1. Is Part I. we have been occupied with the epistemology of our subject. that is to say, with what we know about the characteristics and the justitication of probable Kinowledge. In Part II. I pass to its Formal Longic. I am not certain of how much positive value this lart will prove to the reader. My object in it is to show that. starting from the philosophical ideas of Part I., we can deduce hy rigorous methods out of simple and precise definitions the usually accepted results. such as the theorems of the addition and multiplication of probabilities and of inverse probability. The reader will readily pereecive that this Part would never have been written except under the influence of Mr. Russell's Principia Wathematicu. But I am simsible that it may suffer from the over-claboration and artificiality of this method without the justification which its grandeur of scale affords to that great work. In common, however, with other examples of formal method, this attempt has had the negative advantage of compelling the author to make his ideas precise and of discovering fallacies and mistakes. It is is part of the spade-work which a conscientious author has to undertake; though the process of doing it masy be of greatur value to him than the results can be to the reader, who is concerned to know, as a saf coruard of the reliability of the rest of the construction, that the thing can be done, rather than to examine the architectural plans in detail. In the development of my own thought, the following chapters have been of great importance. For it was through trying to prove the fundamental theorems of the subject on the hypothesis that Trobability was a relation that I first worked my way into the subject: and the rest of this Treatise has arisen out of attempts to solve the successive questions to which the ambition to treat Probability as a branch of Formal Logic first gave rise.

A further occasion of diffidence and apology in introducing this Part of my Treatise arises out of the extent of my debt to Mr. W. E. Johnson. I worked out the first scheme in complete independence of his work and ignorant of the fact that he had thought, more profoundly than I had, along the same lines; I have also given the exposition its final shape with my own hands. But there was an intermediate stage, at which I submitted what I had done for his criticism, and received the benefit not only of criticism but of his own constructive exercises. The result is that in its final form it is difficult to indicate the exact extent of my indebtedness to him. When the following pages were first in proof, there seemed little likelihood of the appearance of any work on Probability from his own pen, and I do not now proceed to publication with so good a conscience, when he is announcing the approaching completion of a work on Logic which will include " Problematic Inference."

I propose to give here a brief summary of the five chapters following, without attempting to be rigorous or precise. I shall then be free to write technically in Chapters XI.-XV., inviting the reader, who is not specially interested in the details of this sort of technique, to pass them by.
2. Probability is concerned with arguments, that is to say, with the "bearing" of one set of propositions upon another set. If we are to deal formally with a generalised treatment of this subject, we must be prepared to consider relations of probability between any pair of sets of propositions, and not only between sets which are actually the subject of knowledge. But we soon find that some limitation must be put on the character of sets of propositions which we can consider as the hypothetical subject of an argument, namely, that they must be possible subjects of knowledge. We cannot, that is to say, conveniently apply our theorems to premisses which are self-contradictory and formally inconsistent with themselves.

For the purpose of this limitation we have to make a distinction between a set of propositions which is merely false in fact and a set which is formally inconsistent with itself. ${ }^{1}$ This leads

[^44]us to the conception of a gromp of propositions. which is defined as a set of propooitions such that-(i.) if a lorieal principhe belongs to it. all propmsitions which are instances of that logical principle also belong to it; (ii.) if the propnsition $p$ and the proposition ' not- $p$ or $q$ ' both belong to it, then the proposition $q$ also belongs to it ; (iii.) if an! promesition phemesto to it. then the contradictory of $p$ is excludel from it. If the gremp defined by one part of a set of propesitions excludes a propensition which belomes to a group defined by another part of the set. then the set taken as a whole is ircomsistent with itself and is incapable of forming the premiss of an argument.

The conception of a group leads on to a precise definition of one proposition requiring another (which in the realm of assertion corresponds to relerence in the realm of probability), and of lowical priority as being an order of propositions arising out of their relation to thense siecial groups, or real groups, which are in fact the subject of knowledge. Lamical priority has no ahsolute signification, hut is relative to a specific body of knowledre. or, as it has been termed in the traditional logic, to the I fiverse of Reference.

It also enables us to reach a definition of inference distinct from impliention, as defined by Mr. Russell. This is a matter of very great impertance. Readery who are acquainted with the work of Mr. Russell and his followers will probably have noticed that the contrast between his work and that of the traditional lewice is by no means wholly due to the greater precision and more mathomatical character of his technique. There is a differmene also in the design. His ohject is to diseover what assumpmions are required in order that the formal propositions generally amoptod her mathematicians and logivians may low obtainali,

[^45]as the result of successive steps or substitutions of a few very simple types, and to lay bare by this means any inconsistencies which may exist in received results. But beyond the fact that the conclusions to which he seeks to lead up are those of common sense, and that the uniform type of argument, upon the validity of which each step of his system depends, is of a specially obvious kind, he is not concerned with analysing the methods of valid reasoning which we actually employ. He concludes with familiar results, but he reaches them from premisses, which have never occurred to us before, and by an argument so elaborate that our minds have difficulty in following it. As a method of setting forth the system of formal truth, which shall possess beauty, inter-dependence, and completeness, his is vastly superior to any which has preceded it. But it gives rise to questions about the relation in which ordinary reasoning stands to this ordered system, and, in particular, as to the precise connection between the process of inference, in which the older logicians were principally interested but which he ignores, and the relation of implication on which his scheme depends.
' $p$ implies $q$ ' is, according to his definition, exactly equivalent to the disjunction ' $q$ is true or $p$ is false.' If $q$ is true, ' $p$ implies $q^{\prime}$ holds for all values of $p$; and similarly if $p$ is false, the implication holds for all values of $q$. This is not what we mean when we say that $q$ can be inferred or follows from $p$. For whatever the exact meaning of inference may be, it certainly does not hold between all pairs of true propositions, and is not of such a character that every proposition follows from a false one. It is not true that 'A male now rules over England' follows or can be inferred from 'A male now rules over France'; or 'A female now rules over England' from 'A female now rules over France'; whereas, on Mr. Russell's definition, the corresponding implications hold simply in virtue of the facts that 'A male now rules over England' is true and 'A female now rules over France' is false.

The distinction between the Relatival Logic of Inference and Probability, and Mr. Russell's Universal Logic of Implication, seems to be that the former is concerned with the relations of propositions in general to a particular limited group. Inference and Probability depend for their importance upon the fact that in actual reasoning the limitation of our knowledge presents us
with a particular set of propositions. to which we must relate any other proposition ahout which we seek knowledge. The conrse of an argument and the results of reasoning depent, not simply on what is true, but on the particular body of knowledee from which we have set onut. Iltimately, indeed, Mr. Russell canmot avoid concerning himself with groups. For his aim is to disconer the smallest set of propositions which specify our formal knowledere and then to show that ther do in fact specify it. In this enterprise, being human. he must confine himself to that part of formal truth which we know, and the question, how far his axioms comprehend all formal truth, must remain insoluble. But his object, nevertheless, is to establish a train of implieations between formal truths ; and the character and the justification of rational argument as such is not his subject.
3. Passing on from these preliminary reflections. war first task is to establish the axioms and definitions which are to make operative our symbolical processes. These processes are almost entirely a development of the idea of representing a probability by the symbel a $h$. where $h$ is the premiss of an argument and " its conclusion. It might have been a motation more in acourd ance with our fundammal ideas. to have empleved the symbel "h tudesignate the argument from $h$ to $a$, and to have represthtad the probability of the argument, or rather the degree of ratiomal bedief about a which the argument anthorises. by the scouble $\mathrm{P}^{\prime}\left(\mathrm{n}^{\prime}\right)$. This would correspond to the symbel $\mathrm{V}^{\prime}\left(\right.$ a'h $\left.^{\prime}\right)$ which has been emploved in Chapter VI. for the evidential value of the argument as distinct from its probability. But in a section where we are only concerned with probabilities. the use of $P^{\prime}\left({ }^{\prime} / 1\right)$ would have been unneressarily cumberos, and it is, therefore convenient to drop the prefix P' and to denote the probatility itself by $a / h$.

The diecosery of a consenient symbol, like that of an cescom: al word. hars oftom proved of more than verbal importance. Ilwar thinking on the subjere of Prohatility is not pescihbe withome at symbol which takes an explicit account of the premiss of the argument as well as of its conclusion : and emilless confusion has arisen through diamssims about the probability of :a comblasion without referemee th the argument as a whele. I claim. therefone. the introduction of the symbol $a /$ as an essential stop tomaris any progress in the subject.
4. Inasmuch as relations of Probability cannot be assumed to possess the properties of numbers, the terms addition and multiplication of probabilities have to be given appropriate meanings by definition. It is convenient to employ these familiar expressions, rather than to invent new ones, because the properties which arise out of our definitions of addition and multiplication in Probability are analogous to those of addition and multiplication in Arithmetic. But the process of establishing these properties is a little complicated and occupies the greater part of Chapter XII.

The most important of the definitions of Chapter XII. are the following (the numbers referring to the numbers of Chapter XII.) :
II. The Definition of Certainty: $a / h=1$.
III. The Definition of Impossibility : $a / h=0$.
VI. The Definition of Inconsistency: ah is inconsistent if $a / h=0$.
VII. The Definition of a Group : the class of propositions $a$ such that $a / h=1$ is the group $h$.
VIII. The Definition of Equivalence: if $b / a h=1$ and $a / b h=1$ $(a \equiv b) / h=1$.
IX. The Definition of Addition: $a b / h^{2}+a b / h^{1}=a / h$.
X. The Definition of Multiplication: $a b / h=a / b h . b / h=$ $b / a h . a / h$. The symbolical development of the subject largely proceeds out of these definitions of Addition and Multiplication. It is to be observed that they give a meaning, not to the addition and multiplication of any pairs of probabilities, but only to pairs which satisfy a certain form. The definition of Multiplication may be read: 'the probability of both $a$ and $b$ given $h$ is equal to the probability of $a$ given $b h$, multiplied by the probability of $b$ given $h$.'
XI. The Definition of Independence: if $a_{1} / a_{2} h=a_{1} / h$ and $a_{2} / a_{1} h=a_{2} / h, a_{1} / h$ and $a_{2} / h$ are independent.
XII. The Definition of Irrelevance: if $a_{1} / a_{2} h=a_{1} / h, a_{2}$ is irrelevant to $a_{1} / h$.
5. In Chapter XIII. these definitions, supplemented by a few axioms, are employed to demonstrate the fundamental theorems of Certain or Necessary Inference. The interest of this chiefly lies in the fact that these theorems include those which the
traditional Logic has termed the Loters of Thought, as for example the Law of Contradiction and the Law of Exeluded Middle. These are here exhibited as a part of the ereneralised theory of Inference or Rational Argument, which includes probable Inference as well as certain Inference. The object of this chapter is to show that the ordinarily accepted rules of Inferences can in fact be deduced from the definitions and axioms of chapter XII.
6. In Chapter XIV. I proceed to the fundamental Theorems of Probable Inference, of which the following are the most interesting :

Addition Theorom: ( $n: l$ )/h-u/h $b / / h-a h / h$, which reduces to $(a+b) / h=a^{\prime} h+b / h$. where $a$ and $b$ are mutually exclusive : and, if $p_{1} p_{2} \ldots p_{n}$ form, relative to $h$, a set of exclusive and exhaustive alternatives, $a / h=-p_{r}^{n} r_{r} / h_{\text {. }}$.

Theorem of Imelewane: If a/ $h_{1} h_{2}$ wi/h. then a/ $/ h_{1} h_{2}=a, h_{1}$; i.e. if a proposition is irrelecant, its contradictory also is irrelevant.

Theorem if Independence: If $a_{2} /\left(a_{1} h=a_{2} / h, a_{1}^{\prime}, c_{2} h=a_{1} / h\right.$; i.e. if $a_{1}$ is irrelevant to $a_{2} / h$. it follows that $a_{2}$ is irrelevant to $a_{1} / h$ and that $a_{1} / h$ and $a_{2} / h$ are independent.

Multiylication Theorem: If $a_{1} / h$ and $a_{2} / h$ are independent, $a_{1} a_{2} / h=a_{1} / h . a_{2} / h$.

Theorem of Inverse Probability: $\frac{u_{1} / b h}{u_{2} l h / h}=\frac{b_{1}, u_{1} h}{b_{1} / u_{2} h} \cdot a_{1} / h / a_{2} / h$. Further, if $a_{1} / h-p_{1}, c_{2} / h=p_{2}, b / a_{1} h_{1} q_{1}, b_{1}^{\prime} u_{2} h-q_{2}$, and $a_{1} h h_{h} \cdot u_{2} / l_{h} h-1$, then $n_{1} l_{1} h=l_{1}^{\prime} \eta_{1}$; ant if $a_{1} / h n_{2} l_{1} l_{1} n_{1} l_{1} t_{2} q_{1}$. which $\mu_{1}^{\prime} q_{1}+l_{2} I_{2} \quad \eta_{1}: q_{2}$
is equivalent to the siatement that the probability of ${ }_{1}$ when we know $b$ is equal to $\quad q_{1}$, where $q_{1}$ is the prohability of $b$ when $q_{1}: q_{2}$
we know $a_{1}$ and $q_{2}$ its probability when we know $a_{2}$. This theorem emmetated with varying deoremis of inaceuramy appears in all Treatises on Probability, but is not generally proved.

Chapter XIV. concludes whith some clatorate themems on the combination of premioses based on a terdmical symbelic device. knewn as the ('ummintire Formmula, which is the werk of Mr. II. E. Johnson.
7. In Chapter XV. I bring the non-numerical theory of probability developed in the preseding chapters inter commention with the usual mumerical conception of it, and demmontrate how
and in what class of cases a meaning can be given to a numerical measure of a relation of probability. This leads on to what may be termed numerical approximation, that is to say, the relating of probabilities, which are not themselves numerical, to probabilities, which are numerical, by means of greater and less, by which in some cases numerical limits may be ascribed to probabilities which are not capable of numerical measures.

## CHAPTER XI

## THE THEORY OF GROUPS, WITH SPECIAL REFERENCE TO

LOGICAL CONSISTENCE, INFFRENCE, AN゙1) I.O(¥ICAL PRIORITY

1. The Thenry of Probability deals with the relation between $t$ wo sets of propensitions, such that, if the first set is known to be true, the second can he known with the appropriate degree of probability be argument from the first. ${ }^{1}$ 'The relation. however, also exists when the first set is not known to be true and is hypothetical.

In a symbolical treatment of the subject it is important that we should be free to consider hymerlutionl premisses, and to take account of relations of probahility as existing between amy pair of sets of propositions, whe ther or mot the premiss is actually part of knowledge. But in actine thus we must he careful to avoid two possible sources of error.
2. The first is that which is liable to arise wherever variables are concerned. This was mentioned in pasing in $\$ 18$ of (hapter IV. We must remember that whenever we substitute for a variable come particular value of it, this may so aftect the rele vant evidence as to modify the probability. This danger is ahways present except where, as in the first half of Chapter Nlll.. the conclusions respecting the variable are certain.
3. The second difficulty is of it different character. (Our premisses may be hypetherteal and mot actually the suljoent of knowledre. Put must they not be possible subjects of know ledee? How are we to deal with hepothetieal premisses which are self contradictory or formally inconsistent with themsilues. and which cammet be the subject of rational helief of ann deveres?

[^46]Whether or not a relation of probability can be held to exist between a conclusion and a self-inconsistent premiss, it will be convenient to exclude such relations from our scheme, so as to avoid having to provide for anomalies which can have no interest in an account of the actual processes of valid reasoning. Where a premiss is inconsistent with itself it cannot be required.
4. Let us term the collection of propositions, which are logically involved in the premisses in the sense that they follow from them, or, in other words, stand to them in the relation of certainty, ${ }^{1}$ the group specified by the premisses. That is to say, we define a group as containing all the propositions logically involved in any of the premisses or in any conjunction of them ; and as excluding all the propositions the contradictories of which are logically involved in any of the premisses or in any conjunction of them. ${ }^{2}$ To say, therefore, that a proposition follows from a premiss, is the same thing as to say that it belongs to the group which the premiss specifies.

The idea of a 'group' will then enable us to define ' logical consistency.' If any part of the premisses specifies a group containing a proposition, the contradictory of which is contained in a group specified by some other part, the premisses are logically inconsistent; otherwise they are logically consistent. In short, premisses are inconsistent if a proposition 'follows from' one part of them, and its contradictory from another part.
5. We have still, however, to make precise what we mean in this definition by one proposition following from or being logically involved in the truth of another. We seem to intend by these expressions some kind of transition by means of a logical principle. A logical principle cannot be better defined, I think, than in terms of what in Mr. Russell's Logic of Implication is termed a formal implication. ' $p$ implies $q$ ' is a formal implication if ' not- $p$ or $q$ ' is formally true ; and a proposition is formally true, if it is a value of a propositional function, in which all the constituents other

[^47]than the arguments are logical constants, and of which all the values are true.

We might define a group in such a way that all logical principles belonged to every group. In this case all formally true propositions would belong to every group. This definition is logically precise and would lead to a coherent theory. But it possesses the defect of not closely corresponding to the methods of reasoning we actually employ, because all logical principles are not in fact known to us. And even in the case of those which we do know, there seems to be a logical order (to which on the above definition we cannot give a sense) amongst propositions, which are about logical constants and are formally true, just as there is amongst propositions which are not formally true. Thus, if we were to assume the premisses in every argument to include all formally true propositions, the sphere of probable argument would be limited to what (in contradistinction to formally true propositions) we may term empirical propositions.
6. For this reason, therefore, I prefer a narrower definitionwhich shall correspond more exactly to what we seem to mean when we say that one proposition follows from another. Let us define a group of propositions as as set of propositions such that:
(i.) if the proposition " $p$ is formally true' belongs to the group, all propositions which are instances of the same formal propositional function also belong to it ;
(ii.) if the proposition $p$ and the proposition ' $p$ ' implie's $q$ both brlong to it, then the proposition $q$ also belongs to it :
(iii.) if any proposition $p$ belongs to it, then the contradictory of $p$ is excluded from it.

According to this definition all processes of certain inference are wholly composed of steps cach of which is of one of two simple types (and if we like we might perhaps regard the first as comprehending the other). I donot feel certain that these conditions may not be narrower than what we mean when we say that one proposition follows from another. But it is not necessary for the purpose of defining a group, to dogmatise as to whether any other additional methods of inference are, or are mot, open to us. If we define a group as the propositions logically involved in the premisses in the above sense, and preseribe that the premisses of an argument in probability must specify a group not less extensive than this, we are placing the minimum amount of restriction upon
the form of our premisses. If, sometimes or as a rule, our premisses in fact include some more powerful principle of argument, so much the better.

In the formal rules of probability which follow, it will be postulated that the set of propositions, which form the premiss of any argument, must not be inconsistent. The premiss must, that is to say, specify a 'group ' in the sense that no part of the premiss must exclude a proposition which follows from another part. But for this purpose we do not need to dogmatise as to what the criterion is of inference or certainty.
7. It will be convenient at this point to define a term which expresses the relation converse to that which exists between a set of propositions and the group which they specify. The propositions $p_{1} p_{2} \ldots p_{\text {" }}$ are said to be fundumental to the group $h$ if (i.) they themselves belong to the group (which involves their being consistent with one another) ; (ii.) if between them they completely specify the group ; and (iii.) if none of them belong to the group specified by the rest (for if $p_{r}$ belongs to the group specified by the rest, this term is redundant).

When the fundamental set is uniquely determined, a group $h^{\prime}$ is a sub-group to the group $h$, if the set fundamental to $h^{\prime}$ is included in the set fundamental to $h$.

Logically there can be more than one distinct set of propositions fundamental to a given group ; and some extra-logical test must be applied before the fundamental set is determined uniquely. On the other hand, a group is completely determined when the constituent propositions of the fundamental set are given. Further, any consistent set of propositions evidently specifies some group, although such a set may contain propositions additional to those which are fundamental to the group it specifies. It is clear also that only one group can be specified by a given set of consistent propositions. The members of a group are, we may say, rationally bound up with the set of propositions fundamental to it.
8. If Mr. Bertrand Russell is right, the whole of pure mathematics and of formal logic follows, in the sense defined above, from a small number of primitive propositions. The group, therefore, which is specified by these primitive propositions, includes the most remote deductions not only amongst those known to mathematicians, but amongst those which time
and skill have not yet served to solve. If we define certainty in a lugical and mot a pisycholorical sense, it seems necressary, if cour promisses include the essential axioms, to regrard as certain all propmsitions which follow from these, whether or not they are known to us. Yet it seems as if there must ber some locical sense in which unproved mathematical theorens-sone of those, for instance, which deal with the theory of numbers can be likely or unlikely, and in which as proposstion of this kind, which has been suggested to us by analogy or supportul by induction, can possess an intermediate degree of probability.

There can the no douht, I think, that the logical relation of certainty does exist in these cates in which lack of skill or insight prevents our apprembling it, in spite of the fact that sullicient promisess, imcluding -nticiont logical principles, are linmwn to us. In these cases we must say, what we are not permitted to say when the indeterminacy arises from lack of premisese, that the protahility is unlimoun. There is alil a sense. hwover, in which in such a case the knowledge we actually possess can be, in a logical sense, only probable. While the relation of certainty exi-t.s betwem the fundamental axioms and every mathematical hepothesis (or it contradictory), there are other data in relation t) which these hypothoes prosess intermediate degrees of prohability. If we are mable through lack of skill to discover the relation of pobahility which an hypothesis does in fact bear towards one set of data, this set is practically useless, and we must fix our attontion on som other set in relation to which the probability is mot unknown. When Newton held that the binomial theorem persessed for empirical reasons sufticient probability to warrant is lurther investigation of it, it was not in relation to the aximes of mathematics, whet her he knew them or not, that the probatility existed, hut in relation to his empirical evidence combimed, perhaps, with some of the axioms. There is, in short, an exeption th the rule that we must alwas consider the probability of any conclusion in relation to the whole of the data in our perssessim. When the relation of the conclusion to the whele of our evidence cannot be known, then we must be gudded ly its relation to some part of the evidence. When, therefore, in later chapters 1 spuak of a formal propensition as possiosimes an internediate degh.... prolahility, this will alwas lo. in rolation
to evidence from which the proposition does not logically follow in the sense defined in $\S 6$.
9. It follows from the preceding definitions that a proposition is certain in relation to a given premiss, or, in other words, follows from this premiss if it is included in the group which that premiss specifies. It is impossible if it is excluded from the group-if, that is to say, its contradictory follows from the premiss. We often say, somewhat loosely, that two propositions are contradictory to one another, when they are inconsistent in the sense that, relative to our evidence, they cannot belong to the same group. On the other hand, a proposition, which is not itself included in the group specified by the premiss and whose contradictory is not included either, has in relation to the premiss an intermediate degree of probability.

If $a$ follows from $h$ and is, therefore, included in the group specified by $h$, this is denoted by $a / h=1$. The relation of certainty, that is to say, is denoted by the symbol of unity. The reason why this notation is useful and has been adopted by common consent will appear when the meaning of the product of a pair of relations of probability has been explained. If we represent the relation of certainty by $\gamma$ and any other probability by $a$, the product $a \cdot \gamma=a$. Similarly, if $a$ is excluded from the group specified by $h$ and is impossible in relation to it, this is denoted by $a / h=0$. The use of the symbol zero to denote impossibility arises out of the fact that, if $\omega$ denotes impossibility and $a$ any other relation of probability, then, in the senses of multiplication and addition to be defined later, the product $a \cdot \omega=\omega$, and the sum $a+\omega=a$. Lastly, if $a$ is not included in the group specified by $h$, this is written $a / h \neq 1$ or $a / h<1$ : and if it is not excluded, this is written $a / h \neq 0$ or $a / h>0$.
10. The theory of groups now enables us to give an account, with the aid of some further conceptions, of logical priority and of the true nature of inference. The groups, to which we refer the arguments by which we actually reason, are not arbitrarily chosen. They are determined by those propositions of which we have direct knowledge. Our group of reference is specified by those direct judgments in which we personally rationally certify the truth of some propositions and the falsity of others. So long as it is undetermined, or not determined uniquely, which propositions are fundamental, it is not possible to discover
a noenssary ont r amomest propmsitions on to show in what way a true propmsition 'follows from' one true premiss rather than another. But when we have determined what propesitions are fundamental, l,y ardering theo which we kuew directly to he true. or in summ other was. thon a meanige can be attached to priority and to the disinctinn lwanen inference and implication. When the propositims which we know directly are given, there is a logical wher ammes: these nther propositions which we how indirectly and by argument.
11. It will be useful to distinguish between those groups which are heymbimieal and thew of whish the fundamental ent is known th her true. Wia with term the former hypothetical gromes and the latter real groups. To the real group, which contains all the propuritims which are known to be true. We may asciom the old logical term Universe of Reference. While knowledge is here takest the criterion if a real group. What follows will the equally valid whatever criterion is taken, so long as the fundamental set is in some manner or other determined uniquely.

If it is impossible for us to know a proposition $p$ except by inference from a knowledge of $q$, so that we cannot know $p$ to be
 that " $\mu$ tomuins q. Mure precisely requirement is defined as follows:
$p$ does not require $q$ if there is some real group to which $p$ belongs and $q$ does not belong, i.e. if there is a real group $h$ such that $p / h=1, q / h \neq 1$; hence
$p$ requires $q$ if there is no real group to which $p$ belongs and $q$ does not belons.
$p$ does not require $q$ within the group $h$, if the group $h$, to which $p$ belongs, contains a subgroup ${ }^{1} h^{\prime}$ to which $p$ belongs and $q$ does not belong; i.e. if there is is group $h^{\prime}$ such that $h^{\prime} / h=1, p / h^{\prime}=1$, $q / h^{\prime} \neq 1$. This reduces to the proposition next but one above if $h$ is the L'niverse of Reference. In § 13 these definitions will be generalised to cover intermediate degrees of probabilits.
12. Inference and logical priority can he defined in terms of requirement and real groups. It is convenient to distinguish two types of inference corresponding to hypothetical and real

[^48]groups-i.e. to cases where the argument is only hypothetical, and cases where the conclusion can be asserted :

Hypothetical Inference. - 'If $p^{\prime}, q$,' which may also be read ' $q$ is hypothetically inferrible from $p$,' means that there is a real group $h$ such that $q / p h=1$, and $q / h=1$. In order that this may be the case, $p h$ must specify a group; i.e. $p / h \neq 0$, or in other words $p$ must not be excluded from $h$. Hypothetical inference is also equivalent to: ' $p$ implies $q$,' and ' $p$ implies $q$ ' does not require ' $q$.' In other words, $q$ is hypothetically inferrible from $p$, if we know that $q$ is true or $p$ is fu!se and if we can know this without first knowing either that $q$ is true or that $p$ is false.

Assertoric Inference.- ' $p \therefore q$,' which may be read ' $p$ therefore $q$ ' or ' $q$ may be asserted by inference from $p$ ', means that ' if $p, q$ ' is true, and in addition ' $p$ ' belongs to a real group ; i.e. there are proper groups $h$ and $h^{\prime}$ such that $p / h=1, q / p h^{\prime}-1, q / h^{\prime} \neq 1$, and $p / h^{\prime} \neq 0$.
$p$ is prior to $q$ when $p$ does not require $q$, and $q$ requires $p$, when, that is to say, we can know $p$ without knowing $q$, but not $q$ unless we first know $p$.
$p$ is prior to $q$ within the group $h$ when $p$ does hiot require $q$ within the group, and $q$ does require $p$ within the group).

It follows from this and from the preceding definitions that, if a proposition is fundamental in the sense that we can only know it directly, there is no proposition prior to it ; and, more generally, that, if a proposition is fundamonta! to a given group, there is no proposition prior to it within the group.
13. We can now apply the conception of requirement to intermediate degrees of probability. The motation adopted is, it will be remembered, as follows :
$p / h=a$ means that the proposition $p$ has the probable relation of degree a to the proposition $h$; while it is postulated that $h$ is self-consistent and therefore specifies a group.
$p / h=1$ means that $p$ follows from $h$ and is, therefore, included in the group specified by $h$.
$p_{i}^{\prime} h=0$ means that $p$ is excluded from the group specitied by $h$.
If $h$ specifies the Uniwerse of Reference, i.e. if its group comprehends the whole of our knowledge, $p / h$ is called the ulsolute probability of $p$, or (for short) the probubility of $p$; and if $p / h=1$ and $h$ specifies any real group, $p$ is said to be absolutely certain
or (for short) cerlain. Thus $p$ is 'eurtain' if it is a member of a real group, and is certain ' proposition is one which we know to be true. Similarly if $p / h=0$ under the same conditions, $p$ is absolutely impossible, or (for short) impossible. Thus an 'impossible proposition is one which we know to be false.

The definition of requirement, when it is generalised so as to take account of intermediate degrees of prohability, becomes, it will be seen, equivalent to that of relevance:

The probability of $p$ does not require $q$ within the group $h$, if there is a subroup $h^{\prime}$ such that, for every subgroup $h^{\prime \prime}$ which
 amd ! the - $y^{\prime h}$.

When $p$ is included in the group $h$, this defmition reduces to the defmition of requirement given in $\$ 11$.
14. The importance of the theory of groups arises as soon as we admit that there are some propositions which we take for granted without arsument, and that all arguments, whether demonstrative or probable, consist in the relating of other conclusions to these as premisse's.

The particular propositions, which are in fact fundamental to the Universe of Reference, vary from time to time and from person to person. Our theory must also be applicable to hypothetical Universes. Although ab particular Universe of Reference may be defined by considerations which are partly psycholonical, when once the Universe is given, our theory of the relation in which other propositions stand towards it is entirely logical.

The formal development of the theory of argument from
 chapters, resembles in its general method other parts of formal logic. We seek to astablish implications between our primitive axioms and the derivative propositions, without specifie reference to what particular propositions are fundamental in our actual Universe of Ruference.

It will be seen more clearly in the following chapters that the laws of inference are the laws of probiability, and that the former is a particular case of the latter. The relation of a proposition to a group depends upon the relevanee to it of the group, and a group is relevant in so far as it contains a necessabry or sufficient condition of the popmsilion, of a hecestary or sulliciont condition

being neccssary if every hypothetical group, which includes the proposition together with the Universe of Referenc\%, includes the condition, and sufficient if every hypothetical group, which includes the condition together with the Universe of Reference, includes the proposition.

## (HIP1PRR XII

## THE IUEFINITIONS AND ANIOMS OF INFERFNCE AND

 PROIBABILITY1. It is not necessary for the validity of what follows to decide in what manner the set of propositions is determined, which is fundamental to our Universe of Reference, or to make definite assumptions as to what propositions are included in the group which is specified by the dutu. When we are investigating an empirical problem, it will be natural to include the whole of our loyical apparatus, the whole body, that is to say, of formal truths which are known to us, toge ther with that part of our empirical knowledge which is relevant. But in the following formal developments, which are designed to display the logical rules of probahility we need only assume that our data always include those logical rules, of which the steps of our proofs are instances, together with the axioms relating to probability which we shall cnunciate.

The objecet of this and the chapters immediately following is to show that all the usually assumed conclusions in the fundamental logic of inference and probability follow rigorously from a few axioms, in accordance with the fundamental conceptions expounded in Part I. This body of axioms and theorems corresponds, I think, to what logicians have termed the Lowes of Thought, when they have meant hy this something narrower than the whole system of formal truth. But it goes beyond what has heen usual, in dealing at the same time with the laws of probable. as well as of neressarv, inference.
2. This and the following chapters of Part II are largely independent of many of the more controversial issues raised in the preceding chapters. They do not prejudge the question as
to whether or not all probabilities are theoretically measurable ; and they are not dependent on our theories as to the part played by direct judgment in establishing relations of probability or inference between particular propositions. Their premisses are all hypothetical. (ficen the existence of ceriain relations of probability, others are inferred. Of the conclusions of Chapter III., of the criteria of equiprobability and of inequality discussed in Chapters IV. and V., and of the criteria of inference discussed in $\S 55,6$ of Chapter XI., they are, 1 think, wholly independent. They deal with a different part of the subject, not so closely connected with epistemology.
3. In this chapter 1 cuntine myself to Definitions and Axioms.

Propositions will be denoted ber small letters and relations by capital letters. In accordance with common usage, a disjunctive combination of propositions is represented by the sign of addition, and a conjunctive combination by simple juxtaposition (or, where it is necessary for clearness, by the sign of multiplication) : e.g. ' $a$ or $b$ or $c$ ' is written ' $a+b+c$,' and ' $a$ and $b$ and $c$ ' is written ' $a b c$.' ' $a+b$ ' is not so interpreted as to exclude ' $a$ and $b$.' The contradictory of $a$ is written $\bar{a}$.
4. Preliminary Dejinitions:
I. If there exists a relation of probability $P$ between the proposition $a$ and the premiss $h$

$$
u_{i} h=\mathrm{P} \quad \mathrm{D} \cdot \mathrm{f}
$$

II. If P is the relation of certainty ${ }^{1}$

$$
\mathrm{P}=1 \quad \text { Wef. }
$$

III. If P is the relation of impossibility ${ }^{1}$

$$
P=0 \quad \text { Def. }
$$

IV. If P is a relation of probability, but not the relation of certainty $\mathrm{P}<1$. 1) f.
V. If P is a relation of probability, but not the relation of impossibility $\quad \mathrm{P}>0$. Def.
VI. If $a / h=0$, the conjunction $a h$ is inconsistent. Def.
VII. The class of propositions $a$ such that $a / h=1$ is the group specified by $h$ or (for short) the group $h$. Def. VIII. If b/ath 1 and albh $=1$, ( 11 b), h 1 . D

This may be rewarded as the definition of Equiralence. Thus we see that equivalence is relative for aremiss $h$. $\quad$ is equivalunt to $b$, given $h$, if $b$ follows from $a h$, and $a$ from $b h$.

[^49]
## 5. P'relimin ary Arimus.

We shall assume that there is included in every premiss with Which w. are conconed ther fomal implications which allow us to assert the following axioms :
(i.) Provided that $a$ and $h$ are propositions or conjunctions of fropmeitions or disjunctions of propositions, and that $h$ is not an inconsistent conjunction, there exists one and only one relation of probatilite P' hetweon "tas conclusion and / as premiss. Thns any conclusion " bears thany consistent premiss hone and only one relation of probability.
(ii.) If $(n \equiv h) / h=1$, and $x$ is a proposition, $x / a h=x / b h$. This is the Axiom of Equivalence.
(iii.)

$$
\left.\begin{array}{rl}
(u+l,=\bar{u} \bar{b})^{\prime} h & =1 \\
(u, u & \prime \prime
\end{array}\right)=1 .
$$

if $a$ is included in the group specified by $h, h$ and ah are equivalent
6. Addition and IIultiplication.- If we were to assume that probabilities are mumbers or ratios, these operations could be given their usual arithmetical signification. In adding or multiplying probabilities we should be simply adding or multiplying numbers. But in the absence of such an assumption. it is necessary to give a meaning by definition to these processes. I shall define the addition and multiplication of relations of probabilities only for certain types of such relations. But it will be shown later that the limitation thus placed on our operations is not of practical importance.

We define the sum of the probable relations $a b / h$ and $a b / h$ as being the probable relation of ; and the product of the probable relations $a$ 'bl and $b / h$ as being the probable relation $a b / h$. That is to say :


Bufore we proced to the axioms which will make these symbols operative, the definitions may be restated in more familiar languace. IX. may be read: "The sum of the probabilitios of 'both $a$ and $b$ ' and of ' $a$ but not $b$, relative to the same hypothesis, is equal to the probability of ' $a$ ' relative to this hypo-
thesis." X. may be read: "The probability of 'both a and $b$.' assuming $h$, is equal to the product of the probability of $b$, assuming $h$, and the probability of $a$, assuming both $b$ and $h$." Or in the current terminology ${ }^{1}$ we should have: "The probability that both of two events will occur is equal to the probability of the first multiplied by the probability of the second, assuming the occurrence of the first." It is, in fact, the ordinary rule for the multiplication of the probabilities of events which are not 'independent.' It has, however, a much more central position in the development of the theory than has bem usually reengnised.

Subtraction and division are, of course, defined as the inverse operations of addition and multiplication :

$$
\begin{array}{ll}
\text { XI. If } \mathrm{P} Q=R, P=R & \text { Def. } \\
\text { XII. If } P+Q=R, P=R-Q . & \text { Def. }
\end{array}
$$

Thus we have to introduce as definitions what would be axioms if the meaning of addition and multiplication were already defined. In this latter case we should have been able to apply the ordinary processes of addition and multiplication without any further axioms. As it is, we need axioms in order to make these symbols, to which we have given our own meaning, operative. When certain properties are associated, it is often more or less arbitrary which we take as defining properties and which we associate with these by means of axioms. In this case I have found it more convenient, for the purposes of formal development, to reverse the arrangement which would come most natural to commonsense, full of preconceptions as to the meaning of addition and multiplication. I define these processes, for the theory of probability, by reference to a comparatively unfamiliar property, and associate the more faniliar properties with this one by means of axioms. These axioms are as follows :
(iv.) If $P, Q, R$ are relations of probability such that the products $\mathrm{PQ}, \mathrm{PR}$ and the sums $\mathrm{P}+\mathrm{Q}, \mathrm{P}+\mathrm{R}$ exist, then :
(iv. a) If PQ exists, (QP' (xists, and $\mathrm{P}^{\mathrm{Q}} \mathrm{Q}$ - QP. If $\mathrm{P}+\mathrm{Q}$ exists, $\mathrm{Q}+\mathrm{P}$ exists and $\mathrm{P}+\mathrm{Q}=\mathrm{Q}+\mathrm{P}$.
(iv. b) $\mathrm{PQ}<\mathrm{P}$ unless $\mathrm{Q}=1$ or $\mathrm{P}=0 ; \mathrm{P}+\mathrm{Q}>\mathrm{P}$ unless $\mathrm{Q}=0$. $\mathrm{PQ}=\mathrm{P} \quad$ if $\quad \mathrm{Q}=1$ or $\mathrm{P}=0 ; \mathrm{P}+\mathrm{Q}=\mathrm{P} \quad$ if $\quad \mathrm{Q}=0$.
(iv. c) If $\mathrm{PQ}, \mathrm{PR}$, then Q R unless $\mathrm{P}=0$. If $\mathrm{P}+\mathrm{Q} \mathrm{P}+\mathrm{R}$, then $\mathrm{Q} \cdot \mathrm{R}$ and conversely.

A meaning has not been given, it is important to notice, to the signs of addtition and multupliation between pmokhilitios in all mases. Acemding th the definitions we have givm. I' - (e) and $P Q$ have not an interpretation whenever $P$ and $Q$ are relations of pohblilit! but in certain conditions only. Faurther more, if $P-Q=R$ and $Q-S+T$, it does not follow that $\mathrm{P}+\mathrm{S}+\mathrm{T} \mathrm{R}$, since no meaning has been assigned to such an expression as $\mathrm{P}+\mathrm{S}+\mathrm{T}$. The equation must be written $\mathrm{P}+(\mathrm{S}+\mathrm{T})$ $=P$, and we cannot infer from the foregoing axioms that $(\mathrm{P}+\mathrm{S})+\mathrm{T}-\mathrm{R}$. The following axioms allow us to make this and other inferences in cases in which the sum P -S exists, i.e. when $\mathrm{P}+\mathrm{S}=\mathrm{A}$ and A is a relation of probability.
(v.) $[ \pm \mathrm{P} \pm(\mathrm{Q}]+[ \pm \mathrm{R} \pm \mathrm{S}]=[ \pm \mathrm{P}=\mathrm{R}]-[\mp(\mathrm{Q} \pm \mathrm{S}]=\lfloor \pm \mathrm{P} \pm \mathrm{R}]+$

$$
Q \sim-\quad|\quad P: Q|-1 \quad R ; S
$$

in every case in which the probabilities $[ \pm \mathrm{P} \pm \mathrm{Q}],[ \pm \mathrm{R} \pm \mathrm{S}]$, $[=\mathrm{P} \pm \mathrm{R}]$, etc., exist, i.e., in which these sums satisfy the conditions necessary in order that a meaning may be given to them in the terms of our definition.
(vi.) $P(R \pm S) P R \pm P S$, if the sum $R \pm S$ and the products PR and PS' exist as probabilities.
7. From these axioms it is possible to derive a number of propositions respecting the addition and multiplication of prob)-

 and $\mathrm{S} \cdot(\mathrm{Q}$ - xist ; and that $(\mathrm{P}+(\mathrm{Q})(\mathrm{R}+\mathrm{S})=(\mathrm{P}+\mathrm{Q}) \mathrm{R}+(\mathrm{P} \cdot(\mathrm{Q}) \mathrm{S}=$
 the sums and products in question exist. In general any rearrangement which would be legitimate in an equation between arithmetic quantities is also legitimate in an equation between probabilities, provided that our initial equation and the equation which finally results from our symbolic operations can both be expressed in a form which contains only products and sums which have an interpretation as probabilities in accordance with the definitions. If, therefore, this condition is ohserved, we need not complicate our operations by the insertion of brackets at every stage, and no result can be obtained as a result of leaving them out, if it is of the form preseribed above, which could not be obtained if they had been rigorously inserted throughout. We can only be interested in our final results when they deal with actually existent and intelligible probabilities- for our object is,
always, to compare one probability with another-and we are not incommoded, therefore, in our symbolic operations by the circumstance that sums and products do not exist between every pair of probabilities.
8. Independence:
XIII. If $a_{1} / a_{2} h=\prime_{1} / h$ and $a_{2} / a_{1} h=a_{2} / h$, the probabilities $a_{1} / h$ and $a_{2} / h$ are independent.

Def.
Thus the probabilities of two arguments having the same premisses are independent, if the addition to the premisses of the conclusion of either leaves them unaffected.

Lirelevance : ${ }^{1}$
XIV. If $a_{1} / a_{2} h=a_{1} / h, a_{2}$ is irrelevant on the whole, or, for short, irrelevant to $a_{1} / h$.

Def.
1 This is repeated for convenience of reference from Chapter IV. §14. It is only necessary here to take account of irrelevance on the whole, not of the more precise sense.

## CHAPTER XIII

## THE FUNDAMENTAL THEOREMS OF NECESSARY INFERENCE

1. In this chapter we shall be mainly concerned with deducing the existram of relatonis of erertanty or impossibility, wiven ot her relations of certainty or impossibility, -with the rules, that is to say, of Certain or, as De Morgan termed it, of Necessary Inference. But it will be convenient to include here a few theorems dealing with intermediate degrees of probability. Except in one or two important cases I shall not trouble to translate these theorems from the symbolism in which they are expressed, since their interpretation presents no difficulty.

## $\therefore$ (1) a 4 it if !



[^50]\[

$$
\begin{equation*}
a b / h<b / h \text { or } a b / h=b / h \quad \text { by } X \text {. and (iv. } b \text { ). } \tag{4}
\end{equation*}
$$

\]

(5) If P and Q are rolations of probability and $\mathrm{P}+\mathrm{Q}=0$, then $\mathrm{P}=0$ and $\mathrm{Q}=0$.
and

$$
\begin{array}{cc}
\mathrm{P}+\mathrm{Q}>\mathrm{P} \text { unless } \mathrm{Q}=0 & \text { by (iv. } h \text { ) } \\
\mathrm{P}>0 \text { unless } \mathrm{P}=0 & \text { by } \mathrm{V} .
\end{array}
$$

$$
\therefore P+Q>0 \text { unless }(\ell=0 \text {. }
$$

Hence, if $\mathrm{P}+\mathrm{Q}=0, \mathrm{Q}=0$ and similarly $\mathrm{P}=0$.
(6) If $\mathrm{PQ}=0, \mathrm{P}=0$ or $\mathrm{Q}=0$,

$$
Q>0 \text { unless } Q=0 \quad \text { by } V \text {. }
$$

Hence

$$
\mathrm{PQ}>\mathrm{P} .0 \text { unless } \mathrm{Q}=0 \text { or } \mathrm{P}=0
$$ by (iv.c), i.e. $\quad \mathrm{PQ}>0$ unless $\mathrm{Q}=0$ or $\mathrm{P}=0 \quad$ by (iv. b).

Whence, if $\mathrm{PQ}=0$, the result follows.
(7) If $\mathrm{PQ}=1, \mathrm{P}=1$ and $\mathrm{Q}=1$,

$$
\begin{array}{cc}
\mathrm{PQ}<\mathrm{P} \text { unless } \mathrm{P}=0 \text { or } \mathrm{Q}=1 & \text { by (iv. b), } \\
\mathrm{PQ}=\mathrm{P} \text { if } \mathrm{P}=0 \text { or } \mathrm{Q}=1 & \text { by (iv. } b) . \\
\mathrm{P}<1 \text { unless } \mathrm{P}=1 & \text { by } \mathrm{IV} .,
\end{array}
$$

and

$$
\therefore \mathrm{PQ}<1 \text { unless } \mathrm{P}=1 .
$$

Hence $\mathrm{P}=1$; similarly $\mathrm{Q}=1$.
(8) If $a / h=0, a b / h=0$ and $a / b /=0$ if $h / h$ is not inconsistent.

and since $a / h=0, \quad b / u t h . a / h=0 \quad$ by (iv. b),

$$
\begin{aligned}
& \therefore \text { wh/ji=0 and } \| \mid\left(h / l_{1}, / / h=0,\right. \\
& \therefore \text { unless }\|/ h \cdots,\| / h=0) \quad \text { ly }(5) .
\end{aligned}
$$

whence the result by VI.
Thus, if a conclusion is impossible, we may add to the conclusion or add consistently to the premisses without affecting the argument.
(9) If $u / h=1$. a/hik $=1$ if $l / /$ is inot inconsistmit.
since $\quad " 1 i=1$,

$$
\bar{a} / / \overline{=}=0
$$

by (1.1),
$\therefore \bar{a} / b h=0$ by (8) if $\bar{b} h$ is not inconsistent, whence
" $/ \|, h=1$
by (1.4).
Thus we may add to premisses, which make a conclusion certain, any other premisses not inconsistent with them, without affecting the result.


Since $\quad \| / i=1, \quad$ / $/ h / h=1$ hy ( 9 ) unless $h, l_{1}=(1)$,
$\therefore$ illuth. "/lh - li/nh and "/bh. hilh =-l!!h by (iv.b).
whence the result. manss iflh=0.


This is the principle of equivalence. In virtue of it and of
 wherever they occur in a probability whose premisses include $h$.

ド,


(13.1) $\bar{a} / u=0$, unless $"$ is inconsistent. This follows from (13) atm (isl).
(13.2) "/ $\bar{a}=()$, unless $\bar{a}$ is inconsistent. This follows from (iii). lis whiting af fin a in (13.1).

Le $f$ be the group of assumptions, common to $a$ and $b$, which we have supposed to be included in every real group; 110
and

$$
\begin{aligned}
& \text { auth th? atm the live live (iii) and (12). } \\
& \text { whf def ide wis } \\
& \text { lis. }
\end{aligned}
$$

Site:

> why " ha lypmhars.
and

May 1 .
whom,
14. 11

Thus, if a is impmesihle given h, then $b$ is impossible given a.
 $h_{1}^{\prime} 2^{\prime h} h_{1} h_{2} h^{2} \cdot h_{2} h \quad$ ley 11 .
 the result by (iv. h) unless $/ 1 / h_{2}=0$.

If

$$
\begin{equation*}
h_{1} / h_{2}=0 . \quad h_{2} / h_{1}=0 \tag{14}
\end{equation*}
$$

since we assume that $h$ is not inconsistent, and hence

$$
\begin{equation*}
h_{1} h_{2} / h=0 \tag{8}
\end{equation*}
$$

Thus, if $h_{1}$ is impossible given $h_{2}, h_{1} h_{2}$ is always impossible and is excluded from every group.
(15.1) If $h_{1} h_{2} / h_{0}=0$ and $h_{2} h_{h}$ is not inconsistent, $h_{1} / h_{2} h=0$. This, which is the converse of (15), follows trom $X$. and (6).
(16) If $h_{1} / h_{2}=1,\left(h_{1}+\bar{h}_{2}\right) / h_{h}=1$,

$$
\begin{aligned}
& \bar{h}_{1} / h_{2}=0 \quad \text { by (1). } \\
& \therefore \bar{h}_{1} h_{2} / h=0 \text { by (15), } \\
& \therefore \Pi_{1} h_{2} / h_{1}=1 \quad \text { by }(1.3) \text {, } \\
& \therefore\left(/_{1}+\vec{K}_{2}\right) / h=1 \quad \text { hy (12) and (iii.). }
\end{aligned}
$$

(16.1) We may write (16):

If $h_{1} / h_{2}=1,\left(h_{2} \partial h_{1}\right) / h=1$, where ' ${ }^{\prime}$ ' symbolises 'implies.' Thus if $h_{1}$ follows from $h_{2}$, then it is always certain that $h_{2}$ implies $h_{1}$.
(16.2) If $\left(h_{1}+\bar{h}_{2}\right) / \hbar=1$ and $h_{2} h_{i}$ is not inconsistent, $h_{1} / h_{2} h_{h}=1$.

$$
\bar{h}_{1} h_{2} / i \prime=0, \text { as in }(16),
$$

$\therefore \bar{h}_{1} / h_{2} h=0$ by (15.1), since $h_{2} h_{2}$ is not inconsistent,

$$
\begin{equation*}
\therefore h_{1} / h_{2} h_{i}=1 \tag{1.4}
\end{equation*}
$$

This is the converse of (14).
(16.3) We may write (16.2) :

If $\left(h_{2} \supset h_{1}\right) / h_{i}=1$ and $h_{2} / h_{\text {is }}$ not inconsistent, $h_{1} / h_{2} h_{1}: 1$. Thus, if we define a ' group' as a set of propositions, which fullow from and are certain relatively to the proposition which specifies thern, this proposition proves that, if $h_{2} \partial h_{1}$ and $h_{2}$ belons to a group $h_{2} h_{\text {, then }} h_{1}$ also belongs to this group.
(17) If ( $\left.h_{1} \supset: a \equiv b\right) / h=1$ and $h_{1} h^{2}$ is not inconsistent, $/ / h_{1} h^{h}$ $=b / h_{\mathbf{1}} \mathrm{h} . \quad$ This follows from (16.3) and (12).
(18) $a / a=1$ or $\bar{a} / \bar{a}=1$.

$$
a / a=1 \text {, unless } \text { a is inconsistent, by (13). }
$$

If $a$ is inconsistent, $" / h==\|$, where $~ h$ is not inconsistent. and therefore

$$
\begin{equation*}
\bar{a} / h=1 \tag{1.3}
\end{equation*}
$$

Thus unless $a$ is inconsistent. $\bar{a}$ is not ineonsistent, and tirerefore

$$
\begin{equation*}
\bar{a} / \bar{a}==1 \tag{13}
\end{equation*}
$$

(19) $\quad n a \bar{a} / h-0$,

$$
\begin{aligned}
& \bar{a} / \bar{a}=1 \text { or } " / c .1 \\
\therefore & \text { by }(18), \\
& \text { by }=0
\end{aligned} \text { or } \bar{a} / " \ldots n \text { by (1.1) and }(1.2) .
$$

In either case $\omega \bar{a} / \hbar=0$ by (15).

Thus it is impossible that both a and its contradictory should be true. This is the Law of Contradiction.
$(\because 1)(\cdots+\pi)!=1$.
Since

Thus it is certain that whor a or its contradiotory is true. This is the Law of Excluded Middle.


and $\bar{l} \hat{l}_{2} \underbrace{}_{2} i_{2} \bar{a}_{1} \cdot \bar{a} b_{1}$ by X.,


$$
\therefore A_{1} 1_{1}=0 \text { or } 1_{1} H_{2} 0 \text {. }
$$

and

$$
" 11_{2} 1 \text { ur } 2 i_{1} 11 \text {. }
$$

$$
\therefore A_{2} \text { " or } 2 / 1{ }^{11} \text {. }
$$

In either case $h_{1} h_{2} / h_{1}=0$ by (15).
Thus, if a proposition is certain relatively to one set of premissers, athe impossible relatively 10 another set, the two sto are incompatible.

$$
\begin{aligned}
& \therefore \text { ill ally (asise of } 0 \text {. }
\end{aligned}
$$

$$
\begin{aligned}
& \therefore \text { at iol } 11 \text { or hat } 11 \text {. }
\end{aligned}
$$

$$
\begin{aligned}
& \text { whent•隹 (1) by (1.4). }
\end{aligned}
$$

$$
\begin{aligned}
& (a \bar{a} \equiv(l+\bar{a}) / h=1 \quad \text { by (iii.). }
\end{aligned}
$$

$$
\begin{aligned}
& \therefore(1, \bar{a}) 1 \text { h h (1.3). }
\end{aligned}
$$

## CHAPTER XIV

THE FUNDAMENTAL THEOREMS OF PROBABLE INFERENCE

1. I shall give proofs in this chapter of most of the fundamental theorems of Probability, with very little comment. The bearing of some of them will be discussed more fully in Chapter XVI.

## 2. The Addition Theorems:

(21) $(u+b) / h=u / h+u / h-u l_{1} / h$.

In IX. write $(a+b)$ for $a$, and $\bar{a} b$ for $b$.
Then

$$
\begin{array}{ll}
\text { Then } & (a+b) \bar{a} b / h+(a+b) \bar{a} \bar{b} / h=(a+b) / h, \\
\text { whence } & \bar{a} b / h+(a+b)(\alpha+\bar{b}) / h=(a+b) / h \quad \text { by (iii.), } \\
& \bar{a} / h \cdot h \cdot l / h+(l / h=(k+h) / h \quad \text { by (iii.) and IX. }
\end{array}
$$

That is to say, $(a+b) / h=a / h+(1-a / b h) \cdot b / h$,

$$
=a / h+b / h-c b / h .
$$

In accordance with the principles of Chapter XII. §6, this should be written, strictly, in the form $a / h+(b / h-a b / h)$, or in the form $b / h+(a / h-a b / h)$. The argument is valicl. since the probability $(b / h-a b / h)$ is equal to $\bar{b} b / h$, as appears from the preceding proof, and, therefore, exists. This important theorem gives the probability of " $a$ or $b$ ' relative to a given hypothesis in terms of the probabilities of ' $a$,' ' $b$,' and ' $a$ and $b$ ' relative to the same hypothesis.
(24.1) If $a b / h=0$, i.e. if $a$ and $b$ are exclusive alternatives relative to the hypothesis, then

$$
\left(11+h_{1}\right) / / h=u / h+l_{1} / h .
$$

This is the ordinary rule for the addition of the probabilities of exclusive alternatives.
$(21.2) \quad\left(c_{1} / h+\bar{a} h_{1} / h \quad, \quad, / h\right.$,
since

$$
a l_{1}+\bar{a} l_{l} \equiv l, \quad \text { by (iii.), }
$$

and

$$
\alpha \bar{a} \bar{b} / h=0
$$

by (19) and (8).
(24.3) $(a+b) / h=a / h+b \bar{a} / h$. This follows from (24) and (24.2).





$$
\cdot(-1) \quad / 1!2 \cdots /, \quad i
$$

(2.6) If $p . p_{t} / h{ }^{(1)}$ for all pairs of values of $s$ and $t$, it follows by repeaterd application of $X$. that

$$
1 \cdot l_{2} \ldots f_{1}+\frac{\ddot{1}}{1}
$$


 exhaustive alternatives, then

$$
\because \quad 1
$$

(25) If $p_{1} p_{2} \ldots p$ form, relative to $h$, a set of exclusive and exhatustive alternatives,
́:




$$
\therefore \text { raini } 11 \text { h } \because \text { is is is am inmomist nt. }
$$

11..1世":
בy in ! !

## 1.

Also
F...

Summin:

If $a h$ is inconsistont, i.e. if $a / h=0$ (for $h$ is by hypothesis consistent), the result follows at once by (8).


$$
\begin{gathered}
1 \\
2 \\
1
\end{gathered}
$$

(26) $\quad(1 / h=(a+\bar{h}) / h$.

For

$$
\begin{aligned}
(\prime+\bar{h}) / h & =a / h+\bar{h} / h-(\bar{h} / / h \quad \text { by }(24), \\
& =a / h \quad \text { by (13.1) and (8). }
\end{aligned}
$$

(26.1) This may be written

$$
a / h=(h \partial a) / h .
$$

(27) If $(u+b) / h=0, u / h=0$.

$$
\begin{gather*}
\text { u/h+ }[t / h-(t h / h]=0, \text { by }(24) \text { and bopothesis } \\
\therefore u / h \cdots u \tag{r.}
\end{gather*}
$$

(27.1) If $\quad 1 / h=0$ and $i / h=0,(a+b) / h=0$. This follows from (24).
(28) If $u / k=1,(a+\bar{b}) / k=1$,

$$
(a+\bar{b}) / h=u / h+\bar{b} \bar{a} / h \quad \text { by }(2.4 .3) \text {, }
$$

whence $(a+\bar{b}) / h=a / i=1$ by (1.1) and (8), together with the hypothesis. That is to say, a certain proposition is implied by every proposition.
(28.1) If $a / h=0,(\bar{a}+b) / h=1$ by substituting $\bar{a}$ for: $a$ and $b$ for $\bar{b}$ in (28). That is to say, a certainly false proposition implies every proposition.
(29) If $\quad /\left(h_{1}+l_{2}\right)=1, u / h_{1}=1$,

$$
\bar{a} /\left(h_{1}+h_{2}\right)=0,
$$

and

$$
\therefore \bar{a}\left(h_{1}+h_{2}\right) / h_{1}=0 \quad \text { by }(15) \text {. }
$$

Hence

$$
\begin{equation*}
\bar{a} h_{1} / h_{1}=0 \tag{27}
\end{equation*}
$$

whence the result.
(29.1) If $a / h_{1}=1$ and $a / h_{2}=1, a /\left(h_{1}+h_{2}\right)=1$.

As in $(20) \quad \bar{a} h_{1} /\left(h_{1}+h_{2}\right)=0$ and $\bar{a} h_{2} /\left(h_{1}+l_{2}\right)=0$.
Hence

$$
\bar{a}\left(h_{1}+h_{2}\right) /\left(h_{1}+h_{2}\right)=0 \quad \text { by }(2-1)
$$

whence the result.
(29.2) If $a /\left(h_{1}+h_{2}\right)=0, a / h_{1}=0$. This follows from (29).
(29.3) If $a / h_{1}=0$ and $a / h_{2}=0, a /\left(h_{1}+h_{2}\right)=0$. This follows from (29.1).

## 3. Irrelevance and Independence:

(30) If $a / h_{1} h_{2}=a / h_{1}$, tlien $u / h_{1} \bar{h}_{2}=a / h_{1}$, if $h_{1} \bar{h}_{2}$ is not inconsistent.

$$
\begin{align*}
& w_{1}^{\prime} h_{1}=\left(u h_{2} / h_{1}+u \bar{h}_{2}^{\prime} / \|_{1}\right.  \tag{24.2}\\
& =r_{1}^{\prime} / h_{1} h_{2} \cdot h_{2} h_{1}+a_{1}^{\prime} h_{1} \overline{h_{2}} \cdot \overline{h_{2}} / h_{1}, \\
& =a l^{\prime} / l_{1} \cdot l_{2}^{\prime} h_{1}+a / h_{1} J_{2} \cdot J_{2} / l_{1} \text {, }
\end{align*}
$$

whence $"!/ h_{1}=r!/ h_{1} \bar{h}_{2}$, unless $\bar{h}_{2} / h_{1}=0$, i.e. if $h_{1} \bar{h}_{2}$ is not inconsistent.

Thas, if a propmition is inconsant of an argument, then the contradictory of the proposition is also irrelevant.
(31) If $u_{2} u_{1} h \quad a_{2} h$ and $u_{2} h$ is ton inemsistent, $t_{1} / u_{2} h a_{1} h$.

This follows by (iv. c), since $a_{2} / u_{1} h \cdot a_{1} / h=a_{1} / a_{2} h \cdot a_{2} / h$ by X . If, that is to say, $a_{1}$ is irrelevant to the argument $a_{2} / h$ (see XIV.) and $n_{2}$ is mot imon-istmot with $h_{h}$ : then $n_{2}$ is imeleromet (t) the argumenc $n_{1}$ h and $n_{1} / h$ and $n_{2}$ /h arn indepemtont (s.ee Xlll.)
4. Theorems of Relevance:



ah. ho h he hyotheris:
$h_{1}{ }^{\prime} a h>h_{1}{ }^{\prime} h$
Thus if $h_{1}$ is favourably relevant to the argument $a / h, a$ is favourably relevant to the argument $h_{1} / h$.

This constitutes a formal demonstration of the generally accepted principle that if a hypothesis helps to explain a phenomenon, the fact of the phemomenon supports the reality of the hyporthesis.

In the following theorems $p$ will be said to be more favourable to $a / h$, than $q$ is to $b / h$, if ${ }_{a / h}^{a / p^{h}>}>_{b / h}^{b / h}$, i.e. if, in the language of $\S 8$ below, the coefficient of influence of $p$ on $a / h$ is greater than the coeflicient of influence of $q$ on $b / h$.
(33) If $c$ is favourable to $a / h$, and $h_{1}$ is not less favourable to $a / h x$ than $x$ is to $a / h h_{1}$, then $h_{1}$ is favourable to $a / h$.
 second term on the right is greater tham unity and the product of the third and fourth terms is greater than or equal to unity.
(33.1) A forliori, if $s$ is favourable to a'h and not favourable to $a / h h_{1}$. and if $h_{1}$ is not unfavourable to $a / h x$, then $h_{1}$ is favourable to $a^{\prime} h$.
(3-1) If $x$ is fasourable to aith and $h_{1}$ is not less favourable to $x / h a$ than $x$ is to $h_{1}$ 'ha, then $h_{1}$ is favourable to $a / h$.

This follows by the same reasoning as (:3). since by an application of the Maltiplication Theorem
(35) If $x$ is favourable to $a / h$, but not more favourable to it than $h_{1} x$ is, and not less avonmable to it than to $a / h h_{1}$, then $h_{1}$ is favourable to a! ! $h_{1}$.

This result is a little more substantial than the two preceding. By judgimes the influence of $x$ and $h_{1} r$ on the arguments $a / h$ and alih. We can: inter the intuence of $h_{1}$ by itself on the argument $a / h$.

## 5. The Multiplication Theorems:


 and since $\alpha_{1} / h$ and $\mu_{2} / h$ are independent,

Therefore

$$
{ }^{\prime} 1_{2}^{\prime \prime} / h="_{1} \quad \cdot{ }^{\prime}{ }_{2} \|
$$

Hence, when $a_{1} / h_{\text {and }}$ and are independent, we can arrive at the probability of $a_{1}$ and $a_{2}$ jointly on the same hypothesis by simple multiplication of the probabilities $a_{1} / h$ and $a_{2} / h$ taken separately.
(37) If $\mu^{\prime} 1^{\prime} h=\mu^{\prime} 2^{\prime} I^{\prime} 1^{i} \cdots l^{\prime} 3^{\prime} t^{\prime} I^{\prime} 2^{\prime}:=\cdots$.

$$
I^{\prime} 1 I_{2}^{\prime}!_{3}^{\prime} 3 \cdot l^{\prime}!l^{\prime} / 1=\left\{/_{1} / l_{1}\right\}
$$

 applications of $X$.
6. The Inverse Principle:
 each consistent.
For
and

$$
{ }_{1}\left|l_{1,1}, l_{1} l_{i} \quad\right| l_{1} l_{1} \cdot "_{1} l_{1},
$$

$$
"_{2}!l_{1} \cdot i l_{1} \quad i_{1}^{\prime} m_{2} l_{1} \cdot \|_{2}!h_{1} \quad \text { by X. }
$$ whence the result follows, since $\quad, / h \neq 0$, unless $b / h$ is inconsistent.

 $a_{1} / b h+a_{2} / b h=1$, then it easily follows that

$$
\begin{aligned}
& \text { a/h.r al'mimex } x_{1} \text { liar }^{\prime} h_{1} \text { /haxi }
\end{aligned}
$$


ant


The proposition is easily extended to the cases in which the number of $a$ 's is greater than two.

It will be worth while to translate this theorem into familiar langnage. Let $b$ represent the occurrence of an event $\mathrm{B}, a_{1}$ and $a_{2}$ the hypotheses of the existence of two possible causes $A_{1}$ and $A_{2}$ of B , and $h$ the reneral data of the problem. Then $p_{1}$ and $p_{2}$ are the $\dot{a}$ priori probabilities of the existence of $\Lambda_{1}$ and $A_{2}$ respectively, when it is not known whether or not the event $B$ has occured; $q_{1}$ and $q_{2}$ the probabilities that each of the causes $A_{1}$ and $A_{2}$, if it exists, will be followed by the event B. Then
$\not \rho^{\prime} 1 / 1$ and 1212 are the probabilities of the existence $\mu_{1} \xi_{1} \because \mu \eta_{2} \quad p_{1} q_{1} \quad l_{2} ?$ ?
of $A_{1}$ and $A_{2}$ respectively after the event, i.e. when, in addition to our other data, we know that the event B has occurred. The initial condition, that bh must not be inconsistent, simply ensures that the problem is a possible one, i.e. that the occurrence of the event $B$ is on the initial data at least possible.

The reason why this theorem has renerally been known as the Inverse Princijele of Probability is obvious. The causal problems to which the C'alculus of Probability has been applied are naturally divided into two classes.- - the direct in which, given the cause, we deduce the effect: the indirect or inverse in which, given the effect, we investigate the causr. The Inverse Principle has heen usmally employed to doal with the latter class of problem.

## 

The Multiplieation Theorems given above deal with the com-
 relation of ${ }^{1} 1_{1} h_{2} / h$ to these probabhilitios. In this paragraph the correspondinet prothen of the combination of premisses will be
treated; given $a / h_{1}$ and $a / h_{2}$ we shall consider the relation of $a / h_{1} h_{2}$ to these probabilities.

where " is the a prior probability of the conclusion a and both hypotheses $h_{1}$ and $h_{2}$ jointly, and $v$ is the $i$ prioni probability of the contradictory of the conclusion and looth hypotheses $h_{1}$ and $h_{2}$ jointly.
where $p=\mu / 1 h_{1}$ and $y=-1 / h_{2}$.
and increases with

$$
/_{2}{ }^{\prime \prime \prime l_{1}}{ }_{1}
$$

$$
/_{2} / \bar{a} h_{1}
$$

These results are not very valuable and show the need of an original method of reduction. This is supplied by Mr. W. E. Johnson's Cumulative Formula, which is at present, umpublished but which I have his permission to print below. ${ }^{1}$
8. It is first of all necessary to introduce an new symbol. Let us write
XV. wilh $=\left\{a^{h} h\right\}^{4}!/!$ Def.

We may call $\left\{a^{\prime} b\right\}$ the confficient of influcme of $b$ upon a on hypothesis $h$.

 These coefficients, thus belong be definition to a memeral class of operators, which we may call separative factors.

since

$$
\begin{equation*}
a h / h=a / b h \cdot b / h . \tag{41}
\end{equation*}
$$

1 The substance of propositions (41) to (49) below is derived in its entirety from his notes,- the exposition only is mine.

$$
\begin{aligned}
& h_{2} / a h_{1} \cdot{ }^{\prime} \\
& =h_{2} /\left(\mu h_{1} \cdot l^{\prime}+h_{2} / \bar{a} h_{1} \cdot(1-\mu)^{\prime}\right.
\end{aligned}
$$

Thus we mav also call $\{$ a'b\} the coeffirient of deqendence butwern $a$ and $b$ on hypothesis $h$.

(11.2) And in general
 {音}.
sim
sime
wh th...he wh .wh.il

```
(42.2) And in general we have a commutative rule, by which the order of the terms may be always commuted e.f.
(43) As a multiplier the separative factor operates so as to separate the terms that may be associated (or joined) in the multiplicand.
T:ns
for

Similarly (for example)

For

(44.1) If \(\left\{a^{\prime} b\right\}=1\), it follows that \(a / h\) and \(b / h\) are independent arguments; and conversely.

For arlilh = whllh by (vi.) and (12).
(!:) The ('mmmiatire Formuln
, .... : ". ". "n
1, \(1: 3-1\)
Take " +1 propositions ", \(n, \cdots \ldots\) Then by repetition

which mav be written

\section*{\(n+1 \quad n+1 \quad n+1\)}

\[
=(x / h)^{n+1} \Pi \quad \Pi a / x h:\left(x^{\prime} / h\right)^{n+1} \Pi u / x^{\prime} h: \ldots
\]

Now
\(x / h a b e . . .: x^{\prime} /\) habe . . . : \(x^{\prime \prime} /\) habe . . .
\[
=x / h \cdot(a b c \ldots) / x h: x^{\prime} / h \cdot(r b c \ldots) / x^{\prime} h: \ldots \text { by (38), }
\]
and
\[
a b c \ldots\left|x h=\left\{a^{x h} b^{x h} c \ldots\right\} \Pi a\right| x h \quad \text { by (41.2), }
\]



which may be written
\[
(x / h)^{n} . x / h a b c \ldots \propto\left\{a^{\text {fli}} b^{n h} c \ldots\right\} . x / a h . x / b h . x / c h \ldots
\]
where variations of \(x\) are involved.
The cumulative formula is to be applied when, having accumulated the evidence \(a, b, c \ldots\), we desire to know the comprarative probabilities of the various possible inferences \(x, x^{\prime}\). . which may be drawn, and already know determinatrly the foree of each of the items \(a, b, c \ldots\) separately as evidence for \(x, x^{\prime} \ldots\)

Besides the factors \(x / a h, x / b h\), etc., we require to know two other sets of values, viz.: (1) \(x / h\), etc., i.e. the \(\grave{a}\) priori probabilities of \(x\), etc., and (2) \(\left\{a^{\text {sh }} b^{\text {th }} c \ldots\right\}\), etc., i.e. the coefficients of dependence between \(a, b\), an! \(c \ldots\) on hypotheses \(x h\), etc. It may be remarked that the values \(\left\{r^{\text {sh }} b^{\text {th }} c \ldots\right\}\), \(\left\{a^{\prime \prime} l^{\prime \prime \prime}{ }^{\prime} \ldots\right\} \ldots\) are not in any way related, even when \(x^{\prime} \equiv \bar{x}\).

What corresponds to the cumulative formula has been employed, sometimes, by mathematicians in a simplified form which is, except under special conditions, incorrect. First, it has been tacitly assumed that \(\left\{"^{\prime \prime} l^{\prime} \ldots \ldots\right\} \cdot\left\{"^{\prime \prime \prime} \|_{1} "^{\prime} \ldots \ldots\right\} \ldots\) are all unity: so that
\[
(x / h)^{n} x / h a b c e . . \infty x / a h b \cdot x / h / h \cdot x / c h \ldots
\]

Secondly, the factor \((x / h)^{n}\) has been omitted, so that
\[
x / h+1 b r \ldots \propto x / a h . a x / b h \cdot x / c h \ldots
\]

It is this second incorrect statement of the formula which leads to the fallacious rule for the comimation of the testimonies of independent witnesses ordinarily given in the text-imoks. \({ }^{1}\)
(46.1). If \(a b c^{\prime} \ldots / a h=\left\{a^{x h} b^{x h} c \ldots\right\} a / x h . b / x^{x h} . c / a h \ldots\)


\footnotetext{
1 seep p liso butuw.
}

This result is axemedinulv intorestimes. Mr. Johnson is the tiest to
 and the invirse formule: viz. that the same coefficient is required for correcting the simple formulæ of multiplication in both cases. As he remarbs, however, while the direct formula gives the required probability directly by multiplication, the inverse formula gives only the comparative probability.
(46.2) If \(x, x^{\prime}, x^{\prime \prime} \ldots\) are exclusive and exhaustive alternatives,

whence the result, since etc.
(47.1) The above formula may be written in the condensed form


This follows at once from (46.2), since \(x\) and \(\bar{x}\) are exclusive and exhaustive alternatives. (It is assumed that sh, \(\bar{x} h\), and ah, etc., are not inconsistont.)

This formula gives \(x / h a b c \ldots\) in terms of \(x / a h, x / b h\), ete., together with the three values \(x / h,\left\{\mu^{\prime \prime t} \mid,^{\prime \prime h}, r^{\prime \prime \prime} \ldots\right\}\), and

\[
\begin{align*}
& \bar{x} / h a \bar{b} c d \ldots \bar{x} / h b e \bar{d} \ldots=\left\{r^{\bar{z} h} h c d \ldots\right\} \cdot \bar{x} / a h{ }^{\prime} \bar{x} / h{ }^{\prime} \tag{48.1}
\end{align*}
\]

This gives the effect on the odds (prob. \(x:\) prob. \(\bar{x}\) ) of the extra knowledge \(\boldsymbol{a}\).
(4.9) When several data co-operate as evidence in favour of a proposition, they continually strengthen their own mutual probabilities, on the assumption that when the proposition is known to be true or to be false the data jointly are not counterdependent.
I.e. if \(\left\{a^{x h} b^{r h} c \ldots\right\}\) and \(\left\{a^{z h} b^{z h} c \ldots\right\}\) are not less than unity, and \(x / k h>x / h\) where \(k\) is any of the data \(a, b, c \ldots\), then \(\left\{a^{\prime \prime} b^{\prime \prime} c^{\prime \prime} d \ldots\right\}\) beginning with unity, continually increases, as the number of its terms is increased.
\[
\begin{aligned}
& a b c \ldots / h=x a b c \ldots / h+\bar{x} a b c \ldots / h \\
& \text { by (24.2). } \\
& =x / h \cdot a b c \ldots \mid x h+\bar{x} / h \cdot a b c \ldots / \bar{x} h .
\end{aligned}
\]
\[
\begin{aligned}
& \text { (since }\left\{a^{3 h} b^{\text {xh }} c \ldots\right\} \text { and }\left\{a^{3 h} b^{\text {h }} c \ldots\right\} \text { are not less than unity), }
\end{aligned}
\]
\[
\begin{aligned}
& 11\left|\begin{array}{cl}
\bar{x} / \text { uh } \\
\bar{x} / k & \bar{x}_{1}^{\prime}, \bar{x}_{1}, \cdots
\end{array}\right|
\end{aligned}
\]

We can show that each additional piece of evidence \(a, b, c \ldots\) increases the value of this expression. For let \(x / h \cdot G+\bar{x} / h \cdot G^{\prime}\) be its value when all the evidence up to \(k\) exclusive is taken, so that
\[
x / \operatorname{lih} \cdot\left(\hat{i}+\bar{x} / l i h . \hat{C}^{\prime}\right.
\]
is its value when \(k\) is taken. Now \(\mathrm{G}>\mathrm{G}^{\prime}\) since \(x_{1} / a / h>x / h\), etc., and \(\bar{x} /\) ahh \(-\bar{x} / / /\). etc.. by the hypothesis that the evidener favours \(x\); and for the same reason \(x / k h-x / h\), which is equal to \(\bar{x} / h-\bar{x} / k h\), is positive.
whence the result.
(19.1) The above proposition can be generalised for the case of exclusive alternatives \(x, x^{\prime}, r^{\prime \prime} \ldots\) (in place of \(\left.r, \bar{x}\right)\).

F,

from which it follons that. if !atha... Ime. ; 1, and it
 ( \({ }^{\prime}\) ) \(\ldots\) ) is increasing (with the number of letters) from unity.

Mr. Johnson describes this result as a generalisation of the corrected " middle term fallacy" (see Chap. V. § 4).

\section*{APPENDIX}

ON SYMBOLIC TREATMENTS OF PROBABILITY
The use of the symbol 0 for impossibility and 1 for certainty was first introduced hy leibnitz in a very early pamphlet, entitled

 Leibuitz, 1, 553). Leibnitz represented intermediate degrees of prohalility by the sign \(\frac{1}{2}\), meaning, however, by this symbol a membl. butwon 11 and 1

Several modern writers have made some attempt at a symbolic treatment of Probability. But with the exception of Boole, whose methods I have discussed in detail in Chapters XV., XVI., and XVII., no one has worked out anything very elaborate.

Mr. Mecoll published a number of brief notes on Probability of considerable interest-see esperially his symbolic Logic, Sixth Paper
 of a S'ymbolical Language. The conception of probability as a relation between propesitions underlies his symbolism, as it does mine. \({ }^{1}\) The probability of \(a\), relative to the \(\dot{a}\) priori premiss \(h\), he writes "; and the probability, given \(b\) in addition to the a priori premiss, he writes
 in the probability of a brought about by the addition of \(b\) to the evidene, he calls " the dependence of the statement \(a\) upon the state-

\footnotetext{
\({ }^{1}\) I did not come: acrosu these notes until my own method was considerably
 Probinbility.
}
 ology, \(b\) is irrelevant to \(a\) on evidence \(h\). The multiplication and addition formulæ he gives as follows: \(\begin{array}{cccc}a b & a & b \\ \epsilon & b & b & = \\ \epsilon\end{array}{ }^{\prime \prime}\)

Also
\[
\begin{aligned}
& \begin{array}{cccc}
a+b & a & b & a b \\
\epsilon & = & & \\
\epsilon & \epsilon
\end{array} . \\
& \delta_{b}^{a}=A_{B}^{A}{ }^{i}{ }^{b} \text {, whene } A={ }^{a} \text {. }
\end{aligned}
\]

It is surprising how little use he succeeds in making of those grood results. He arrives, however, at the inverse formmla in the shaje--
\[
\begin{aligned}
& c_{i} \text {. } \\
& c_{r} \quad \in c_{\text {, }} \\
& v=\Sigma_{1} c_{r} \quad \epsilon \quad{ }_{c}
\end{aligned}
\]
where \(c_{1} \ldots c_{n}\) are a series of mutually exclusive causes of the event \(v\) and include all possible causes of it ; reaching it as a sencralisation of the proposition

In a paper entitled "Operations in Relative Number with Applications to the Theory of Probabilities," \({ }^{1}\) Mr. B. I. Gimman attempter a symbolic treatment based on a frequence theny simila (o) Vemos. but made more precise and more consistent with itself: " Prohmability has to do, not with individual events, hut with classes of events ; and not with one class, but with a pair of classes, -the one containing, the other contained. The latter being the one with which we are principally concerned, we speak. by an ellipsis, of its probatibility: without mentioning the containing class: but in veality probahilityis a ratio, and to define it we must have both comelates given." Bint Mr. Gilman's symbolic treatment leads to very little. Howe rewently R. Laemmel. in his Untersuchurgen äber die Eirmithnug ro, Hidirscheinlichkeiten, made a beriming on somewhat similar lines; but in his case also the symbolic treatment lears to no substantin tresults.

Apart from the writers mentioned above, there are a ferv who have incidentally made use of a probability symbol. It will be sufficient to cite Czuber. \({ }^{2}\) He denotes the probability of an event

\footnotetext{
\({ }^{1}\) Published in the volume of Johns Hopkins Studies in Loric.
\({ }^{2}\) Wahrscheinlichkeitsrechnung, vol. i. pp. 43-48.
}

E by W(E), and the probability of the event E given the occurrence of an event F he \(\mathrm{W}_{1}(\mathrm{E})\). He uses this symbol to give \(\mathrm{W}_{1}(\mathrm{E})=\mathrm{W}_{\overline{\mathrm{F}}}(\mathrm{E})\) as the criterion of the independence of the events E and F ( F denoting the non-occurrem of \(F) ; W(E)=1\), as the expression of the fact that E is a necessary consequence of F ; and one or two other similar ncille:

Finally there is in the Bulletin of the Physico-mathematical Society of Kazan for 1857 a memoir in Russian by Platon S. Porvtzki entitled "A solution of the General Problem of the Theory of Probability by Means of Mathematical Logic." I have seen it stated that Schroder intended to publish ultimately a symbolic treatment of Probability. Whether he had prepared any manuseript on the subject before his death I do not know.

\section*{CHAPTER XV}

\section*{NUMERICAL MEASUREMENT AND APPROXIMATION OF PROBABILITIES}
1. The possibility of numerical measurement, mentioned at the close of Chapter III., arises out of the Addition Theorem (24.1). In introducing the definitions and the axiom, which are required in order to make the convention of numerical measurement operative, we may appear, as in the case of the orisinal definitions of Addition and Multiplication, to be arguing in an artificial way. This appearance is due, here as in Chapter XII., to our having given the names of addition and multiplication to certain processes of compounding probabilities in advance of postulating that the processes in question have the properties commonly associated with these names. As common sense is hasty to impute the properties as soon as it hears the names, it may overlook the necessity of formally introducing them.
2. The definitions and the axiom which are needed in order to give a meaning to numerical measurement are the following :--.
XVII. \(a / h+\{a / h+[a / h+(a / h+\ldots r\) terms \()]\}=r \cdot a / h\). Def.
 XIX. If \(b / f=q \cdot c / g\), then \(\frac{1}{r} \cdot b / f={ }_{,}^{q} c / g . \quad\) Def.

Thus if \(b / h=a / h+a / h+\ldots\) to \(r\) terms, then the probability \(b / h\) is said to be \(r\) times the probability \(a / h\); hence if \(a b / h=0\) and \(a / h=b / h\), the probability \((a+b) / h\) is twice the probability \(a / h\). If \(a\) and \(b\) are exhaustive as well as exclusive alternatives relatively to \(h\). so that \((a+b) / h=1\), since we take the relation of certainty as our unit, then \(a / h=b / h=\frac{1}{2}\).

We also need the following axiom postulating the existence of relations of probability corresponding to all proper fractions:
(vii.) If \(q\) and \(r\) are any finite integers and \(q<r\), there exists as relation of probability which con tee expressed. by means of the convention of the foregroing definitions, as \({ }_{i}{ }_{i}\).
3. From these axioms and definitions combined with those of (Whatwr XIl, it is cany the stow (owrtaint! himes mpresented by unity and impossibility by zero) that we can manipulate according to the ordinary laws of arithmetic the " numbers" which be matan of a sperial moventon we have thms introduced? to represent probabilities. Of the kind of proofs necessary for the complete demonstration of this the following is given as an example:

Let the probability \({ }_{\text {/In }}=\mathrm{P}\), which exists by (vii.),
thion
\[
\begin{aligned}
& \text { (...) }{ }_{\text {/ }}^{1} \text { i i. }
\end{aligned}
\]
\[
\begin{aligned}
\therefore \text { alf }+l / h= & n . \mathrm{P}+\prime \prime . \mathrm{P}, \text { if this probability exists, } \\
& \mathrm{P}+\mathrm{P} \ldots \text { to } n \text { terms }+\mathrm{P}+\mathrm{P} \ldots \text { to } m \text { terms, } \\
& \mathrm{P}+\mathrm{P} \ldots \text { to } m+n \text { terms } .
\end{aligned}
\]
(ㅎ․ - n) P
hy (XIX.

This probalility exists in virtue of (vii.).
4. Many probabilities-in fact all those which are equal to the probability of some other argument which has the same premiss and of which the conclusion is incompatible with that of the original argument-are numerically measurable in the sonse that there is some wther pobkabley whth whel they are comparathe in the mather deseritnad atocio. But the? ere mot nume firally measurable in the mest usual semse, untus 11 . pro batility with which they are thus compromble is the rolatom of cortaints: The conditions undio which a probability at is numericably measurable and equal to the aresily seen. It
is necessary that there should exist probabilities \(a_{1} / h_{1}, a_{2} / h_{2} \ldots\). \(a_{4} / h_{,}, \ldots u_{1} / h_{0}\), such that
\[
\begin{aligned}
& "_{1} / /_{1} \quad \quad_{2} / /_{2} \quad \cdots-\mu_{i / 1} / h_{q}=\ldots=\|_{!} / h_{1},
\end{aligned}
\]
\[
\begin{aligned}
& \text { If } a / h=\frac{q_{1}}{r_{1}} \text { and } b / h \quad \frac{q_{2}}{r_{2}} \text {, it follows from (32) that } a b / h=\frac{q_{1} q_{2}}{r_{1} r_{2}}
\end{aligned}
\]
only if \(u_{j} / h\) and \(b / h\) are independent arguments. Unless, therefore, we are dealing with independent arguments, we camot apply detailed mathematical reasoning even when the individual probabilities are numerically measurable. The greater part of mathematical probability, therefore, is concerned with arguments which are both independent and numerically measurable.
5. It is evident that the cases in which exact numerical measurement is possible are a very limited class, generally dependent oil evidence which warrants a judgment of equiprobability by an application of the Irinciple of Indifference. The fuller the evidence upon which we rely, the less likely is it to be perfectly symmetrical in its bearing on the various alternatives, and the more likely is it to contain some piece of relevant information favouring one of them. In actual reasoning, therefore, perfectly equal probabilities, and hence exact numerical measures, will occur comparatively seldom.

The sphere of inexact numerical comparison is not, however, quite so limited. Many probabilities, which are incapable of numerical measurement, can be placed nevertheless between numerical limits. And by taking particular non-numerical probabilities as standards a great number of comparisons or approximate measurements become possible. If we can place a probability in an order of magnitude with some standard probability, we can obtain its approximate measure by comparison.

This methor is frequently adopted in common discourse. When we ask how probable something is, we often put our question in the form-Is it more or less probable than so and so ?where 'so and so' is some comparable and better known probability. We may thus obtain information in cases where it would be impossible to ascribe any number to the probability in question. Darwin was giving a numerical limit to a non-numerical prob-
ability when he said of a conversation with Lyell that he thousht it no more likely that he should be right in nearly all points than that he should toss up a penny and wet heads twenty times running. \({ }^{1}\) Similar cases and others aks. Where the probability which is taken as the standard of emmparison is itwelf nonnumerical and not, as in Darwin's instance, a numerical one, will readily occur to the reader.

A spmecially important cast of approximate comparison is that of 'practical certainty.' This differs from logical errtainty since its contratictory is not impossible, but we are in practice completely satisfied with any probability which approaches such a limit. The phrase has naturally mot been used with complete precision ; but in its most useful sense it is essentially nonnumerical wie camot mesare practical certainty in troms of losical certainty: We can only explain how ereat practical certainty is by givine instances. We may say, for instance, that it is measured hy the probability of the sumis risine to-morrow. The type which we shall be most likele to take will be that of a well-verified induction.
6. Most of such comparisons must be based on the principles of Chapter V. It is possible, however, to develop a systematic method of approximation which may be oceasionally useful. The theorems given below are chiefly suggested by some work of Booles. His theorems were introduced fur a diflerent pur pose, and he does not seem to have realised this interesting application of them; but analytically his problem is idmotical with that of approximationi. This methen of approximation is alson substantially the same amalytically as that dealt with l, Mr. Yule under the heading of Consistence. \({ }^{3}\)

\footnotetext{
\({ }^{1}\) Life and Letters, vol. ii. p. 240.
\({ }^{2}\) In Boole's C'alculus we are apt to be left with an equation of the second or of an even higher degree from which to derive the probability of the conclusion; and Boole introduced these methods in order to determine which of the several roots of his equation should be taken as giving the true solution of the problem in probability. In each cuse he shows that that root must be chosen which lies betwern certain limits, and that only one rout satisties this condition. The general theory to bo applied in such cases is expounded by him in Chapter XIX. of The Laus of Thought, which is contitled "On Statistical Conditions." But the solution given in that chapter is awkward and unsatisfactory, and he subsequently published a much better method in the Philosophical Magazine for 1854 (4th serics, vol. viii.) under the title "On the Conditions hy which the Solutions of Questions in the Theory of Probabilities are limited."
- Theory of stativtics, chap. ii.
}
(51) \(x y / h\) always lies between \({ }^{1} x / h\) and \(x / h+y / h-1\) and between \(y / h\) and \(x / h+y / h-1\).
For
\[
\begin{array}{rlrl}
x y / h & =x / h-x \bar{y} / h & \text { by }(24.2), \\
& =x / h-\bar{y} / h \cdot x / \bar{y} h & & \text { by X. }
\end{array}
\]

Now
\[
x / \bar{y} h \text { lies between } 0 \text { and } 1 \text { by (2) and (3), }
\]
\(\therefore x y / h\) lies between \(x / h\) and \(x / h-\bar{y} / h\),
i.e. between \(x / h\) and \(x / h+y / h-1\).

As \(x y / h \nleftarrow 0\), the above limits may be replaced by \(x / h\) and 0 , if \(x / h+y / h-1<0\).

We thus have limits for \(x y / h\), close enough sometimes to be useful, which are available whether or not \(x / h\) and \(y_{i}^{\prime h}\) are independent arguments. For instance, if \(y / h\) is nearly certain, \(x y / h\) \(=x / h\) nearly, quite independently of whether or not \(x\) and \(y\) are independent. This is obvious ; but it is useful to have a simple and general formula for all such cases.
(52) \(x_{1} x_{2} \ldots x_{n}\), \(/ h\) is always greater than \(\leq x_{1} / h-n\).

For by (51) \(._{1} r_{2} \ldots e_{n+1} / h>r_{1} r^{\prime} r_{2} \ldots x_{n} / h+x_{n} / h-1\)
\[
>_{1} r_{1} n_{2} \ldots m_{n-1} / h+x_{n} / h_{1}+m_{n+1} \mid h-2 \text {, }
\]
and so on.
(53) \(x y!/ h+\bar{x} \bar{y} / h\) is always less than \(x / h-y / h+1\), and less than \(y / h-x / h+1\).
For as in (..1)
and
\[
r y / h=x / h-x \bar{y} / \|
\]
\[
\bar{x} \bar{y} / h=\bar{y} / h-\omega \bar{y} / h,
\]
\[
\therefore x y / h+\bar{x} \bar{y} / h=x / h-y / h+1-2 x \bar{y} / h,
\]
whence the required result.
(54) \(x y / h-\bar{x} \bar{y} / h=x / h+y / h-1\).

This proposition, which follows immediately from the above, is really out of place here. But its close connection with conclusions (51) and (53) is obvious. It is slightly unexpected, perhaps, that the difference of the probabilities that both oi two events will occur and that neither of them will, is independent of whether or not the events themselves are independent.
7. It is not worth while to work out more of these results here. Some less systematic approximations of the same kind are given in the course of the solutions in Chapter XVII.

In seeking to compare the degree of one probatility with that of another we may desire to get rid of one of the terms, on account

\footnotetext{
\({ }^{1}\) In this and the following theorems the term 'between' includes the limits.
}
of its not beinu comparable with any of our stamdard prol abibities． Thus our object in general is to eliminate a given symbel of quantity from a set of equations or ineyuations．If，for instance． we are to ohtain numerical limit－within which our probability must lie，we must eliminate from the result those probabilitio which are non－numerical．This is the general problem for solution．
（5i）A meneral method of solvine these problems when we can throw our equations into a linear shape so far as all symbols of probatility are concerned，is best shown in the following example：－
\begin{tabular}{|c|c|c|}
\hline cuppose we have & \(\lambda: 1 \cdot "\) & （i．） \\
\hline & 入．\()^{\text {i }}\) & （ii．） \\
\hline & \(\lambda: r: 9\), & （iii．） \\
\hline & \(\lambda: \mu-r-\mu-1\) & （iv．） \\
\hline & \(\lambda+\mu: \sigma+\tau\). & （i．） \\
\hline & ＋！\(+\sigma \cdot \tau: u 1\) & （ i i．） \\
\hline
\end{tabular}
where \(\lambda ., . r, \cdots, \quad, \quad-v\) represent probabilities which are to be eliminated，athl limits are to bo．found fore e in terms of the standard probabilities \(a, b, d, e\) and 1 ．
\(\lambda, \mu\) ，etc．，must all lie between 0 and 1 ．
From（i．）and（iii．）\(\sigma=c-a\) ；from（ii．）and（iii．）\(\nu=c-b\) ．
From（i．），（ii．）and（iii．）\(亠 \quad a-b \quad\) ．

substituting for \(\sigma . l^{\prime}, \lambda\) in（iv．），（v．），and（vi．）


We have still to，liminat，\(\mu \cdot \mu \cdot{ }^{\prime} /-\cdots, \mu-\)－
\[
\mu, .1 \ldots \quad ., \quad-1
\]

Hence we have ：

Lower limits of \(c:-n, l\)（whichever is greatest）．
This example，which is only slightly modition from one given hy Pombe，reprosents the andmal conditions of a well－known problem in probability．

\section*{CHAPTER XVI}

\section*{OBSERVATIONS ON THE THEOREMS OF CHAPTER XIV. AND THEIR DEVELOPMENTS, INCLUDING TESTIMONY}
1. In Definition XIII. of Chapter XII. a meaning was given to the statement that \(a_{1} / h\) and \(a_{2} / h\) are independent arguments. In Theorem (33) of Chapter XIV. it was shown that, if \(a_{1} / h\) and \(a_{2} / h\) are independent, \(a_{1} a_{2} / h=a_{1} / h \cdot a_{2} / h\). Thus where on given evidence there is independence between \(a_{1}\) and \(a_{2}\), the probability on this evidence of \(a_{1}{ }^{\prime \prime}\), jointly is the product of the probabilities of \(a_{1}\) and \(a_{2}\) separately. It is difficult to apply mathematical reasoning to the Calculus of Probabilities unless this condition is fulfilled; and the fulfilment of the condition has often been assumed too lightly. A good many of the most misleading fallacies in the theory of Probability have been due to a use of the Multiplication Theorem in its simplified form in cases where this is illegitimate.
2. These fallacies have been partly due to the absence of a clear understanding as to what is meant by Independence. Students of Probability have thought of the independence of events, rather than of the independence of arguments or propositions. The one phraseology is, perhaps, as legitimate as the other; but when we speak of the dependence of events, we are led to believe that the question is one of direct causal dependence, two events being dependent if the occurrence of one is a part cause or a possible part cause of the occurrence of the other. In this sense the result of tossing a coin is dependent on the existence of bias in the coin or in the method of tossing it, but it is independent of the actual results of other tosses; immunity from smallpox is dependent on vaccination, but is independent of statistical returns relating to immunity ; while the testimonies of two witnesses about the same occurrence are independent, so long as there is no collusion between them.

This semse, which it is mut casy to define quite precisely, is at any rate not the sense with which we are concerned when we deal with independent probabilities. We are concermed, not with direct causation of the kind described above, but with 'dependence for knowledure,' with the question whe ther the limencledye of one fact or event affords any rational ground for expecting the existence of the wher. The dependence for knowledge of two events usually arises. no doubt, out of causal comection, or what we term such, of some kind. But two events are not independent for knowledere merely because there is an absence of direct causal connection between them; nor, on the other hand, are they necessarily dependent because there is in fact a causal train which brings them into an indiecet connection. The question is whether there is any know probable comnection, direct or indirect. I knowledge of the resulte of other tosings of a coin may he hardly. less relevant than a knowledge of the bias of the coin; for a knowledge of these result may be a gromul for a probable knowledge of the bias. There is a similar connection between the tatistice of immmity from smallpos and the causal relations betwern varcination and smallpox. The truthful testimonies of two witnesses about the same occurrence have a common cause, namely the occurrence howerer imbemant (in the legal sense of the abseme of (e)llusion) the witneses mate ber For the purposes of probability two facts are omly independent if the existence of one is now imlication of ansthing which might be a part cause of the other.
3. While dependence and independence may be thus connected with the conception of causality, it is not convenient to found our definition of indepmdenee upen this commetion. I partial or pos-ihle canse involver idea- which are -till obecure, and I have prefereed to define independence he referenten to the com.
 there really are material extornal cansal laws, how far causal connection is distinct from logical connection, and other such questionk are profomadly a-amiated with the uhtmate prohlems of logic and protrability and with mans of the topics. .-peredlls. those of Part III., of this treatise. But I have nothing useful to say about them. Nearly everything with which I deal can he. expressed in terms of logical relevance. And the relations between logical relevance and material cause muat be left domboful.
4. It will be useful to give a fer examples out of writers who, as I conceive, have been led into mistakes through misapprehending the significance of Independence.

Cournot, \({ }^{1}\) in his work on Probability, which after a long period of neglect has come into high favour with a modern school of thought in France, distinguishes between ' subjective probability " based on ignorance and 'objective probability' based on the calculation of 'objective possibilities,' an 'objective possibility ' being a chance event brought about by the combination or convergence of phenomena belonging to independent series. The existence of objectively chance events depends on his doctrine that, as there are series of phenomena causally dependent, so there are others between the causal developments of which there is independence. These objective possibilities of Cournot's, whether they be real or fantastic, can have, however, small importance for the theory of probability. For it is not known to us what series of phenomena are thus independent. If we had to wait until we knew phenomena to be independent in this sense before we could use the simplified multiplication theorem, most mathematical applications of probability would remain hypothetical.
5. Cournot's 'objective probability,' depending wholly on objective fact, bears some resemblances to the conception in the minds of those who adopt the frequency theory of probability. The proper definition of independence on this theory has been given most clearly by Mr. Yule \({ }^{2}\) as follows :
" Two attributes A and B are usually defined to be independent, within any given field of observation or 'universe,' when the chance of finding them together is the product of the chances of finding either of them separately. The physical meaning of the definition seems rather clearer in a different form of statement, viz. if we define \(A\) and \(B\) to be independent when the proportion of A's amongst the B's of the given universe is the same as in that unicerse at large. If, for instance, the question were put, ' What is the test for independence of smallpox attack and vaccination?' the natural reply would be, 'The percentage of vaccinated amongst the attacked should be the same as in the general population.' . . ."
\({ }^{1}\) For some account of Cournot, see Chapter XXIV. § 3.
2 "Notes on the Theory of Association of Attributes in Statistics," Biometrika, vol. ii. p. 125.

This definition is consistent with the rest of the theory to which it belongs, but is, at the same time, open to the general objections to it. \({ }^{1} \mathrm{Mr}\). Yule admits that A and B may be indeperndent in the world at large but not in the world of ("s. The question therefore arises as to what world given evidence specifies, and whether any step forward is possible when. as is generally the case, we do not know for certain what the propertions in a given world actually are. As in the case of Cournot's independent suries it is in ereneral impessible that we should know whe ther A and B are or are not independent in this sense. The lowical indepmance for knowlodge which justifies our reasoning in a certain way must be something different from either of these objective forms of independence.
6. I come now to Boole's treatment of this subject. The contral urror in hiss system of probability arises out of his giving iwn ine onsistent definitions of 'independmene.' 2 He first wins the reader's accquiescence by giving a profectly correct definition: "Two events are said to be independent when the probability of the hapmening of either of them is unaffected by our ceperctution of the neequrence or failure of the other." \({ }^{3}\) But a mement later he interprets the term in quite a different sense ; for, according to Boole's second definition. we must regard the events as independent unless wer are told either that they must concur or that they comot concur. That is to says they are independent unless we know for certain that there is, in fact, an invariable conmection between them. "The simple events, \(x, y, z\), will be said to be conditimad when they are not free to oceur in wery possible combination ; in other words, when some com-

\({ }^{1}\) See (hapter \III.
\({ }_{2}^{2}\) Boole's mistake was pointed out, accurately though somewhat obscurely by H. Willoraham in his review "On the Theory of (hances deweloped in Professor
 failed to understand the point of Wilbraham's criticism, and replied hotly, challenzing him to impugn any individual results ("Reply to some Observations
 returned to the same question in a paper entitled "On a (ieneral Methed in the Theory of Probabilities," Phil. Mag. 4th series, vol. viii., 1854, where he endeavours to support his theory by an appeal to the I'rinciple of Indifference. McColl, in his "sixth Paper on Calculus of Equivalent Statements," raw that Boole's fallacy turned on his definition of Independence; but 1 do not think he understood, at least he does not explain, where precisely Boole's mistake lay
\({ }^{3}\) Laws of Thought, p. 255. The italics in this quotation nre mine.
. . Simple unconditioned events are by definition independent. " \({ }^{1}\) In fact as long as \(x z\) is possible, \(x\) and \(z\) are independent. This is plainly inconsistent with Boole's first definition, with which he makes no attempt to reconcile it. The consequences of his employing the term independence in a double sense are far-reaching. For he uses a method of reduction which is only valid when the arguments to which it is applied are independent in the first sense, and assumes that it is valid if they are independent in the second sense. While his theorems are true if all the propositions or events involved are independent in the first sense, they are not true, as he supposes them to be, if the events are independent only in the second sense. In some cases this mistake involves him in results so paradoxical that they might have led him to detect his fundamental error. \({ }^{2}\) Boole was almost certainly led into this error through supposing that the data of a problem can be of the form, "Prob. \(x=p\)," i.e. that it is sufficient to state that the probability of a proposition is such and such, without stating to what premisses this probability is referred. \({ }^{3}\)

It is interesting that De Morgan should have given, incidentally, a definition of independence almost identical with Boole's second definition: "Two events are independent if the latter might have existed without the former, or the

\section*{\({ }^{1}\) Op. cit. p. 258.}
\({ }^{2}\) There is an excellent instance of this, Laws of Thought, p. 286. Boole discusses the problem : Given the probability \(p\) of the disjunction 'either Y is true, or X and Y are false,' required the probability of the conditional proposition, 'If X is true, Y is true.' The two propositions are formally equivalent; but Boole, through the error pointed out above, arrives at the result \(\frac{c p}{1-p+c p}\), where \(c\) is the probability of 'If either Y is true, or X and Y false, X is true.' His explanation of the paradox amounts to an assertion that, so long as two propositions, which are formally equivalent when true, are only probable, they are not necessarily equivalent.
\({ }^{3}\) In studying and criticising Boole's work on Probability, it is very important to take into account the various articles which he contributed to the Philosophical Magazine during 1854, in which the methods of The Laws of Thought are considerably improved and modified. His last and most considered contribution to Probability is his paper "On the application of the Theory of Probabilities to the question of the combination of testimonies or judgments," to be found in the Edin. Phil. Trans. vol. xxi., 1857. This memoir contains a simplification and general summary of the method originally proposed in The Laws of Thought, and should be regarded as superseding the exposition of that book. In spite of the error already alluded to, which vitiates many of his conclusions, the memoir is as full as are his other writings of genius and originality.
former without the latter. for anvthing that we know to the contrars." \({ }^{1}\)
7. In many other cases urmers have arisen, not through a misapprehension of the meaning of independence, but merely through careless assumptions of it, of through enunciating the Theorem of Multiplication without its qualifying comdition. Mathematicians have been tow vamer to assume the lewitimary of those complicated processes of multiplying probrabilities. for which the greater part of the mathematies of probability is engaged in supplying simplifications and approximate solutions. Even 1) Morgan was careless cmough in one of his writings \({ }^{2}\) to rnmeliate the Multiplication Theorem in the following form: * The prohability of the happening of two, there or mome ewents is the product of the probabilitios of the ir hapmenine separately (p. 398). . . . Knowing the probability of a compound event, and that of one of its components, we find the probability of the other by dividing the first by the second. This is a mathematical result of the last ten ohvions to require further proof (p. 401).

An excellent and classic instance of the danger of wrongful assumptions of independence is given by the problem of determining the probability of throwing heads twice in \(t\) wo consecutive tosses of a coin. The plain man generally assumes without hesitation that the chance is \(\left(\frac{1}{2}\right)^{2}\). For the it mioni chance of heads at the firat tosis is \(\frac{1}{2}\). and we might maturally suppese that the two exents are indernudent, shluce the mere fact of heads having appeared once can have no influence on the next toss. But this is not the case muless we know foe certain that the coin is free from bias. If we do not know whether there is bias, or which way the hias lies, then it is remomathe top pht the probability. somewhat higher than \(\left(\frac{1}{2}\right)^{2}\). The fuct of heads havene appearel at the firat toms is mot the cause of heads appearine at the stomet alse. but the limertolyge, that the woin has fallen heads already, affeets cur foremant of its falling thas in the future, sime heads in the past mat have been due to a camse which will favour head a in the future. The pmsithility of hias in a coin, it may loe moticent.
 is not very consistent with himself in his various distinct treatises on this subject, and other definitions may be found elsewhere. Boole's second defini. tion of Independence is alao adopeded by Macfarlane, Algebra of Lergic, p. 2l

always favours ' runs ' ; this possibility increases the probability both of 'runs' of heads and of 'runs' of tails.

This point is discussed at some length in Chapter XXIX. and further examples will be given there. In this chapter, therefore, I will do more than refer to an investigation by Laplace and to one real and one supposed fallacy of Independence of a type with which we shall not be concerned in Chapter XXIX.
8. Laplace, in so far as he took account at all of the considerations explained in \(\S 7\), discussed them under the heading of Des inégalités inconnues qui peuvent exister entre les chances que l'on suppose égales. \({ }^{1}\) In the case, that is to say, of the coin with unknown bias, he held that the true probability of heads even at the first toss differed from \(\frac{1}{2}\) by an amount unknown. But this is not the correct way of looking at the matter. In the supposed circumstances the initial chances for heads and tails respectively at the first toss really are equal. What is not true is that the initial probability of 'heads twice' is equal to the probability of 'heads once' squared.

Let us write 'heads at first toss' \(=h_{1}\); 'heads at second toss' \(==h_{22}\). Then \(h_{1} / h=h_{2} / h=\frac{1}{2}\), and \(h_{1} h_{2} / h=h_{2} / h_{1} h \cdot h_{1} / h\). Hence \(h_{1} h_{2} / h=\left\{h_{1} / h\right\}^{2}\) only if \(h_{2} / h_{1} h=h_{2} / h\), i.e. if the knowledge that heads has fallen at the first toss does not affect in the least the probability of its falling at the second. In general, it is true that \(h_{2} / h_{1} h\) will not differ greatly from \(h_{2} / h\) (for relative to most hypotheses heads at the first toss will not much influence our expectation of heads at the second), and \(\frac{1}{4}\) will, therefore, give a good approximation to the required probability: Laplace suggests an ingenious method by which the divergence may be diminished. If we throw two coins and define 'heads' at any toss as the face thrown by the second coin, he discusses the probability of 'heads twice running' with the first coin. The solution of this problem involves, of course, particular assumptions, but they are of a kind more likely to be realised in practice than the complete absence of bias. As Laplace does not state them, and as his proof is incomplete, it may be worth while to give a proof in detail.

Let \(h_{1}, t_{1}, h_{2}, t_{2}\) denote heads and tails respectively with the first and second coins respectively at the first toss, and \(h_{1}{ }^{\prime}, l_{1}{ }^{\prime}, h_{2}{ }^{\prime}, l_{2}{ }^{\prime}\) the corresponding events at the second toss, then
\({ }^{1}\) E'ssai philosophique, p. 49. See also "Jémoire sur les Prohabilités," Mém. de l'Acad. p. 228, and cp. D'Alembert, "Sur le calcul des probabilités," opuscules mathématiques (1781), vol. vii.
the probability (with the aimen convention) of 'hemls twiee rumning,' i.e. agreement betwees the two cons twiee rumines is

\[
\left(i_{1} 11_{1}\right.
\]
 of Indifference, and \(h_{2} 2_{2} 2_{2} \quad 1\).

Similarly
\[
\left(h_{1}^{\prime \prime} 1_{1}^{\prime} \cdot t_{1}^{\prime} 1_{1}^{\prime}\right)^{\prime} h_{1}=2 h_{1}{ }^{\prime \prime} 1_{1}^{\prime} h .
\]

We may assume that \(h_{1} / h_{1}{ }^{\prime} h=h_{1} / h\), i.e. that heads with one coin is irrele vant to the probability of heads with the other: and \(h_{1} / h=h_{1}^{\prime} / h=\frac{1}{2}\) by the Principle of Indifference, so that
since. \(\left(h_{1} h_{1}^{\prime} \cdots t_{1}^{\prime} 1^{\prime}\right)\) beime irrelevant to \(h_{2}^{\prime} / h_{1}, h_{2}^{\prime}\left(h_{1} h_{1}^{\prime}+t_{1} 1_{1}^{\prime}, h_{1}\right)\) \(h_{2}^{\prime} / h_{2}^{1} \cdot\)

Now \(h_{2}\left(h_{2}^{\prime}, h_{1} h_{1}^{\prime}{ }^{\prime}-I_{1}^{\prime} 1^{\prime}=h_{1}\right)\) is wreater tham \(\frac{1}{2}\), since the fact of
 they will agree again. But it is less than \(h_{2} / h_{1} h\) : for we may assume that \(h_{1},\left(h_{2}^{\prime}, h_{1} h_{1}^{\prime} \cdot t_{1}^{\prime} \prime_{1}^{\prime}, h\right)\) is lass thath \(h_{2}^{\prime}\left(h_{2}^{\prime}, h_{1}^{\prime} \prime_{1}^{\prime}, h_{1}\right)\). and also that \(h_{2} /\left(h_{2}^{\prime}, h_{1} h_{1}^{\prime}, h_{1}\right)=h_{2} h_{1} h\), i.e. that honle twi... runniner with one coin does not increase ther prohathility of heads \(t\) wice rumning with a different eonin. Latplatees methanh of tossines. therefore yields with these assumptions, more or lase lemitimate accordines \(t 0\) the content of \(h\), a peobability nearee to \(\frac{1}{4}\) than is
 exatctly \(\frac{1}{1}\).
9. Two other examples will complete this rather discursion commentary. It has been supposed that by the Principle of Indifference the pron abibity of the existome of iron upon sirius is \(\frac{1}{2}\), and that smilarly the probability of the wistenm there of any other element is also \(\frac{1}{2}\). The probability, therefore, that not one of the 68 terrestrial elements will be found on Nirius is ( \(\left.\frac{1}{2}\right)^{\circ}\). and that at least onn will bee fonmel theme is I (2 \()^{\prime}\) wr approximately certain. This argument, or a similar one, has

many other things, that at least one college exactly resembling some college at either Oxford or C'ambridge will ahmost certainly be found on Sirius. The fallacy is partly due, as has been pointed out by Von Kries and others, to an illegitimate use of the Principle of Indifference. The probability of iron on Sirius is not \(\frac{1}{2}\). But the result is also due to the fallacy of false independence. It is assumed that the known existence of 67 terrestrial elements on Sirius would not increase the probability of the sixty-eighth's being found there also, and that their known absence would not decrease the sixty-eighth's probability. \({ }^{1}\)
10. The other example is that of Maxwell's classic mistake in the theory of gases. \({ }^{2}\) According to this theory molecules of gas move with great velocity in every direction. Both the directions and velocities are unknown, but the probability that a molecule has a given velocity is a function of that velocity and is independent of the direction. The maximum velocity and the mean velocity vary with the temperature. Maxwell seeks to determine, on these conditions alone, the probability that a molecule has a given velocity. His argument is as follows:

If \(\phi(x)\) represents the probability that the component of velocity parallel to the axis of X is \(x\), the probability that the velocity has components \(x, y, z\) parallel to the three axes is \(\phi(x) \phi(y) \phi(z)\). Thus if \(\mathrm{F}(r)\) represents the probability of a total velocity \(v\), we have \(\phi(x) \phi(y) \phi(z)=\mathrm{F}(r)\), where \(x^{2}=x^{2}+y^{2}+z^{2}\). It is not difficult to deduce from this (assuming that the

\footnotetext{
\({ }^{1}\) See Von Kriss, Jir Principien der Wharscheinlichleitsrechnung, p. 10. Stumpf (Über den Begriff der mathom. Wahwehoinlichlirit, pp. 71-74) argues that the fallacy results from not taking into account the fact that there might be as many metals as atomic weights, and that therefore the chance of iron is \(\frac{1}{z}\), where \(z\) is the number of possible atomic weights. A. Nitsche (Vierteljsch.f. wissensch. Philos., 1892) thinks that the real alternatives are 0 , or only 1 , or only \(2 \ldots\) or 68 terrestrial elements on Sirius, and that these are equally probable, the chance of each being \(\stackrel{1}{69}\)
\({ }^{2}\) I take the statement of this from Bertrand's Calcul des probabilités, p. 30. Let me here quote a precocious passage on Prohability regarded as a branch of Logic, from a letter written by Maxwell in his ninetcenth year (1850), before he came up to Cambridge: "They say that Understanding ought to work by the rules of right reason. These rules are, or ought to be, contained in Logic ; but the actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind " (Life, page 143).
}
functions are analytical) that \(\phi(.)^{\prime}\) must be of the form Ge

It is generally agreed at the present time that this result is erronerous. But the nature of the error is, I think. quite different from what it is commonly supposed to be.

Bertrand. \({ }^{1}\) Poincare. \({ }^{2}\) and Von Kries. \({ }^{3}\) all cite this arsument of Maxwell's as an illustration of the fallacy of Independence: and argue that \(\phi(r), \phi(y)\), and \(\phi(z)\) camot, as he assumes. represent indepembent prohatilities, if, as he also assumes, the probability of a velocity is a function of that velocity: But it is not in this way that the error in the result really arises. If we do not know what function of the velocity the probability of that velocity is, a knowled fe of the velocity parallel to the axes of \(s\) and \(y\) tells us nothing about the velocity parallel to the axis of \(z\). Maxwell was, I think, quite right to hold that a mere assumption that the prohability of a relocity is some function of that velocity, does not interfere with the mutual independenee of statements as to the velocity parallel to cach of the three ases.s. Let us denote the proposition. "the relecity parallel to the axis of X is \(x\) " by \(\mathbf{X}(x)\), the corresponding propesitions relative to the axes of Y and \(Z\) by \(Y^{\prime}(y)\) and \(Z(z)\), and the proposition 'the total velocity is \(v^{\prime}\) by \(\mathrm{V}(v)\); and let \(h\) represent our à priori data. Then if \(\left.\mathrm{X}(,)^{\prime}\right) h(x)\) it is a justifiable inference from the Principle of Indifference that \(V(y) h \quad \phi(y)\) and \(Z(z) h=\phi(z)\). Maxwell infors from this that \(X(s)\rangle(y) Z(z) / h \quad \phi(s) \phi(y) \phi(z)\). That is to say, he assumes that \(\mathrm{Y}(y) / \mathrm{X}(x) . h=\mathrm{Y}(y) / h\) and that \(Z(z) / Y(y) \cdot \mathbf{X}(x) \cdot h=Z(z) / h\). I do not agree with the authorities cited abow that this is illegitimate. So long as we do not know what function of the total velocity the prohability of that velocity is. a kowwedee of the velocities parallel to the axes of \(x\) and \(y\) has no bearing on the probability of a given velocity parallel to the axis of \(z\). But Maxwell groes on to infer that \(\mathrm{X}(r) Y^{\prime}(y) Z(z), h \quad V^{\prime}(r) h\), where \(r^{2}-r^{2}+y^{2}=z^{2}\). It is here. and in a very chementary way that the error crens in. The propositions \(X(x) Y(y) Z(z)\) and \(\backslash(r)\) are not equivalent. The latter follows from the former, hut the former does not follow from the latter. There is more than one set of values \(x, y, z\),

\footnotetext{
 - 'intion dra promathiliti (2bul al.), |p. 1111

}
which will yield the same value \(v\). Thus the probability \(V(v) / h\) is much greater than the probability \(\mathrm{X}(x) \mathrm{Y}(y) \mathrm{Z}(z) / h\). As we do not know the direction of the total velocity \(v\), there are many ways, not inconsistent with our data, of resolving it into components parallel to the axes. Indeed I think it is a legitimate extension of the preceding argument to put \(\mathrm{V}(v) / h=\phi(v)\); for there is no reason for thinking differently about the direction V from what we think about the direction X .

A difficulty analogous to this occurs in discussing the problem of the dispersion of bullets over a target-a subject round which, on account of a curiosity which it seems to have raised in the minds of many students of probability, a literature has grown up of a bulk disproportionate to its importance.
11. I now pass to the Principle of Inverse Probability, a theorem of great importance in the history of the subject. With various arguments which have been based upon it I shall deal in Chapter XXX. But it will be convenient to discuss here the history of the Principle itself and of attempts at proving it.

It first makes its appearance somewhat late in the history of the subject. Not until 1763, when Bayes's theorem was communicated to the Royal Society, \({ }^{1}\) was a rule for the determination of inverse probabilities explicitly enunciated. It is true that solutions to inductive problems requiring an implicit and more or less fallacious use of the inverse principle had already been propounded, notably by Daniel Bernoulli in his investigations into the statistical evidence in favour of inculation. \({ }^{2}\) But the appearance of Bayes's Memoir marks the beginning of a new stage of development. It was followed in 1767 by a contribution from Michell \({ }^{3}\) to the Plilosophical Transactions on the distribu-

\footnotetext{
\({ }^{1}\) Published in the Phil. Trans. vol. liii., 1763, pp. 376-398. This Memoir was communicated by Price after Bayes's death; there was a second Memoir in the following year (vol. liv. pp. 29S-310), to which Price himself made some contributions. See Todhunter's History, pp. 299 et seq. Thomas Bayes was a dissenting minister of Tunbridge Wells, who was a Fellow of the Royal Society from 1741 until his death in 1761. A German edition of his contributions to Probability has been edited by Timerding.

2 "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir," Hist. de l'Acad., Paris, 1760 (published 1766). Bernoulli argued that the recorded results of inoculation rendered it a probable cause of immunity. 'This is an inverse argument, though Bayes's theorem is not used in the course of it. See also D. Bernoulli's Memoir on the Inclinations of the Planetary Orbits.
\({ }^{3}\) Michell's argument owes more, perhaps, to Daniel Bernoulli than to Bayes.
}
tion of the stars, to which further reference will he made in Chapter XXI. And in 1774 the rule was clearly, theorth mot quite accurately, enunciated bey Laplace in his ․ Ménoire sur la probabilité des causes prar les éron-mens" (Mémuires
 the principle as follows (p. 623) :
"Si un évènement peut être produit par un nombre \(n\) de causes différentes, les frobabilités do lexistonce du (res causers prises de l'évènement sont mintre clles comme liss prohabilités de l'évènement prises do ces causes: it la prohabilité de loxist nuen de chacune d'elles sst égalu à la probabilité de l'évèmoment prise de cette cause, divisée par la somme do toutos liss prohabilités de l'évènement prises de chacune de ces causes."

He speaks as if he intended to prove this principle, hut he only give explanations and instances without proof. The principle is not strictly true in the form in which he enunciates it, as will low seen on reference to theorems (38) of (hapter NIV.; and thw omission of the necessary qualification has led to a number of fallacious arguments, some of which will be considered in Chapter XXX
12. The value and originality of Bayes's Memoir are considerable, and Laplace's method prohathly owes much more to it than is eremally recomised or than was acknowledged by Laplace. The principle, often called be Payes's name, dows not appear in his Memen in the shapee given it by Laplace and usually adopted since ; but Bayess emumetiatom is strictly correct and his mothod of arriving at it shows its true horical connection. with more fundamental principles, whereas Laplace's enunciation gives it the appearance of a new principle specially introduced for the solution of causal problems. The following passate \({ }^{1}\) gives, in my opinion, a rieht method of approachines the problem: "If there be two subsequent events, the probability. of the second \(\frac{{ }^{\prime}}{\mathrm{N}}\) and the probability of both tore ther \({ }^{\mathrm{P}} \mathrm{N}^{\prime}\) and. it being first discovered that the sorond went has happerned. from hence: I guess that the first mont has also happened, the prob)


\footnotetext{


}
is denoted by \(a\) and of the second by \(b\), this corresponds to \(a b / h=a / b h . b / h\) and therefore \(a / b h=\frac{a b / h}{b / h}\); for \(a b / h=\frac{\mathrm{P}}{\mathrm{N}}, b / h=\frac{\mathrm{b}}{\mathrm{N}}\), \(a / b h=\frac{\mathrm{P}}{b}\). The direct and indeed fundamental dependence of the inverse principle on the rule for compound probabilities was not. appreciated by Laplace.
13. A number of proofs of the theorem have been attempted since Laplace's time, but most of them are not very satisfactory, and are generally couched in such a form that they do no more than recommend the plausibility of their thesis. Mr. McColl \({ }^{1}\) gave a symbolic proof, closely resembling theorem (38) when differences of symbolism are allowed for : and a very similar proof has also been given by A. A. Markoff. \({ }^{2}\) I am not acquainted with any other rigorous discussion of it.

Von Kries \({ }^{3}\) presents the most interesting and careful example of a type of proof which has been put forward in one shape or another by a number of writers. We have initially. according to this view, a certain number of hypothetical possibilities, all equally probable, some favourable and some unfavourable to our conclusion. Experience, or rather knowledge that the event has happened, rules out a number of these alternatives, and we are left with a field of possibilities narrower than that with which we started. Only part of the original field or Spielraum of possibility is now admissible (zulässig). Causes have à posteriori probabilities which are proportional to the extent of their occurrence in the now restricted field of possibility.

There is much in this which seems to be true, but it hardly amounts to a proof. The whole discussion is in reality an appeal to intuition. For how do we know that the possibilities admissible à posteriori are still. as they were assumed to be \(\grave{a}\) priori, equal possibilities? Von Kries himself notices that there is a difficulty : and I do not see how he is to avoid it, except by the introduction of an axiom.

This was in fact the course taken by Professor Donkin in 1851, in an article which aroused some interest in the Philosophical

\footnotetext{
1 "Sixth Paper on the Calculus of Equivalent Statements," Proc. Lond. Math. Soc., 1897, vol. xxviii. p. 567. See also p. 155 above.

2 W'ahrscheinlichkeit.rechnung, p. 178.
\({ }^{3}\) Die Principien der Wahrscheinlichkeitsrechnuny, pp. 117-121. The above account of Von Kries's argument is much condensed.
}

Magazine at the time，but which has since been forentem． Donkin＇s theory is，however，of considerahlm intorest．He laid down as one of the fundamental principhes of probathlity the following：\({ }^{1}\)
－If there he any number of mutnally exelusive hypothess \(h_{1} h_{2} h_{3} \ldots\) of which the probabilities relative to a particular state of information are \(\mu_{1} \mu_{2} \mu_{3}\) ．．．and if new information be gained which chaneres the probatilities of some of thems suppose of \(h\) and all that follow，without havine otherwise any reference． to the rest．Then the probathitites of these latem habie the same ratios to one another，afler the new information，that they hat before．，＂ 2

Donkin wors on to say that the most impertant case is wher． the now information comsists in the knowledere that somm of the hypotheses must be rejected，without any further information as to these of the orisinal set which are retained．This is the proposition which Von Kries requires．

As it stands，the phrase＂without having otherwise any reference to the rest＂obvionsly lacks precision．In interperetat tion，however，can be put upon it，with which the principle is true．If，given the old information and the truth of one of the hypothesses \(h_{1} \ldots\) ．to the exclusion of the rest，the probability of what is comsiced bが the カール information is the stme whicherer of the hypotheses \(h_{1} \ldots h_{\text {a }}\) has been taken，then Donkin＇s principle is valid．For let a be the old information．\(a^{\prime}\) the new， and let．\(h^{\prime \prime} \mu, h / \operatorname{lan}^{\prime}=\rho^{\prime}:\) then
 explained．

14．Difficulties connected with the Incersis Principh have arisen，howewr，not son much in attwmpts to prove the principle as in these to emmente it theneth it may have been the law

\footnotetext{
 4th ．．．i．．vol．i．，Lail．
－It is interesting to notice that an axiom，practically equivalent to this， has been Inid down more lately by A．A．Markoff（W＇ahrscheinlichkeikerehnung． p．8）under the title • Cinabhangigkeitsaxiom．＇
}
of a rigorous proof that has been responsible for the frequent enunciation of an inaccurate principle.

It will be noticed that in the formula (38.2) the à priori probabilities of the hypotheses \(a_{1}\) and \(a_{2}\) drop out if \(p_{1}=p_{2}\), and the results can then be expressed in a much simpler shape. This is the shape in which the principle is enunciated by Laplace for the general case, \({ }^{1}\) and represents the uninstructed view expressed with great clearness by De Morgan: \({ }^{2}\) "Causes are likely or unlikely, just in the same proportion that it is likely or unlikely that observed events should follow from them. The most probable cause is that from which the observed event could most easily have arisen." If this were true the principle of luverse Probability would certainly be a most powerful weapon of proof, even equal, perhaps, to the heavy burdens which have been laid on it. But the proof given in Chapter XIV. makes plain the necessity in general of taking into account the à priori probabilities of the possible causes. Apart from formal proof this necessity commends itself to careful reflection. If a cause is very improbable in itself, the occurrence of an event, which might very easily follow from it, is not necessarily, so long as there are other possible causes, strong evidence in its favour. Amongst the many writers who, forgetting the theoretic qualification, have been led into actual error, are philosophers as diverse as Laplace, De Morgan, Jevons, and Sigwart, Jevons \({ }^{3}\) going so far as to maintain that the fallacious principle he enunciates is "that which common sense leads us to adopt almost instinctively."
15. The theory of the combination of premisses dealt with in \(\S \S 7,8\) of Chapter XIV. has not often been discussed, and the history of it is meagre. Archbishop Whately \({ }^{4}\) was led astray

\footnotetext{
\({ }^{1}\) See the passage quoted above, p. 175.
2 "Essay on Probabilities," in the Cabinet Encyclopardia, p. 27.
\({ }^{3}\) Principles of Sci nce, vol. i. p. 280.
\({ }^{4}\) Logic, 8th ed. p. 211: "As in the case of two probable premisses, the conclusion is not established except upon the supposition of their being both true, so in the case of two distinct and independent indications of the truth of some proposition, unless both of them fail, the proposition must be true : we therefore multiply together the fractions indicating the probability of the failure of each-the chances against it-and, the result being the total chances against the establishment of the conclusion by these arguments, this fraction being deducted from unity, the remainder gives the probability for it. E.g. a certain book is conjectured to be by such and such an author, partly, 1st, from its resemblance in style to his known works; partly, 2 nd , from its being attri-
}
by a superficial error, and 1) Nomgan, adopting the some miss taken rule, pushed it to the peint of ahsundity. \({ }^{1}\) Bishop Terren \({ }^{2}\) approached the question mone eritically. Boole's \({ }^{3}\) last and most considered contribution to ther subject of probability dealt with the same topic. I know of no discossom of it durine the past sixty years.

Boole's treatment is full and detailed. He states the problem as follows: "Required the probahility of an event z when two circumstances \(x\) and \(y\) are known to be present, the probability of the event \(z\), when we know only of the existence of the circumstances \(x\), being \(p\), and the probability, when we moly know of the existence of \(y\), being \(9 .{ }^{-4}\) His solution, howewer, is vitiated by the fundamental error examined in \(\$ 6\) ahove. Two of his conclusions may be mentioned for their plansibility, but neither is valid.
" If the causes in operation, or the testimonies borne," he
buted to him by some one likely to be pretty well informed. Let the probability of the conclusion, as deduced from one of these arguments by itself, be supposed ?, and in the other case z; then the opposite probabilities will be ? and t. which multiplied together give \(\frac{3}{2}\) as the probability against the conclusion. . . .

The Archbishop's error, in that a negative can always be turned into an affirmative by a change of verbal expression, was first pointed out by a mere diocesan, Bishop Terrot, in the Edin. Phil. T'rans, vol. xxi. The mistake is well explained by Boole in the same volume of the Edin. Phit. Trans. : "A confusion may here be noted between the probability that a conclusion is proved, and the probability in favour of a conclusion furnished by evidence which does not prove it. In the proof and statement of his rule, Archbishop Whately adopts the former view of the nature of the probabilities concerned in the data. In the exemplification of it, he adopts the latter."
 by means of it that "if any assertion appear neither likely nor unlikely in itself, then any logical argument in favour of it, however weak the premises, makes it in some degree more likely than not-a theorem which will be readily admitted on its own evidence." He then pives an example: " \(i\) priori vegetation on the planets is neither likely nor unlikely ; suppose argument from analogy makes it \(1_{1}^{3}\); then the total probability is \(\frac{1}{2} \frac{1}{1}\) or or \(\frac{1}{2}\) " \({ }^{\prime \prime}\) ) Morgan seems to accept without hesitation the conclusion to be derived from this, that everything which is not impossible is as probable as not.
\({ }^{8}\) "On the Possibility of combining two or more Probabilities of the same Event, so as to form one definite Prohability," Edin. Phil. Trane., 1856 , vol. xxi.
\({ }^{3}\) "On the Application of the Theory of Probabilities to the Question of the Combination of Testimonies or Judgments," Edin. Phil. T'rans., 1857, vol. xxi.
\({ }^{4}\) Loc. cit. p. 631. Boole's principle (loc. cit. p. 620) that " the mean strongth of any probabilities of an event which are founded upon different judgments or observations is to be measured by that supposed prolability of the event in priori which those judgments or observations following thereupen would not tend to alter," is not correct if it means more than that the mean strenifh of \(z / x\) and \(z / y\) is to be measured by \(z / x y\).
argues, ". are, separately, such as to leave the mind in a state of equipoise as respects the event whose probability is sought, united they will but produce the same effect." If, that is to say, \(a / h_{1}=\frac{1}{2}\) and \(a / h_{2}=\frac{1}{2}\), he concludes that \(a / h_{1} h_{2}=\frac{1}{2}\). The plausibility of this is superficial. Consider, for example, the following instance : \(h_{1}=\mathrm{A}\) is black and B is black or white, \(h_{2}=\mathrm{B}\) is black and A is black or white, \(a=\) both A and B are black. Boole also concluded without valid reason that \(a / h_{1} h_{2}\) increases, the greater the a priori improbability of the combination \(h_{1} h_{2}\).
16. The theory of "Testimony" itself, the theory, that is to say, of the combination of the evidence of witnesses, has occupied so considerable a space in the traditional treatment of Probability that it will be worth while to examine it briefly. It may, however, be safely said that the principal conclusions on the subject set out by Condorcet, Laplace, Poisson, Cournot, and Boole, are demonstrably false. The interest of the discussion is chiefly due to the memory of these distinguished failures.

It seems to have been generally believed by these and other logicians and mathematicians \({ }^{1}\) that the probability of two witnesses speaking the truth, who are independent in the sense that there is no collusion between them, is always the product of the probabilities that each of them separately will speak the truth. \({ }^{2}\) On this basis conclusions such as the following, for example, are arrived at:

X and Y are independent witnesses (i.e. there is no collusion between them). The probability that \(\mathbf{X}\) will speak the truth is \(x\), that Y will speak the truth is \(y\). X and Y agree in a particular statement. The chance that this statement is true is
\[
\begin{gathered}
x y \\
x y+(1-x)(1-y)
\end{gathered}
\]

For the chance that they both speak the truth is \(x y\), and the chance that they both speak falsely is \((1-x)(1-y)\). . \(s\) s, in this

\footnotetext{
\({ }_{1}\) Perhaps M. Bertrand should be resistered as an honourable exception. At least he points out a precisely analogous fallacy in an example where two meteorologists prophesy the weather, Calcul des Probabilités, p. 31.
\({ }^{2}\) E.g., Boole, Laws of Thought, p. 279.
De Morgan, Formal Logic, p. 195.
Condorcet, Essai, p. 4. Lacroix, Traité, p. 248. Cournot, Exposition, p. 354. Poisson, Recherches, p. 323.
Lais list could be greatly extended.
}
case, our hypothesis is that they auree these two altermatives are exhanstive: whence the abow result, which may low foumd in almost every discussion of the subject.

The fallacy of such reasoning is easily exposed by a more exact statement of the problem. For let \(a_{1}\) stand for " \(\mathrm{X}_{1}\) asserts \(a\)," and let \(a / n_{1} h=x_{1}\), where \(h\), our general data, is bv itself irrelevant to \(a, i . e ., x_{1}\) is the probability that a statement is true of which we only know that \(X_{1}\) has assorted it. Similarly lut. ms write \(b_{1} h_{2} h_{2}=x_{2}\), whare \(h_{2}\) stands for " \(X_{2}\) assiets \(h_{\text {, " }}\) The above argument then asmunes that, if \(\mathrm{X}_{1}\) and \(\mathrm{X}_{2}\) are witnesses who are cansally independment in the sense there is no collusion between them direct or indirect. ah \(h_{1} h_{2} h_{2}-a h_{1} h_{1}, l_{1} h_{2} h_{2} \quad a_{1} r_{2}\).

But ah, \(c_{1} h_{2} h \quad a /\left(a_{1} b b_{2} h\right.\). \(h_{1}\left(a_{1} h_{2} h\right.\), and this is not equal to \(x_{1} x_{2}\) unless \(a_{1}^{\prime} u_{1} h_{1} h_{2} h_{1} \quad a_{1}^{\prime} n_{1} h\) and \(l_{1}{ }^{\prime} m_{1} l_{2} h=h_{1} / l_{2} h\). It is not a sufficient condition for this, as seme usually to be supposed, that \(\mathrm{X}_{1}\) and \(\mathrm{X}_{2}\) should be witmesses causally independent of one another. It is also necessary that " and \(b\), i.e. the propositions asserted hy the witnesses, shoukd ho irrelerant to one amother and also pach of them irrelevant to the fact of the assertion of the other by a witness. If a knowledge of a affects the probability pither of \(b\) or of \(h_{1}\). it is evident that the formula braks down. In the ore extreme case, where the assertions of the two contradict one another, \(a b / a_{1} b_{2} l_{1}-1\). In the other extreme. where the two atree in the same assertion, \(i\).e. where \(a \equiv b, a / a_{1} b b_{2} h=1\) and not \(=a / a_{1} h\).
17. The special problem of the agreement of witnesses, who make the same statement, can be best attacked as follows. a certain amount of simplification being introduced. Let the general data hof the problem indude the hymothesis that \(\mathrm{X}_{1}\) and \(X_{2}\) are each asked and reply twa querion to which theme is mily one correct answer. Let \(a_{i}=\) " \(\mathrm{X}_{i}\) asserts \(a\) in reply to the question," and \(m_{i}=\) " \(X_{i}\) gives the correct answer to the question." Then
\[
m_{1} / a_{1} h=x_{1} \text { and } m_{2}^{\prime} / a_{2} h-s_{2}
\]
\(x_{1}\) and \(x_{2}\) being, in the conventional language of thi problem, the "credibilities" of the witnesses. We have, since the witnesses agraw and since a follows from in a and in frillows from mu
\[
\begin{aligned}
& \text { "1"h I" inl: }
\end{aligned}
\]

Also, since the witnesses are, in the ordinary sense, "ind peradent"
witnesses, \(a_{2} / u_{1} a h=a_{2} / a h\) and \(a_{2} / a_{1} \bar{a} h=a_{2} / \bar{a} h\); that is to say, the probability of \(X_{2}\) 's asserting \(a\) is independent of the fact of \(X_{1}\) 's having asserted \(a\), given we know that \(a\) is, in fact, true or false. as the case may be.

The probability that, if the witnesses agree, their assertion is true is
\[
\begin{gathered}
a / a_{1} a_{2} h=m_{1} / a_{1} a_{2} h=\begin{array}{l}
m_{1} a_{2} / a_{1} h \\
a_{2} / a_{1} h
\end{array} \\
=\frac{a_{2} / a_{1} m_{1} h \cdot m_{1} / a_{1} h}{a_{2} a / a_{1} h+a_{2} \bar{a} / a_{1} h}=\frac{a_{2} / a_{1} a h \cdot x_{1}}{a_{2} / a_{1} a h \cdot a_{1} \cdot a_{2} / a_{1} \bar{a} h} \cdot\left(1-x_{2}\right)
\end{gathered} .
\]

If this is to be equal to \(\begin{gathered}x_{1} x_{2} \\ x_{1} x_{2}+\left(1-x_{1}\right)\left(1-x_{2}\right)\end{gathered}\), we must have
\[
\begin{aligned}
& a_{2} / \mu_{1} a h \\
& a_{2} / \mu_{1} \bar{a} h=x_{2} . \\
& 1-x_{2} .
\end{aligned}
\]

Now \(\frac{a_{2} / a_{1} a h}{a_{2} / a_{1} \bar{a} h}=\frac{a_{2} / a h}{a_{2} / \bar{a} h}\) by the hypothesis of "independence"
\[
\begin{aligned}
& \begin{array}{ll}
r_{2} & \bar{a}_{1} / h \\
1-r_{2} & u / h
\end{array} .
\end{aligned}
\]

This then is the assumption which has tacitly slipped into the conventional formula, - that \(a / h=\bar{a} / h=\frac{1}{2}\). It is assumed, that is to say, that any proposition taken at random is as likely as not to be true, so that any answer to a given question is, à priori, as likely as not to be correct. Thus the conventional formula ought to be employed only in those cases where the answer which the " independent" witnesses agree in giving is, a priori and apart from their agreement, as likely as not.
18. A somewhat similar confusion has led to the controversy as to whether and in what manner the a priori improbability of a statement modifies its credibility in the mouth of a witness whose degree of reliability is known. The fallacy of attaching the same weight to a testimony regardless of the character of what is asserted, is pointed out. of course, by Hume in the Essay on Mirudes, and his argument, that the great a priori improbability of some assertions outweighs the force of testimony otherwise reliable, depends on the avoidance of it. The correct riew is also taken hy Laplace in his Essai phitosophique (pp.

98-1(12). where he argues that a witness is luas to be belineed when he asserts an extraerdinary fact. dectaring the opposit. view (taken by Diderot in the article on "Certitude" in the Encegclopédie) to bue inconcervable before " le simple ben sens.

The manner in which the resultant probabilite is affeeted depends upon the precise meaning we attach to "degree of reliability " or "coefficient of credibility." If a witness's credibility is represented ber do we mean that, if "is the true answer, the probability of his giving it is \(x\). or do we mean that if he answers a the probability of \(a\) shmeng true is \(x\) ? These two thing are not equivalent.

Let \(a_{1}\) stand for " \(a\) is asserted by the witness" ; \(h_{1}\) for our evidence bearing on the witness's veracity; and \(h_{2}\) for other evidence bearing on the truth of 1 . Let ai \(h_{1} h_{2}\). ie. the a primi probability of "apart from our knowledge of the fact that the witness has asserted it, be represented by \(p\).

Let \(a / a_{1} h_{1} \quad r_{1}\) and \(\|_{1} / w h_{1} \quad r_{2}\) : st) that \(x_{1} \quad \begin{gathered}\| \\ h_{1}\end{gathered} \mu_{1} / h_{1}\). In general \(u_{1} / h_{1}+\pi_{3} / h_{1}\). Do we mean by the witnessis credibility \(x_{1}\) or \(x_{2}\) !

We require " \(1_{1}^{\prime} H_{1} h_{2} h_{2}\).
Let \(a_{1} / \bar{a} h_{1}=r\), i.e. the probability, apart from our special knowleden comeremine ", that if "is false, the witness will hit on that particular falsehood.
 knowlodere comerning \(a\). \(h_{2}\) is irmen ant to the probabilit? of \(n_{1}\).
19. Generally speaking, all problems, in regard to the combination of testimonies or to the combination of evidence derived from testimony with evidence derived from other sources, may be treated as special instances of the general problem of the combination of arguments. Beyond pointing out the above plausible fallacies, there is little to add. Mr. W. F. Johnson, howewer, has properal a method of detiming codiblits, which is sometimes valuable, because it regards the witness's credibility not absolutely, but with reference to a given type of question,
so that it enables us to measure the force of the witness's testimony under special circumstances. If a represents the fact of A's testimony regarding \(x\), then we may define A's credibility for \(x\) as \(a\), where \(a\) is given by the equation
\[
x / a h=x / h+a \sqrt{ } x / h \cdot \bar{x} / h ;
\]
so that \(a \sqrt{x} / \bar{h} \cdot \bar{x} / h\) measures the amount by which A's assertion of \(x\) increases its probability.
20. One of the most ancient problems in probability is concerned with the gradual diminution of the probability of a past event, as the length of the tradition increases by which it is established. Perhaps the most famous solution of it is that propounded by Craig in his Theologiae Christianae Principia Mathematica, published in 1699. \({ }^{1}\) He proves that suspicions of any history vary in the duplicate ratio of the times taken from the beginning of the history in a manner which has been described as a kind of parody of Newton's Principin. "Craig," says Todhunter, "concluded that faith in the Gospel so far as it depended on oral tradition expired about the year 880, and that so far as it depended on written tradition it would expire in the year 3150. Peterson by adopting a different law of diminution concluded that faith would expire in 1789." \({ }^{2}\) About the same time Locke raised the matter in chap. xvi. bk. iv. of the Essay Concerning Human Understanding: "Traditional testimonies the farther removed, the less their proof. . . . No Probability can rise higher than its first original." This is evidently intended to combat the view that the long acceptance by the human race of a reputed fact is an additional argument
\({ }^{1}\) See Todhunter's History, p. 54. It has been suggested that the anonymous essay in the Phil. Trans. for 1699 entitled "A Calculation of the Credibility of Human Testimony" is due to Craig. In this it is argued that, if the credibilities of a set of witnesses are \(p_{1} \ldots p_{n}\), then if they are successive the resulting probability is the product \(p_{1} p_{2} \ldots p_{n}\); if they are concurrent, it is :
\[
1-\left(1-p_{1}\right)\left(1 \quad p_{2}\right) \ldots\left(1-p_{\prime \prime}\right) .
\]

This last result follows from the supposition that the first witness leaves an amount of doubt represented by \(1-p_{1}\); of this the second removes the fraction \(p_{2}\), and so on. See also Lacroix, Traité élémentaire, p. 262. The above theory was actually adopted by Bicquilley.
\({ }^{2}\) In the Budget of Paradoxes De Morgan quotes Lee, the Cambridge Orientalist, to the effect that Mahometan writers, in reply to the argument that the Koran has not the evidence derived from Christian miracles, contend that, as evidence of Christian miracles is daily weaker, a time must at last arrive when it will fail of affording assurance that they were miracles at all : whence the necessity of another prophet and other miracles.
in its favour and that a long tradition increans rather than diminishes the strength of an assertion. ". This is certain." savs Locke, " that what in one age was affirmed upon slight grounds, can never after come to be more valid in future ages, by being often repeated." In this connection he calls attention to "a rule observed in the law of England, which is, that though the attested copy of a record be good proof, yet the copy of a copy never so well attested, and by never so credible witnesses, will not ber admitted as a proof in Judicature." If this is still a groul rule of law, it seems to indicate an excessive subservience to the principle of the decay of evidence.

But, although Locke affirms sound maxims, he gives no theory that ran afterd absis for calculation. Crais, howewer. was the more typical professor of probability, and in attempting an algeltraic formula he was the first of a considetable family. The last arand discussion of the problem towk place in the columnof the Educational Times. \({ }^{1}\) Macfarlane \({ }^{2}\) mentions that four different mbutions have been put forward by mathematicians of the problem: " A says that B says that a certain event took place; required the probability that the event did take place, \(\mu_{1}\) and \(\mu_{2}\) being 1 's and B's respective probabilities of spuahing the truth." Of these solutions only ('ayley's is correct.
\({ }^{1}\) Reprinted in Mathematics from the Educational Times, vol. xxvii.
 problem without success. Its solution is not difficult, if enough unknowns are introduced, but of very little interest.

\section*{CHAPTER XVII}

\section*{SOME PROBLEMS IN INVERSE PROBABILITY, INCLUDING AVERAGES}
1. The present chapter deals with 'problems'- that is to say, with applications to particular abstract questions of some of the fundamental theorems demonstrated in Chapter XIV. It is without philosophical interest and should probably be omitted by most readers. I introduce it here in order to show the analytical power of the method developed above and its advantage in ease and especially in accuracy over other methods which have been employed. \({ }^{1}\) § 2 is mainly based upon some problems discussed by Boole. §§ 3-7 deal with the fundamental theory connecting averages and laws of error. \(\S \S 8-11\) treat discursively the Arithmetic Average, the Method of Least Squares, and Weighting.
2. In the following paragraph solutions are given of some problems posed by Boole in chapter xx. of his Laws of Thought. Boole's own method of solving them is constantly erroneous, \({ }^{2}\) and the difficulty of his method is so great that I do not know of any one but himself who has ever attempted to use it. The term 'cause ' is frequently used in these examples where it might have been better to use the term 'hypothesis.' For by a possible cause of an event no more is here meant than an antecedent occurrence, the knowledge of which is relevant to our anticipation of the event; it does not mean an antecedent from which the event in question must follow.
(56) The a priori probabilities of two causes \(A_{1}\) and \(A_{2}\) are \(c_{1}\) and \(c_{2}\) respectively. The probability that if the cause \(A_{1}\)

\footnotetext{
\({ }^{1}\) Such examples as these might sometimes be set to test the wits of students. The problems on Probability usually given are simply problems on mathematical combinations. These, on the other hand, are really problems in logic.
\({ }^{2}\) For the reason given in \(\S 6\) of Chapter XVI. The solutions of problems I.-VI., for example, in the Laws of Thought, chap. xx., are all erroneous.
}
occur, an event E will :teompany it (whether as a consequence of \(A_{1}\) or not) is \(p_{1}\), and the probability that \(E\) will accompany \(A_{2}\). if \(A_{2}\) present itself, is \(p_{2}\). Moreover. the event E camont appar in the absence of both the causes \(A_{1}\) and \(A_{2}\). Revpuied the probability of the event E.

This problem is of great historical interest and has been called Boole's ' (hallenge Problem. Boole originally proposid it for solution to mathematicians in 18.51 in the ('ambritge amd I)ublin Mathemutionl Jomrmel. A result was given by ('ayley \({ }^{1}\) in the Philosophicul Magazine. which Bowde declared to bereroneous. \({ }^{2}\) He then entered the field with his own solution. \({ }^{3}\) "Several attempts at its solution," he says, " have been forwarded ter me, all of them he mathematicians of great eminence, all of them admittine of particular verification, yot differing from each other and from the truth." \({ }^{4}\) After calculations of comsiderable length and great diftioulty he arrives at the conclusion that \(1 /\) is the probability of the event E where " is that roon of the equation
which is not less than \(c_{1} p_{1}\) and \(c_{2} \mu_{2}\) and not greater than \(1 \cdot c_{1}\left(1 \quad \mu_{1}\right), 1 \quad c_{2}\left(1 \quad \mu_{2}\right) \cdot\left(m^{\prime} c_{1} p_{1} \quad c_{2} l_{2}\right.\),

This solution can casily be seen to be wrong. For in the case where \(A_{1}\) and \(A_{2}\) cannot both occur, the solution is \(u=c_{1} p_{1}+c_{2} p_{2}\); whereas Boole's equations do not reduce to

 vol. 1. p. 268). The difference arises out of the extreme ambiguity as to the meaning of the terms as employed by (ayley.

3 "Solution of a Question in the Theory of Probabilities," Phil. Mag. 4th series, vol. vii., 1s.it. This solution is the same as that printed by Boole
 Wilbraham grave as the solution \(\| c_{1} p_{1} \cdot c_{2} p_{2} \quad\) o. where \(z\) is necessarily less than either \({ }^{\prime} p_{1} p_{1}\) or \(c_{0} p_{2}\). This solution is correct so far as it goes, but is not
 p. 154 .
\({ }^{4}\) In proposing the problem Boole had said: "Thee motives which have led me, after much con-ideration, to adopt, with reforence to this question, a course unusual in the present day, and not upon slizht grounds to be revived, are the following: First, I propose the question at a test of the sufficiency of received methods. Seoondly. I antiopate that its discussion will in some measure add to our knowledge of an important branch of pure amalyais. When printing his own solution in the Lauw of Thought, he adds, that the above "led to some interesting private correspondence, but did not elicit a *)
this simplified form. The mistake which Boole has made is the one general to his system, referred to in Chapter XVI., § \(6 .{ }^{\text {. }}\)

The correct solution, which is very simple, can be reached as follows:

Let \(a_{1}, a_{2}, e\) assert the occurrences of the two causes and the event respectively, and let \(h\) be the data of the problem.

Then we have \(a_{1} / h=c_{1}, a_{2} / h=c_{2}, e / a_{1} h=p_{1}, e / a_{2} h=p_{2}\) : we require \(e / h\). Let \(e / h=u\), and let \(a_{1} a_{2} / e h=z\). Since the event cannot occur in the absence of both the causes,
\[
c / \bar{a}_{1} \bar{a}_{2} l_{l}=0 .
\]

It follows from this that \(\bar{a}_{1} \bar{a}_{2} / e h=0\), unless \(e / h=0\), i.e.
\[
\left(\left\|_{1}+\right\|_{2}\right) / p l_{1}=1 .
\]
whence
\[
\begin{equation*}
{ }_{1} / e h+a_{2} / e h_{1}=1+\left({ }_{1}{ }^{\prime \prime}{ }_{2} / \cdot \rho h_{1}\right. \tag{24}
\end{equation*}
\]

Now
\[
\begin{aligned}
& \therefore u=\begin{array}{c}
c_{1} \prime_{1}+c_{2}{ }^{\prime}{ }_{2}, \\
1+z
\end{array}
\end{aligned}
\]
where \(z\) is the probability after the event that both the causes were present.

If we write \(e a_{1} a_{2} / h=y\),
\[
\begin{aligned}
& \text { ! } / \text { - " } 11^{\prime \prime}{ }_{2}{ }^{\prime} c l_{1} \cdot \rho / h_{1}=u \approx \\
& u==\left(r_{1} \mu_{1}+r_{2} \prime^{\prime} 2\right)-\mu .
\end{aligned}
\]
so that
Boole's solution fails by attempting to be independent of \(y\) or \(z\).
(56.1). Suppose that we wish to find limits for the solution which are independent of \(y\) and \(z\) : then, since \(y>0\), \(\mu\) こ \({ }_{1} \|_{1}{ }_{1}: c_{2} \mu_{2}\).

Again

Similarly \(\quad / h \approx 1-r_{2}+e_{2} / \rho_{2}\). From the same equations it appears


\footnotetext{
\({ }^{1}\) Boole's error is pinted thet and a correct solution wiven in Mr. McColl's "Sivth Article on the ('alculus of Equivalent Statements" (P'roc. Lond. Math.
Soc. vol. xxviii. p. \(\overline{\text { Sitel }}\) ).
}
\(\therefore u\) lies between
\[
\text { the greatest of } \left\lvert\, \begin{aligned}
& 1 / 1 / 1 \\
& \mid r_{2} / 2
\end{aligned}\right. \text { and the least of } \left\lvert\, \begin{aligned}
& 1 / 1 \\
& 1 \\
& 1-c_{1}(1-/ 2) \\
& 1
\end{aligned}\right.
\]

It will be seen that these numerical limits are the same as the limits obtained by Boole for the routs of his equations.
(56.2) Suppose that the a prion probatilition of the camses is and \(c_{2}\) are to be eliminated. The only limit we then have is " \(H_{1} H_{2}\).
(.95.3) suppese that one of the it primi probabilities cen is to be
 fore, \(c_{1}\) is large, \(u\) does not differ widely from ' \(r_{1} p_{1}\).
(56.t) Suppose \(p_{2}\) is to be eliminated. We then have
\begin{tabular}{ll}
\(1 / 1\) & \(1 / 2\) \\
\(\cdots\) & \(1 / 1\)
\end{tabular}

If therefore \(c_{1}\) is large or \(c_{2}\) small, \(u\) does not differ widely from ' \({ }^{1} /{ }^{\prime}\) '.
(56.5) If \(a_{1} / h_{2} h=a_{1} / h\), i.e. if our knowledge of each of the causes is independent, we have a closer approximation. For
\[
\begin{aligned}
& \therefore \text { " } 1 / 1+2 / 2{ }^{\prime \prime} 11^{2} \cdot 1^{\prime \prime} 2 \text { 。 }
\end{aligned}
\]
(57) We may now generalise (56) and discuss the case of \(n\) causes. If an event can only happen as a consequence of one or more of certain causes \(\mathrm{A}_{1}, \mathrm{~A}_{2}, \ldots \mathrm{~A}_{\text {, }}\), and if \(c_{1}\) is the a priori probability of the cause \(A_{1}\) and \(p_{1}\) the probability that, if the cause \(\mathrm{A}_{1}\) be known to exist, the event E will occur: required the probability of E .

This is Boole's problem VI. (Laws of Thought, p. 3336). As the result of ten pages of mathematics, he finds the solution to be the root lying between certain limits of an "quation of the \(n^{\prime}\) degree which he camon solve. I know no other discussion of the problem. The solution is as follows :
\[
\begin{aligned}
& 1
\end{aligned}
\]
\[
\begin{aligned}
& \therefore e / h=e \bar{a}_{1} \bar{a}_{2} / h+c_{1} p_{1}+r_{2} \prime_{2}-e \mu_{1} \alpha_{2} / h, \\
& c \bar{a}_{1} \bar{a}_{2} / h=e \bar{a}_{1} \bar{a}_{2} \bar{a}_{3} / h+e \bar{a}_{1} \bar{a}_{2} a_{3} / h, \\
& \epsilon \bar{a}_{1} \bar{a}_{2} \iota_{3} / h=e \bar{a}_{1} \bar{a}_{2} / \alpha_{3} h \cdot c_{3}=c_{3}\left\{c \mid{ }^{\prime} \epsilon_{3} h-e \bar{a}_{1} \bar{a}_{2} / a_{3} h\right\} \\
& =c_{3} p_{3}-e \bar{a}_{1} \bar{a}_{2^{\prime \prime}} / h, \\
& \therefore e / h=e \bar{a}_{1} \bar{a}_{2} \bar{a}_{3} / h+c_{1} p_{1}+c_{2} p_{2}+c_{3} n_{3}-\rho\left(\overline{\bar{T}}_{1} \ell_{2} / h-e \bar{a}_{1} \bar{a}_{2^{\prime}{ }_{3}} / h .\right.
\end{aligned}
\]
and

In general
\[
\begin{aligned}
e \bar{a}_{1} \bar{a}_{2} \ldots \bar{a}_{r-1} / h & =e \bar{a}_{1} \bar{a}_{2} \ldots \bar{a}_{r-1} \bar{a}_{r} / h+e \bar{a}_{1} \bar{a}_{2} \ldots \bar{a}_{r-1} a_{r} / h \\
& =e \bar{a}_{1} \ldots \bar{a}_{r} / h+e \bar{a}_{1} \ldots \bar{a}_{r-1} / c_{1} / h \ldots \\
& =e \bar{a}_{1} \ldots \bar{a}_{r} / h+c,\left\{e / / u_{r} h-e \bar{u}_{1} \ldots \bar{u}_{r-1} / c_{1}, h\right\} \\
& =e \bar{a}_{1} \ldots \bar{u}_{r} / h+c p_{r}-e \bar{a}_{1} \ldots \bar{a}_{r-1} a_{r} / h ;
\end{aligned}
\]

But since the \(n\) causes are supposed to be exhaustive
\[
\begin{gather*}
\left.c \overline{1}_{1} \ldots \bar{a}_{\prime \prime} / h_{r}=0\right) . \\
\therefore \varepsilon / h=\sum_{1}^{n} c, \eta_{r}-\sum_{-}^{n} e \bar{u}_{1} \ldots \bar{a}_{r-1^{\prime \prime}}, / h \tag{ii.}
\end{gather*}
\]

Let
\[
\begin{equation*}
m \bar{u}_{1} \ldots \bar{u}_{r-1} l^{\prime} / / h=n_{r} ; \tag{iii.}
\end{equation*}
\]
then
(57.1) If our knowledge of the several causes is independent, if, that is to say, our knowledge of the existence of any one of them is not relevant to the probability of the existence of any other, so that \(a_{r} / u c_{N} h=u_{r} / h=c_{r}\), then
\[
\begin{aligned}
& \ell \bar{a}_{1} \ldots \bar{a}_{r-1} a_{1} / h=e \bar{a}_{1} \ldots \bar{u}_{1} /{ }_{1} / h, h \cdot c_{r} \\
& =c_{,}, r^{\prime} / \bar{u}_{1} \ldots \bar{u}_{r-1} 1^{\prime \prime}, h\left\{1-\bar{u}_{1} \ldots \bar{u}_{,} 1_{1} / u_{1} h\right\} \\
& =c \cdot\left[1-1 I\left(1-c_{1}\right) \ldots\left(1-c_{,}\right)\right]{ }_{1} / \bar{u}_{1} \ldots \bar{a}_{r} 1^{q}, h .
\end{aligned}
\]

Let
\[
\bar{u}_{1} \ldots \bar{a}_{r} \quad 1^{\prime \prime}, h \quad m_{r}
\]
then
\[
\varepsilon / h=\sum_{r=1}^{\prime \prime} "_{r} \mu_{r}-\sum_{2}^{n}\left[1-\|_{n=1}^{\prime}\left(1-r_{s}\right)\right] m_{r}
\]

These results do not look very promising as they stand, but they lead to some useful approximations on the elimination of \(m_{r}\) and \(n_{r}\) and to some interesting special cases.
(57.2) From equation (i.) it follows that e/h \(c_{1} p_{1}\) and \(e / h \geq 1-c_{1}\left(1-p_{1}\right)\); and from equation (ii.) that elh. Ёc \(p\) :
\(\therefore e / h\) lies between

(57.3) Further. if the canses are independent it follows from (57.1) that
\[
\left.h \geq \prime \vdots\left[\begin{array}{lll}
1 & \|_{1}^{\prime}(1 & 1
\end{array}\right)\right]
\]
so that \(e / h\) lies between

(57.4) Now consider the case in which \(p_{1}=p_{2}=\ldots=p_{n}=1\), i.e. in which any of the causes would be sufficiont, and in which the causes are independent. Then \(m_{r}=1\); so that
\[
\left.\begin{array}{rllll}
1, h & -1 & \leq 11 & 11(1 & 1
\end{array}\right) \mid
\]
(57.5) Let \(c_{1}, c_{2} \ldots c_{n}\) be small quantities so that their squares and products may be neglected.

Then
\[
e^{/ / h} \text { ジ } c, l \text {. }
\]
i.e. the smaller the probabilities of the causes the more do they approach the condition of being mutually exclusive. \({ }^{1}\)
(57.6) The it pmisteriuri probability of a particular cialse a after the event has been observed is
\[
\begin{aligned}
& \text { ", /he } \quad \text {, h h } \\
& \stackrel{\prime}{\prime}
\end{aligned}
\]
(This is Boole's problem IX., p. 357).

(58) The probability of the occurrence of a certain natural phenomenon under given circumstances is \(p\). There is also a probability \(a\) of a permanent cause of the phenomenon, i.e. of a cause which would always produce the event under the circumstances supposed. What is the probability that the phenomenon, being observed \(n\) times, will occur the \(n+1^{\text {th }}\) ?

This is Boole's problem X. (Laws of Thought, p. 358). Boole arrives by his own method at the same result as that given below. It is necessary first of all to state the assumption somewhat more precisely. If \(x\), asserts the occurrence of the event at the \(r^{t h}\) trial and \(t\) the existence of the 'permanent cause' we have
and we require
\[
\begin{gathered}
x_{1} / h=\mu, t / h=a, x_{r} / t h=1, \\
x_{n+1} / x_{1} \ldots c_{n} / h=y_{n=1} .
\end{gathered}
\]

It is also assumed that if there is no permanent cause the probability of \(x_{s}\) is not affected by the observations \(x_{r}\), etc., i.e.
\[
y_{1} y_{2} \cdots y_{1-1}
\]

Also
\[
y_{1}=\mu \text { and }!_{2} \quad\left(t+(\mu-⿲)^{\prime \prime-"} 1-u\right.
\]
\[
!1
\]

\footnotetext{
1 This assumption, which is tacitly introduced by Boole, is not generally justifiable. I use it hore, as my main purpose is to illustrate a method. The same problem, without this assumption, will be discussed in dealing with Pure Induction.
}
\[
\begin{aligned}
& x_{s} / i_{1}, \ldots x_{t}, \bar{t} l_{\ell}=s_{n}\left[T h,^{1}\right.
\end{aligned}
\]
\[
\begin{aligned}
& x_{r} / x_{1} \ldots r_{1}{ }_{1} h=x_{1} t \mid x_{1} \ldots r_{1}{ }_{1} h+b_{1} \bar{t} / n_{1} \ldots r_{1}{ }_{1} h
\end{aligned}
\]
\[
\begin{aligned}
& \left(1+(\prime \prime-")\binom{\prime \cdots-"^{\prime}}{1-u}^{r 1}\right.
\end{aligned}
\]
so that
\[
\text { ": (" ") } 1
\]
- "
(5).1) If \(p=a, y_{n}=1\); for if an event can only occur as the result of a permanent cause, a single occurrence makes future occurrences certain under similar conditions.
\[
\left.(" 1,-\prime \prime)^{\prime}\right|_{1} ^{\prime} \quad=\quad 1-\begin{array}{cc}
l^{\prime} & u^{\prime}  \tag{58.2}\\
1 & \prime \prime
\end{array}
\]
\[
\begin{aligned}
& !1 \text { ! }
\end{aligned}
\]
(by easy algebra) :
and \(\mu\) is always \(\because n\) and .1 .
So that \((\eta-u)\left(\begin{array}{cc}\mu^{\prime}-{ }^{\prime} \\ 1 & -\|\end{array}\right)\) is positive and decreases as \(r\) increases,
\[
\therefore!1 \quad 1 \quad!\text {. }
\]

As \(n\) increases \(y_{n}=1-\epsilon\), where
so that for any value of \(\eta\) however small a value of \(n\) can be found such that \(\epsilon<\eta\) so long as \(a\) is not zero.
(58.3) \(t_{n}\) the \(\dot{a}\) posterior probability of a permanent cause after \(n\) successful observations is
\[
\begin{aligned}
& \text { ide. } \\
& \text { ". (1, ", ! ! }
\end{aligned}
\]
\[
\begin{aligned}
& 1 \text { 1, inhere, }
\end{aligned}
\]

So that \(t_{n}\) approaches the limit unity as \(n\) increases, so long as a is not zero.
3. The following is a common type of statistical problem. \({ }^{1}\) We are given a series of measurements, or observations, or estimates of the true value of a given quantity ; and we wish to determine what function of these measurements will yichl us the most probable value of the quantity. on the basis of this evidence. The problem is not determinate unless we have some good ground for making an assumption as to how likely we are in each case to make errors of given magnitudes. But such an assumption, with or without justification, is frequently made.

The functions of the original measurements which we commonly employ, in order to yield us approximations to the most probable value of the quantity measured, are the various linds of means or averages-the arithmetic mean, for example, or the median. The relation, which we assume, between errors of different magnitudes and the probabilities that we have made errors of those magnitudes, is called a law of error. Corresponding to each law of error which we might assume, there is some function of the measurements which represents the most probable value of the quantity. The object of the following paragraphs is to discover what laws of error, if we assume them, correspond to each of the simple types of average, and to discover this by means of a systematic method.
4. Let us assume that the real value of the quantity is either \(b_{1}, \ldots b_{r} \ldots b_{n}\), and let \(a_{r}\) represent the conclusion that the value is, in fact, \(b_{\text {, }}\). Further let \(x_{\text {, represent the evidence that }}\) a measurement has been made of magnitude \(y_{r}\).

If a measurement \(y_{p}\), has been made, what is the probability that the real value is \(b\) ? The application of the theorem of inverse probability yields the following result:
(the number of possible values of the quantity being \(n\) ), where \(h\) stands for any other relevant evidence which we may have, in addition to the fact that a measurement \(x_{p}\), has been made.

Next, let us suppose that a number of measurements \(y_{1} \ldots y_{\text {II }}\)
\({ }^{1}\) The substance of \(\S \S 3.7\) has been printed in the Journal of the Royal Slatistical Socicty, vol. Ixxiv. p. 323 (February 1911).
have been made: what is now the probability that the real value. is \(b_{*}\) ? We require the value of \(a_{N} / x_{1} x_{2} \ldots x_{m} h\). As before,

1
At this point we must introduce the simplifying assumption that, if we knew the real value of the quantity, the different measurements of it would be independent, in the sense that a knowledge of what errors have actually been made in some of the measurements would mot affect in any way our "stimate of what errors are likely to be made in the others. Wee assume: in fact, that \(x_{r} / x_{\mu} \ldots x_{s} \mu_{r} h=x_{r} / a_{r} h\). This assumption is exceedingly important. It is tantamount to the : assumption that our law of error is unchanged throughout the ser rims of wham ant ions ing question. The general evidence h, that is today: which justifies our assumption of the particular law of error which we do assume, is of such a character that a knowledge of the act hal eros made in a number of measurements, not more numerous than those in question, are absolutely or approximately irmbevant to the question of what form of law we ought to assume. The law of error which we assume will be based, presumably, on an experima of the relative frequency wi th which amon of dilliomon magnitudes have been made under analonenes eiremmstances in the past. The above assumption will not be justified if the additional experience, which a kmesteder of the errors in the new measurements would supply, is sulficionty compmemorive, relay timely the cur former experience, to be capable of mendifyine our assumption as ter the shaper of the lat of error, of if it sumer that the circumstances, in which the measurements are being


With this assumption, ie. that \(x_{1}\), etc., are independent of one another relatively to evidence a ho ce... it follows from the ordinary rule for the multiplication of indepmement probabilities that
\[
r_{1} \ldots s_{\infty} f_{1} h \|_{1} \ldots h_{1}
\]
\[
\begin{aligned}
& \text { a 1. } 1 \mathrm{~L} \cdot \mathrm{ln} h \\
& \geq \sum_{1}\left|\begin{array}{cccc}
11 & 1 & 11 & 11 \\
1 & 11
\end{array}\right|
\end{aligned}
\]

Hence
\[
" x_{1} x_{2} \ldots, h=
\]

The most probable value of the quantity under measurement, given the \(m\) measurements \(y_{1}\), etc.-which is our quaesitum - is therefore that value which makes the above expression a maximum. Since the denominator is the same for all values of \(b\), we must find the value which makes the numerator a maximum. Let us assume that \(a_{1} / h=a_{2} / h=\ldots=a_{n} / h\). We assume, that is to say, that we have no reason \(a\) priori (i.e. before any measurements have been made) for thinking any one of the possible values of the quantity more likely than any other. We require, therefore, the value of \(b\), which makes the expression \({ }^{\text {" }}{ }^{m} x_{1} / a_{s} h\) a maximum. Let us denote this value by \(y\).

We can make no further progress without a further assumption. Let us assume that \(x_{4} / a_{N} h\)-namely, the probability of a measurement \(y_{1 \prime}\) assuming the real value to be \(b\)-is an algebraic function \(f\) of \(y_{1}\), and \(b_{s}\), the same function for all values of \(y_{q}\) and \(b_{\text {s }}\) within the limits of the problem. \({ }^{1}\) We assume, that is to say, \(x_{4} / a h=f\left(y_{1}, b_{*}\right)\), and we have to find the value of \(b_{\text {s }}\) namely \(y\), \begin{tabular}{l}
\(=\) ="' \\
\hline
\end{tabular}
which makes II \(f(y, y)\) a maximum. Equating to zero the \(\eta=1\)
differential coefficient of this expression with respect to \(y\), we have \(\sum_{\|=1}^{"={ }^{\prime \prime}} \frac{f^{\prime}\left(y_{q}, y\right)}{f\left(y_{i,}, y\right)}=0,{ }^{2}\) where \(f^{\prime}=\frac{d f}{d y}\). This equation may be written for brevity in the form \(\Sigma{\frac{f^{\prime}}{f_{q}}}_{f_{i}}=0\).

Ii we solve this equation for \(y\), the result gives us the value of the quantity under observation, which is most probable relatively to the measurements we have made.

The act of differentiation assumes that the possible values of \(y\) are so numerous and so uniformly distributed within the range in question, that we may, without sensible error, regard them as continuous.
5. This completes the prolegomena of the inquiry. We are

\footnotetext{
\({ }^{1}\) Gauss, in obtaining the normal law of error, made, in effect, the more special assumption that \(x_{q} / a_{g} h\) is a function of \(e_{q}\) only, where \(e_{q}\) is the error and \(e_{q}=b_{8}-y_{q}\). We shall find in the sequel that all symmetrical laws of error, such that positive and negative errors of the same absolute magnitude are equally likely, satisfy this condition-the normal law, for example, and the simplest medjan law. But other laws, such as those which lead to the geometrie mean, do not satisfy it.
\({ }^{2}\) Since none of the measurements actually made can be impossible, none of the expressions \(f\left(y_{q}, y\right)\) can vanish.
}
now in a position to discover what laws of error correspond to given assumptions respectine the algabatic relation hetaren the measurements and the most probahle value of the quathtit and vice verse For the law of aronr detemnines the form of \(f(\%, y)\). And the form of \(f(y, y)\) demernines the aloperatic retation \(\frac{f^{\prime}}{f}\)-11 between the measurements and the most prohable value. It may be well to repeat that \(f(y, y)\) denotes the probbibility \({ }^{\prime}\), us that an observer will make a measurement ! in observing a quantity whose true value we know to be \(y\). A law of error tells us what this probability is for all pmosible values of ! 1 and \(\%\) within the limits of the problem.
(i.) If the most prohahbe value of the quantity is equal to the arithnetic mean of the meanurements. What lan of mran dons this imply ?
\[
\begin{aligned}
& \geq f^{\prime} \quad\left(1 \text { must the equivalent } 10 \leq(y-y)-0 \text {, since. } t_{1} .\right. \\
& \text { most probable value !/ must equal } \begin{array}{c}
\text { In } \\
\text { II } \\
\text {-y. }
\end{array} \\
& \therefore f_{f}^{\prime \prime}=\phi^{\prime \prime}(y)\left(y-y_{y}\right) \text { where } \phi^{\prime \prime}(y) \text { is some function which } \\
& \text { is not zero and is independent of } y_{\%} \text {. }
\end{aligned}
\]

Integrating,
\[
\begin{aligned}
& \log f \quad \phi^{\prime \prime}(y)(y-y) d y \cdot y(y) \text { where } \psi(y) \text { is somue func- } \\
& \text { tion independent of } y . \\
& \phi^{\prime}(y)(y-y)-\phi(y) \quad \psi(y) .
\end{aligned}
\]
sis that \(f,(x,-\infty)-() \cdot()\)
Any law of arror of this type, therefore, leads to the arithmetic mean of the measurements as the most probable value of the quantity measured.
 \(f\) - An \(\vec{j}^{2}\). the form mormally amomumal.
\(=A e^{-i z_{2}^{2}}\), where \(z_{11}\) is the absolute magnitude of the error in the measurement \(y_{i q}\).

This is, clearly, only one amongst a number of possible solutions. But with one additional assumption we can prove that this is the only law of error which leads to the arithmetic mean.

Let us assume that negative and positive errors of the same absolute amount are equally likely.

In this case \(f_{q}\) must be of the form \(\mathrm{Be} e^{\theta\left(y-y_{q}\right)^{2}}\),
\[
\therefore \phi^{\prime}(y)\left(y-y_{q}\right)-\phi(y)+\psi\left(y_{q}\right)=\theta\left(y-y_{q}\right)^{2} .
\]

Differentiating with respect to \(y\),
\[
\phi^{\prime \prime}(y)=2_{d\left(y-y_{4}\right)^{2}}^{d} \theta\left(y-y_{4}\right)^{2} .
\]

But \(\phi^{\prime \prime}(y)\) is, by hypothesis, independent of \(y_{q}\).
\(\therefore \stackrel{d}{d\left(y-y_{q}\right)^{2}} \theta\left(y-y_{q}\right)^{2}=\cdots k^{2}\) where \(l_{i}\) is constant; integrating, \(\theta\left(y-y_{4}\right)^{2}=-k^{2}\left(y-y_{4}\right)^{2}+\log C\) and we have \(f_{4}=A e^{-1^{2}(y-\ldots)^{2}}\) (where \(\mathrm{A}=\mathrm{BC})\).
(ii.) What is the law of error, if the geometric mean of the measurements leads to the most probable value of the quantity ?

In this case \(\frac{f^{\prime}}{f_{\eta}^{\prime \prime}}=0\) must be equivalent to \(\prod_{\eta=1}^{\prime=y_{n}}=y^{\prime \prime \prime}\), i.e. to
\[
\leq \log _{y}^{y_{11}}=0
\]

Proceeding as before, we find that the law of error is

There is no solution of this which satisfies the condition that negative and positive errors of the same absolute magnitude are equally likely. For we must have
\[
\begin{aligned}
& \phi^{\prime}(y) \log \frac{y_{1 \prime}}{y}+\int_{y}^{\phi^{\prime}(y)} d y+\psi\left(y_{4}\right)=\phi\left(y-y_{4}\right)^{2} \\
& \text { or } \phi^{\prime \prime}(y) \log _{y}^{y_{y}}=d_{d_{y}}^{d} \phi\left(y-y_{y}\right)^{2},
\end{aligned}
\]
which is impossible.
The simplest law of error, which leads to the geometric mean, seems to be whtained by putting \(\phi^{\prime}(y)=\) kiy, \(\psi\left(y_{4}\right)=0\). This gives \(\int_{4}=\mathrm{A}\binom{y}{y_{4}}^{1 \cdot 1 /} e^{-h^{-1 / 4}}\).

A law of error. which leads to the geometric mean of the observations as the most probable value of the quantity, has been previonsly discussed by Sir Donald McAlister (Proceedings of the Royal Society, vol. xxix. (1879) p. 365). His investigation de pends upon the obvious fact that, if the geometric mean of the
observations yindes the mose pmotable value of the quantity, the. arithmetic mean of thie lesarithmes of the ohorerations musi y whe the most probable value of the herarithm wh the quantity. He.... if we suppose that the logarithms of the observations obey the normal law of error (which leads to their arithmetic mean is the ment probable value of the lograrithme of the quantity), "re call by substitution find a law of error for the observations themselves which must lead to the geometric mean of them as the most probable value of the quantity itself.

If, as before, the observations are denoted by \(y_{11}\), etc., and the quantity by \(y\), let their logarithms be denoted by \(l_{p}\), etc., and by \(l\). Then, if \(l_{l /}\), ete., obey the normal law of error, \(f\left(l_{l, l} l\right)=A e^{-h^{2}\left(l_{2}-1\right)^{2}}\). Hence the law of error for \(y_{4}\), ctc., is determined by
\[
\begin{aligned}
&\left.f(y, y) \quad \text { A } e^{2} \operatorname{lng}-\log ,\right)^{2} \\
& x^{-2}(\log ;)^{2}
\end{aligned}
\]
and the most probable value of \(y\) must, clearly, be the geometric mean of \(y_{t}\), etc.

This is the law of error which was arrived at by Sir Donald McAlister. It can easily be shown that it is a special case of the generalised form which I have given above of all laws of error leading to the geometric mean. For if we put \(\psi\left(y_{1}\right)=-k^{2}\left(\log y_{i}\right)^{2}\), and \(\phi^{\prime}(y) \quad 2 i^{2}\) lung. Wh hath
\[
\begin{aligned}
& \therefore \quad \mathrm{A} \quad=(\mathrm{m},)^{\prime}
\end{aligned}
\]

A similar result has been obtained by Professor J. C. Kapteym. \({ }^{1}\) But he is investigating frequency curves, not laws of error, and this result is merely incidental to his main discussion. His method, however, is not unlike a more generalised form of Sir Donald MeAlister's. In order to discover the frequency curve of certain quantities \(y\), he supposes that there are cortain other quantitics \(z\), functions of the quantities \(y\), which are given by \(z=F(y)\), and that the frequency curve of these quantities \(z\) is normal. By this device he is enabled in the investigation of a type of skew frequency curve, which is likely to be met with often, to utilise certain statistical constants corresponding to

\footnotetext{


}
those which have been already calculated for the normal curve.

In fact the main advantage both of Sir Donald McAlister's law of error and of Professor Kapteyn's frequency curves lies in the possibility of adapting without much trouble to unsymmetrical phenomena numerous expressions which have been already calculated for the normal law of error and the normal curve of frequency. \({ }^{1}\)

This method of proceeding from arithmetic to geometric laws of error is clearly capable of generalisation. We have dealt with the geometric law which can be derived from the normal arithmetic law. Similarly if we start from the simplest geometric law of error, namely, \(f_{l}=\mathbf{A}\left(\frac{y}{y_{l}}\right)^{l^{2} / y} e^{-k^{2} y}\), we can easily find, by writing \(\log y=l\) and \(\log y_{q}=l_{q}\), the corresponding arithmetic law, namely, \(f_{q}=A e^{k^{2} e l\left(l-l_{q}\right)-l^{2} e l}\), which is obtained from the generalised arithmetic law by putting \(\phi(l)=l^{2} \cdot e^{\prime}\) and \(\psi\left(l_{l}\right)=0\). And, in general, corresponding to the arithmetic law
we have the geometric law
where
\[
y=\log z, y_{\|}=\log z_{\|}, \int \frac{\phi_{1}{ }^{\prime}(z)}{z} d z=\phi(\log z) \text { and } \psi_{1}\left(z_{i}\right)=\psi\left(\log z_{\|}\right) .
\]
(iii.) What law of error does the harmonic mean imply ?

In this case, \(\Sigma \frac{f_{\eta}^{\prime}}{f_{\eta}}=0\) must be equivalent to \(\Sigma\left(\frac{1}{y_{q}}-\frac{1}{y}\right)=0\).
 A simple form of this is obtained by puting \(\phi^{\prime}(y)=-k^{2} y^{2} y^{2}\) and
 positive and negative crrors of the same absolute magnitude are not equally likely.
(iv.) If the most probable value of the quantity i.s equal to the median of the measurements, what is the law of error ?

The median is usually defined as the measurement which

\footnotetext{
\({ }^{1}\) It may be added that Professor Kapteyn's monograph brings forward considerations which would be extremely valuable in determining the types of phenomena to which geometric laws of error are likely to be applicable.
}
oceupies the midde pmsition when the measuremmen are ramend in order of magnitude. If the number of measurements as is and the most probahla value ot the quantity is the" \("\). \({ }^{1}\) tha and, if the number is even, all values between the \({ }_{2}^{\prime \prime}\) thand the \(\left.\right|_{2} ^{\prime \prime \prime} 1\) thath equally probable amongst themselves and more probable than any onther. For the present purpose, however. it is mecessary to make use of another property of the median, which was known to Fechner (who first introduced the median into use) but which seldom receives as much attention as it deserves. If \(y\) is the mediun of a number of magnitudes, the sum of the absolute differ rowes (i.e. The diffremee aluays reckomed positive) betusen y and worh of the magnitudes is a minimum. The median \(y\) of \(y_{1} y_{2} \ldots y_{\ldots,}\) is found, that is to say, by making \(\sum y_{q}-y^{\prime}\) a minimum where \(y_{11}-y\) is the difference always reckoned positive between \(y\) and \(y\).

We can now return to the investigation of the law of error corresponding to the median.
\[
\text { Wite } y-y \text { © Then simee }- \text { : is to lin a minime:n w. }
\]
must have \(\sum^{y-y_{\|}}=0\). Whence, proceeding as before, we have \(\varepsilon_{1}\)
\[
f \text { ir: } 1
\]

The simplest case of this is obtained by putting
\[
\begin{aligned}
& \phi^{\prime \prime}(y)-l_{i}{ }^{2} \text {. }
\end{aligned}
\]
whene.
\[
i A=A
\]

This satisfies the additional condition that positive and negative errors of equal magnitude are equally likely. Thus in this impertant respert the median is as satifantory an the arithmelw mean, and the law of error which leads to it is as simple. It also resembles the normal law in that it is a function of the crror only. and not of the magnitude of the measurement as well.

The median lan of ermer. if er where mato atmontere amount of the error always reckoned positive is of some historical
interest, because it was the earliest law of error to be formulated. The first attempt to bring the doctrine of averages into definite relation with the theory of probability and with laws of error was published by Laplace in 1774 in a memoir "sur la probabilité des causes par les événemens." \({ }^{1}\) This memoir was not subsequently incorporated in his Théorie anulylique, and does not represent his more mature view. In the Théorie he drops altogether the law tentatively adopted in the memoir, and lays down the main lines of investigation for the next hundred years by the introduction of the normal law of error. The popularity of the normal law, with the arithmetic mean and the method of least squares as its corollaries, has been very largely due to its overwhelming advantages, in comparison with all other laws of error, for the purposes of mathematical development and manipulation. And in addition to these technical advantages, it is probably applicable as a first approximation to a larger and more manageable group of phenomena than any other single law. So powerful a hold indeed did the normal law obtain on the minds of statisticians, that until quite recent times only a few pioneers have seriously considered the possibility of preferring in certain circumstances other means to the arithmetic and other laws of error to the normal. Laplace's earlier memoir fell, therefore, out of remembrance. But it remains interesting, if only for the fact that a law of error there makes its appearance for the first time.

Laplace sets himself the problem in a somewhat simplified form: "Déterminer le milieu que l'on doit prendre entre trois observations données d'un même phénomène." He begins by assuming a law of error \(z=\phi(y)\), where \(z\) is the probability of an error \(y\); and finally, by means of a number of somewhat arbitrary assumptions, arrives at the result \(\phi(z)={ }_{2}^{m_{2}} e^{-\cdots, n}\). If this formula is to follow from his arguments, \(y\) must denote the absolute error. always taken positive. It is not unlikely that Laplace was led to this result by considerations other than those by which he attempts to justify it.

Laplace, however. did not notice that his law of error led to the median. For, instead of finding the most probable value, which would have led him straight to it, he seeks the " mean of error "- the value, that is to say, which the true value is as likely
to fall short of as to exceed. This value is, for the median law, laborions 1. time and ankward in the reault. Laplace works it out correctly for the case where the olservations are no more than three.
6. I don mon think that it is persible to find by this methent a haw of ermer which leads to the mode. But the following eremeral formulae are easily obtained:
(v.) If ご \(\theta(y . y\).! - - 10 is the law of relation between the measurn monts and the most probable value of the quantity, then the law
 butwen 1 and 1. \(/ \theta(y, y) \phi^{\prime \prime}(y) d y\) • \(\psi(y)\) - log A munt be negrative for all walues of ! and \(y\) that are physically possible : and, since the values of \(y_{\|}\)are between them exhaustive,
where the summation i.s for all terms that can be formed he giving \(y_{1}\) every value à priori possible.
(vi.) The most general form of the law of error, when it is assmmed that peritive and mentive erroms of the same magnitude are equally probable, is \(\mathrm{A} e^{-1.2 f\left(y-y_{l}\right)^{2}}\), where the most probable value of the quantity is given by the equation
 \(f\left(y-y_{H}\right)^{2}=\left(y-y_{4}\right)^{2}\); and the median is a special case obtained by putting \(f\left(y-y_{4}\right)^{2}=+\sqrt{ }\left(y-y_{4}\right)^{2}\).

We can obtain other special cases by putting
\[
f(y \quad y)=(y-y)^{\prime} \text {. }
\]
when the law of errer is \(A e^{-l^{2}\left(y-y_{i}\right)^{a}}\) and the most probable values

 the most probable values the roots of \(\begin{aligned} & , ~ 1 \\ & !\cdots y\end{aligned}=0\). In all these cases the law is a function of the error only.
7. These results may be summarised thus. We have assumed:
(a) That we have no reasem, before making measurements, for
supposing that the quantity we measure is more likely to have any one of its possible values than any other.
(b) That the errors are independent, in the sense that a knowledge of how great an error has been made in one case does not affect our expectation of the probable magnitude of the error in the next.
(c) That the probability of a measurement of given magnitude, when in addition to the a priori evidence the real value of the quantity is supposed known, is an algebraic function of this given magnitude of the measurement and of the real value of the quantity.
(d) That we may regard the series of possible values as continuous, without sensible error.
(e) That the à priori evidence permits us to assume a law of error of the type specified in (c) ; i.e. that the algebraic function referred to in (c) is known to us \(a\) priori.

Subject to these assumptions, we have reached the following conclusions :
(1) The most general form of the law of error is
leading to the equation \(\Xi \because(y, y)=0\), connecting the most probable value and the actual measurements, where \(y\) is the most probable value and \(y_{q}\), etc., the measurements.
(2) Assuming that positive and negative errors of the same absolute magnitude are equally likely, the most general form is \(f_{11}=\mathrm{A} e^{\left.-l_{2}^{2}(\eta)-\mu_{4}\right)^{2}}\), leading to the equation \(\geq\left(y-y_{4}\right) f^{\prime}\left(y-y_{4}\right)^{2}=0\), where \(f^{\prime} z=\frac{d}{d z} f z\). Of the special cases to which this form gives rise, the most interesting were
(3) \(f_{11}=\mathrm{A} e^{-12^{2}\left(y-\mu_{4}\right)^{2}}=\mathrm{A} e^{-1 \cdot 2^{2} 2^{2}}\), where \(z_{y}=y-y_{11}\), leading to the arithmetic mean of the measurements ats the most probable, value of the quantity ; and
(4) \(f_{q}=\mathrm{A} e^{-k^{k} z_{q}}\), leading to the median.
(5) The most general form leading to the arithmetic mean is \(f_{q}=\mathrm{A} e^{\phi^{( }(y)\left(y-y_{q}\right)-\phi(y)+\psi\left(y_{q}\right)}\), with the special cases (3), and

(7) The most general form leading to the geometric mean is \(f_{n}=\mathrm{A} e^{\phi(y)} \log _{y}^{1 / y_{y}}+\int_{y}^{\phi^{\prime}(y)} d y+\psi\left(y_{n}\right)\), with the special cases:
(8) j. . \(\|_{\|}^{2} \quad=\) and
(9) \(f \quad 1 e^{\cdots}(10)\).
(10) The most armaral form leadime to the harmonic mean is

(11) / A . 1
(12) The most general form leading to the median is
\[
\left.f \quad x^{\phi()^{n-6}++1}\right)
\]
with the special case (4).
In eacin of these expressions, \(f\), is the probability of a measure ment \(y_{\text {, }}\), given that the true value is \(y\).
8. The doctrine of Means and the allied theory of Least fquares comprise on extmsibe a subject mather that they cammet be adequately treated except in a volume primarily devoted to them. As, however, they are one of the important practical applications of the theory of probability. 1 ann unwilling to pans them by entirely; and the following discursive observations, chiefly relating to the Normal Law of Error, will serve, taken in conjunction with the paragraphs immediately preceding, to illustrate the connection between the theories of this treatise and the general treatment of averages.
9. The Claims of the Arithmetic Average.-By definition the arithmetic average of a number of quantities is nothing more than their arithmetic sum divided by their number. But the utility of an average generally consists in our supposed right to substitute, in certain cases, this single measure for the varying measures of which it is a function. Sometimes this requires no justification ; the word "average" is in these cases used for the sake of shortness, and merely to summarise a set of facts: as, for instance, when we say that the birth-rate in Eingland is greater than the birth-rate in France.

But there are wher cases in which the aberage makne a mome substantial claim to add to our knowledge. After a number of examiners of equal capacity have given varying marks to a candidate for the same paper, it may be thought fair to allom the candidate the average of the different marks allonted: and in general if several estimates of a magnitude haw ben made.
between the accuracy of which we have no reason to discriminate, we often think it reasonable to act as if the true magnitude were the average of the several measurements. Perhaps De Witt, in his report on Annuities to the States General in 1671, \({ }^{1}\) was the first to use it scientifically. But as Leibniz points ont: "Our peasants have made use of it for a long time according to their natural mathematics. For example, when some irforitance or land is to be sold, they form three bodies of appraisers ; these bodies are called Schurzen in Low Saxon, and each body makes an estimate of the property in question. Suppose. then, that the first estimates its value to be 1000 crowns, the scound. 1400 , the third, 1500 ; the sum of these three estimates is takem, viz. 3900 , and because they were three bodies, the third, i.e. 1300 , is taken as the mean value asked for. This is the axiom: nequalibus aequalin, equal suppositions must have equal consideration." \({ }^{2}\)

But this is a very inadequate axiom. Equal suppositions would have equal consideration, if the three estimanes had been multiplied together instead of being added. The truth is that at all times the arithmetic mean has had simplicity to recommend it. It is always easier to add than to multiply. But simplicity is a dangerous criterion: "La nature," says Fresnel, " ne s'est pas embarassée des difficultés d'analyse, elle n'a évité que la complication des moyens."

With Laplace and Gauss there began a series of attempts to prove the worth of the arithmetic mean. It was diseovered that its use involved the assumption of a particular type of law of error for the it priori probabilities of given errors. It was also found that the assumption of this law led on to a more complicated rule, known as the Method of Least Siquares, for combining the results of observations which contain more than one doubtful quantity. In spite of a popular belief that, whilst the Arithmetic Mean is intuitively obvious, the Method of Least Squares depends upon doubtful and arbitrary assumptions, it can be demonstrated that the two stand and fall together. \({ }^{3}\)

\footnotetext{
\({ }^{1}\) De rardye van de lif-renten na proportie van de lostenten. The Hague, 1671.
2 Noureaux Essais. Eingl. transl. p. 54).
\({ }^{3}\) Venn (Logic of Chance, p. 40) thinks that the Normal Law of Error and the Method of Least Squares "are not only totally distinct things, but they have scarcely even any necessary connection with each other. The Law of Error is the statement of a physical fact. . . . The Method of Least Squares, on the other hand, is not a law at all in the scientifie sense of the term. It is simply a rule or direction. . . ."
}

The analytical theorems of Laplace and Gauss are complicated, but the special assumptions upon which they are based are easily. stated. \({ }^{1}\) Gauss supposes (a) that the probability of a given error is a function of the error only and not also of the magnitude of the observation. (b) that the errors are so small that their cubes and higher powers may be neglected. Assumption (a) is arbitrary \({ }^{2}\) and lauss did not state it explicitly. These two assumptions, together with certain others, lead us to the result. For let \(\phi(z)\) be the law of error where \(z\) is the error, and let us assume. as it always is assumed in these proofs, that \(\phi(z)\) can be expranded
 - 3
, \(\phi^{\prime \prime \prime}(0)+\ldots\) It is also supposed that positive and negative .).
errors are equally protable, i.e. \(\phi(z)=\phi(-z)\), so that \(\phi^{\prime}(0)\) and \(\phi^{\prime \prime \prime}(0)\) vanish. Since we may neglect \(z^{4}\) in comparison with \(z^{2}\),



Gauss's proof looks much more complicated than this, hut he
obtains the form ae " by neglecting higher powers of \(z\), so that this expression is really equivalent to \(a: b z^{2}\). By this approximation he has reduced all the possible laws to an equivalent form. \({ }^{3}\) It is true, therefore, that the normal law of error is, to the second power of the error, equivalent to any law of error.
 negative errors are equally probable. Laplace also introduces assumptions equivalent to these.

While mathematicians have endeavoured to establish the normal law of error and the arithmetio mean as a law of logic.

\footnotetext{
\({ }^{1}\) For an account of the three primipal methods of arriving at the Methex
 first methend is in the Theoria Molus, and his seeond in the 'Throria ('ombinafionis Obsermationum. Laplace's investigations are in chap. iv. of the second Book of the Therrie analytique. Laplace's method was improved by Poisson

 rechnung, p. 299), that, because a larger error is less probable than a smaller, therefore the probability of a given error is a function of its magnitude unly.

}
others have claimed for it the testimony of experience and have deemed it a law of nature. \({ }^{1}\)

That this cannot be so, is evident. For suppose that \(x_{1} r_{2} \ldots r_{1}\) are a set of observations of an unknown quantity \(x\). Then, by this principle, \(x={ }_{n}^{1} \leq x\), gives the most probable value of \(x\). But suppose we had wished to determine \(x^{2}\), our observations, ass.uming that we can multiply correctly, would be \(x_{1}{ }^{2}, x_{2}{ }^{2} \ldots x_{n}{ }^{2}\),
 And in general, \({ }_{n}^{1} \unrhd f\left(x_{i}\right) \neq f\left({ }_{n}^{1} \leq x_{r}\right)\). Nor is this a consideration which can safely be ignored in practice. For our "olsservations" are often the result of some manipulation, and the particular shape in which we get them is not necessarily fixed for us. It is not easy to say what the direct observation is. In particular if any such law of sensation, as that enunciated by Fechner, is true (i.e. that sensation varies as the logarithm of the stimulus), the arithmetic mean must break down as a practical rule in all cases where human sensation is part of the instrument by means of which the observations are recorded. \({ }^{2}\)

Apart, however, from theoretical refutations, statisticians now recognise that the arithmetic mean and the normal law of error can only be applied to certain special classes of phenomena. Quetelet \({ }^{3}\) was, I think, the first to point this out. In Eingland, Galton drew attention to the fact many years ago, and Professor Pearson + has shown "that the Gaussian-Laplace normal distribution is very far from being a general law of frequency distribution either for error's of observation or for the distribution of deviations from type such as occur in organic populations. . . . It is not even approximately correct, for example, in the distribution of baronetric variations, of grades of fertility and incidence of disease."

\footnotetext{
1 This is, of course, a very common point of view indeed. Cf. Bertrand, op. cit. p. 183: "Malgré les objections précédentes, la formule de Gauss doit être adoptée. L'observation la confirme: cela doit suffire dans les applications."
\({ }^{2}\) This was noticed by Galton.
\({ }^{3}\) E.g. Letters on the Theory of Probabilities, p. 114.
\({ }^{4}\) On "Errors of Judgment, etc.," Phil. Trans. A, vol. cxcviii. pp. 235-299. The following quotation is from his memoir On the General Theory of Skew Correlation and Nonlinear Regression, where further references are given.
}

The Arithmetic Mean occupies, therefore, no unique position: and it is worth while. from the point of view of probatility. (1) discuss the properties of other persible means and laws of error. as. for example, on the lines indicated in the carlier part of this chapter.
10. The Methent of Lemes sigumes. The probleme to which this methond is applied, is no more than the application of the same 'onsiderations, as those which we hatw just been disenssinge to cases where the relation between the observed measurements and the quantity whose most probable value we require, involves more than one unknown.

Owing to the surprising character of its conclusions, if they could be accepted as universally valid, and to the obscurity of the mathematical fabric that has been reared on and about it, this method has been surrounded by an unnecessary air of mystery. It is true that in recent times scepticism has grown at the expense of mystery. It is also true that just views have been held by individuals for sixty years past, notably by Leslie Ellis. But the old mistakes are not always corrected in the current text-books, and even so useful and generally used a treatise on Least Squares, as Professor Mansfield Merriman's. opens with a series of very fallafious statements.

The controversial side of the Method of Least Squares is purely logical ; in the later developments there is much elaborate mathematics of whose correctness no one is in doubt. What it is important to state with the utmost possible clearness is the precise assumptions on which the mathematics is based; when these assumptions have been set forth, it remains to determine their applicability in particular cases.

In dealing with averages we supposed ourselves to be presented with a number of direct observations of some quantity which it is desired to determine. But it is ohvious that direct observations will be in many cases either impracticable or inconvenient ; and our natural course will be to measure certain other quantities which we know to bear fixed and invariable relations to the unknowns we wish to determine. In surveying, for instance, or in astronomy, we constantly prefer to take measurements of angles or distances in which we are not interested for their own sakes, but which bear known geometrical relationships to the set of ultimate unknowns.

If we wish to determine the most probable values of a set of unknowns \(x_{1}, x_{2}, x_{3} \ldots x_{i}\), instead of obtaining a number of sets of direct observations of each, we may obtain a number of equations of observation of the following type:
\[
\begin{gathered}
a_{1} x_{1}+a_{2} x_{2}+\ldots+a_{r} x_{r}=\mathrm{V}_{1} \\
b_{1} x_{1}+b_{2} x_{2}+\ldots+b_{r} x_{r}=\mathrm{V}_{2} \\
k_{1} x_{1}+\dot{k}_{2} x_{2}+\ldots+\dot{k}_{r} x_{r}=\mathrm{V}_{n}
\end{gathered}
\]
where \(V_{1}\), etc., are the quantities directly observed, and the \(a\) 's, \(b\) 's, etc., are supposed known ( \(n>r\) ).

We have in such a case \(n\) equations to determine \(r\) uiknowns, and since the observations are likely to be inexact, there may be no precise solution whatever. In these circumstances we wish to know the most probable set of values of the \(x\) 's warranted by these observations.

The problem is precisely similar in kind to that dealt with by averages and differs only in the degree of its complexity. It is the problem of finding the most probable solution of such a set of discrepant equations of observation that the Method of Least Squares claims to solve.

By 1750 the astronomers were obtaining such equations of observation in the course of their investigations, and the question arose as to the proper manner of their solution. Boscovich in Italy, Mayer and Lambert in Germany, Laplace in France, Euler in Russia, and Simpson in England proposed different methods of solution. Simpson, in 1757, was the first to introduce, by way of simplification, the assumption or axiom that positive and negative errors are equally probable. \({ }^{1}\) The Method of Least Squares was first definitely stated by Legendre in 1805, who proposed it as an advantageous method of adjusting observations. This was soon followed by the 'proofs' of Laplace and Ciauss. But it is easily shown that these proofs involve the normal law of error \(y=k e^{-12, k}\), and the theory of Least Squares simply develops the mathematical results of applying to equations of observation, which involve more than one unknown, that law

\footnotetext{
\({ }^{1}\) See Merriman's Method of Least Squares, p. 181, for an historical sketch, from which the above is taken. In 1877 Merriman published in the Transactions of the Connecticut Academy a list of writings rolating to the Method of Least Squares and the theory of accidental errors of observation, which comprised 408 titles-classified as 313 memoirs, 72 books, 23 parts of books.
}
of error which leads to the A rithmetic Mran in the case of a single unknown.
11. The Weighting of Averages.-It is necessary to recur to the distinction made at the hegimming of \(\$ 9\) hetween the two types to which our aremage. or. as it is gemerally termed in sombal inquiries, our index number, may belong. The aserage or index number may simply summarise a set of facts and give us the actual value of a composite quantity, as, for example, the index number of the cost of living. In such cases the composite quantity, in which we are interested, need not comtain precis.ly the same number of units of each of the elementary quantitios of which it is compuried, so that the "weights." which demote the numbers of cadn dementary quantity apprepriate to the composite quantity, are part of the definition of the composite quantit! and can mo bere the diapensed whith than the magnitudes of the elementary quantities themselves. Nor in such cases is the rejomion of dismentant observations permissible: if, that is to say, some of the elementary quantities are subject to much wider variation, or to variations of a different type than the majority, that is no reason for rejecting them.

On the other hand, the individual items, out of which the average is composed, may each be indications or approximate estimates of some one single quantity; and the average, instead of representing the measure of a composite quantity, may be selected as furnishing the most probable value of the single quantity, given, as evidence of its magnitude, the values of the various terms which make up the average.

If this is the character of our average, the problem of weighting depends upon what we know about the individual observations or samples or indications, out of which our average is to be built up. The units in question may be known to differ in respects relevant to the probable value of the quaesitum. Thus there may be reasons, quite apart from the actual results of the individual observations or samples, for trusting some of them more than others. Our knowledge may indicate to us, in fact, that the constants of the laws of error appropriate to the several instances, even if the type of the law can be assumed to be constant shmuld the variod according to the data we pmasess about each. It may also indicate to us that the condition of independence between the instances, which the method of averages
presumes, is imperfectly satisfied, and consequently that our mode of combining the instances in an average must be inodified accordingly.

Some modern statisticians, who, really influenced perhaps by practical considerations, have been inclined to deprecate the importance of weighting on theoretical grounds, have not always been quite clear what kind of average they supposed themselves to be dealing with. In particular, discussions of the question of weighting in connection with index numbers of the value of money have suffered from this confusion. It has not been clear whether such index numbers really represent measures of a composite quantity or whether they are probable estimates of the value of a single quantity formed by combining a number of indejendent approximations towards the value of this quantity. The original Jevonian conception of an index number of the value of money was decidedly of the latter type. Modern work on the subject has been increasingly dominated by the other eonception. A discussion of where the truth lies would lead me too far into the field of a subject-matter alien to that of this treatise.

Theoretical arguments against weighting have sometimes been based on the fact that to weight the items of the average in an irrelevant manner, or, as it is generally expressed, in a random manner, is not likely, provided the variations between the weights are small compared with the variations between the items, to affect the result very much. But why should any one wish to weight an average " at random" ? Such observations overlook the real meaning and significance of weights. They are probably inspired by the fact that a superficial treatment of statistic's would sometimes lead to the introduction of weights which are irrelevant. In drawing a conclusion, for example, from the vital statistics of various towns, the figures of population for the different towns may or may not be relevant to our conclusion. It depends on the character of the argument. If they are relevant, it may be right to employ them as weights. If they are irrelevant, it must be wrong and unnecessary to do so. The fact that wheat is a more important article of consumption than pins may, on certain assumptions, be irrelevant to the usefulness of variations in the price of each article as indications of variation in the value of money. With other assumptions, it may be
extremely relevant. Or aqain, we max know that ohservations with a particular instrument tend to be too large and must, therefore bee weighted down. It is contrary both to theory and to common semse to supperse that the posisession of information as to the relative reliabilite of different statistios in mot usioful. There is no place. therefore in my judguent. for at gromatistal argument as to the propriety or improprione of weichting an average.

It should be added that, where we seek to build up an index number of a conception, which is quantitative but is not itself numerically measurable in any defined or mabatigums sense. by combining a number of numerical quantitios. Which. white thes do not measure our quaesitum are nevertheless indications of its quantitative variations and tend to fluctuate in the same sense, as, for example, by means of what are sometimes called economic barometers of the state of business, or the prosperity of the country or the like, some very confusing questions can arise both as to what sort of a thing our resulting index really is, and as to the mode of compilation appropriate to it.

These confusing questions always arise when, instead of measuring a quantity directly, we seek an index to fluctuations in its magnitude by combining in an average the fluctuations of a series of magnitudes, which are, each of them in a different way, to some extent (but only to some extent), correlated with fluctuations in our quaesitum. I must not burden this book with a discussion of the problems of Index Numbers. But I venture to think that they would be sooner cleared up if the natures and purposes of differing index numbers were more sharply distinguished those, namely, which are simply descriptive of acomposite commodity, those which seek to combine results differing from one another in a way analogous to the variations of an instrument of precision, and those which combine results, not of the quarsitum. itself, but of various other quantities, variations in which are partly due to variations in the quassitum. but which we well know to be also due to other distinguishable influences. Index numbers of the third type are often treated by methods and arguments only appropriate to those of the second type.
12. The Rejection of Discordant Observations. -This differs from the problem just discussed, hecause we have supposed so far that our system of weighting is determined by data which we
possess prior to and apart from our knowledge of the actual magnitude of the items of our average. The principle of the rejection of discordant observations cones in when it is argued that, if one or more of our observations show great discrepancies from the results of the greater number, these ought to be partly or entirely neglected in striking the average, even if there is no reason, except their discrepancy from the rest, for attributing less weight to them than to the others. By some this practice has been thought to be in accordance with the dictates of common sense ; by others it is denounced as savouring even of forgery. \({ }^{1}\)

This controversy, like so many others in Probability, is due to a failure to understand the meaning of 'independence.' The mathematics of the orthodox theory of Averages and Least Squares depend, as we have seen, upon the assumption that the observations are 'independent'; but this has sometimes been interpreted to mean a physical independence. In point of fact, the theory requires that the observations shall be independent, in the sense that a knowledge of the result of some dnes not affect the probability that the others, when known, involve given errors.

Clearly there may be initial data in relation to which this supposition is entirely or approximately accurate. But in many cases the assumption will be inadmissible. A knowledge of the results of a number of observations may lead us to modify our opinion as to the relative reliabilities of others.

The question, whether or not discordant observations should be specially' weighted down, turns, therefore, upon the nature of the preliminary data by which we have been guided in initially adopting a particular law of error as appropriate to the observations. If the observations are, relevant to these data, strictly ' independent,' in the sense required for probability, then rejection is not permissible. But if this condition is not fulfilled, a hias against discordant observations may be well justified.

\footnotetext{
\({ }^{1}\) E.g. G. Hagen's Grundzüge der Wahrscheinlichkeitsrechnung, p. 63: " Die Täuschung, die man durch Verschweigen von Messungen begeht, lässt sich eben so wenig entschuldigen, als wenn man Messungen fälschen oder fingiren wollte."
}
\[
\begin{gathered}
\text { P.Ali'T } 111 \\
\text { INDUCTION AND ANALOGY }
\end{gathered}
\]

\title{
CHAPTER XVIII
}

\section*{INTRODUCTION}

Nothing so like as egge ; yet no one, on acoount of this apparont similarity, expects the same taste and relish in all of them. "Tis only after a long course of uniform experiments in any kind, that we attain a firm reliance and security with regard to a particular event. Now where is that process of reasoning, which from one instance draws a conclusion, so different from that which it infers from a hundred instances, that are no way different from that single instance? This question I propose as much for the sake of information, as with any intention of raising difficulties. I camont find, I cannot imagine any such reasoning. But I keep my mind still open to instruction, if any one will

1. I have described Probability as comprising that part of logic which deals with arguments which are rational but not conclusive. By far the most important types of such arguments are those which are based on the methods of Induction and Analory. Almost all empirical science rests on these. And the decisions dietated by experience in the ordinary conduct of lif. generally depend on them. To the analysis and logical justification of these methods the following chapters are directed.

Inductive processes have formed, of course, at all times a vital, habitual part of the mind's machinery. Whenever we learn by "xperience, we are using them. But in the logie of the schools they have taken their proper place slowly. No clear or satisfactory account of them is to be found anywhere. Within and yet beyond the scope of formal logice, on the line, apparently. between mental and natural philosophys. Induction has been admitted into the organon of scientific proof, without much help) from the logicians, no one quite knows when.
2. What are its distimuishine characteristics? What are the qualities which in ordinary discourse seem to afford atrength to an inductive argument?

\footnotetext{

}

I shall try to answer these questions before I proceed to the more fundamental problem-What ground have we for regarding such arguments as rational ?

Let the reader remember, therefore, that in the first of the succeeding chapters my main purpose is no more than to state in precise language what elements are commonly regarded as adding weight to an empirical or inductive argument. This requires some patience and a good deal of definition and special terminology. But I do not think that the work is controversial. At any rate, I am satisfied myself that the analysis of Chapter XIX. is fairly adequate.

In the next section, Chapters XX. and XXI., I continue in part the same task, but also try to elucidate what sort of assumptions, if we could adopt them, lie behind and are required by the methods just analysed. In Chapter XXII. the nature of these assumptions is discussed further, and their possible justification is debated.
3. The passage quoted from Hume at the head of this chapter is a good introduction to our subject. Nothing so like as eggs, and after a long course of uniform experiments we can expect with a firm reliance and security the same taste and relish in all of them. The eggs must be like eggs, and we must have tasted many of them. This argument is based partly upon Analogy and partly upon what may be termed Pure Induction. We argue from Analogy in so far as we depend upon the likeness of the eggs, and from Pure Induction when we trust the number of the experiments.

It will be useful to call arguments inductive which depend in any way on the methods of Analogy and Pure Induction. But Ido not mean to sugerst by the use of the term inductive that these methods are necessarily confined to the objects of phenomenal experience and to what are sometimes called empirical questions ; or to preclude from the outset the possibility of their use in abstract and metaphysical inquiriss. While the term inductive will be employed in this general sense, the expression Pure Induction must be kept for that part of the argument which arises out of the repetition of instances.
4. Hume's account, however, is incomplete. His argument could have been improved. His experiments should not have been too uniform, and nught to have differed from one another
as much as porsible in all respects sate that of the likemess of the eges. He should have tried ewes in the town and in the comeres. in January and in June. He might then haw diseovered that eggs could he good or had, however like they looked.

This principle of varying those of the characteristics of the instances. which wer erard in the comditions of cour generalisation as non-essential, may be termed Negative Analogy.

It will be argued later on that an increase in the number of "xperments is only valualle in so far as, by increasime or pessibly increasing, the variety found amonest the non-essential characteristics of the instances, it strenethens the Negative Analogy If Hume's experiments had been absolutely uniform, he would have been right to raise doubts about the conclusion. There is no process of reasoning. which from one instance draws a conclusion different from that which it infers from a hundred instances, if the latter are known to be in no way diferent from the former. Hume has unconsciously misrepresented the typical inductive argument.

When our control of the experiments is fairly complete, and the conditions in which they take place are well known, there is not much room for assistance from Pure Induction. If the Negative Analogies are known, there is no need to count the instances. But where our control is incomplete, and we do not know accurately in what ways the instances differ from one another, then an increase in the mere number of the instances helps the argument. For unless we know for certain that the instances are perfectly uniform, each new instance may possibly add to the Negative Analogy.

Hume might also have weakened his argument. He expects no more than the same taste and relish from his egus. He attempts no conclusion as to whether his stomach will always draw from them the same nemishment. He has conserved the force of his generalisation by kepping it narrow.
5. In an inductive argument, therefore, we start with a number of instances similar in some respects AB, dissimilar in others ( \({ }^{\text {. We pack out one or more respects } A \text { in which the }}\) instances are similar, and argue that some of the other respects B in which they are also similar are likely to be associated with the characteristies A in other unexamined cases. The more comprehensive the essential characteristics A, the ereater the
variety amongst the non-essential characteristics C , and the less comprehensive the characteristics \(B\) which we seck to associate with A , the stronger is the likelihood or probability of the generalisation we seek to establish.

These are the three ultimate logical elements on which the probability of an empirical argument depends,- the Positive and the Negative Analogies and the scope of the generalisation.
6. Amongst the generalisations arising out of empirical argument we can distinguish two separate types. The first of these may be termed universal induction. Although such inductions are themselves susceptible of any degree of probability, they affirm invariable relations. The generalisations which they assert, that is to say, claim universality, and are upset if a single exception to them can be discovered. Only in the more exact sciences, however, do we aim at establishing universal inductions. In the majority of cases we are content with that other kind of induction which leads \(u_{p}\), to laws upen which we can generally depend, but which does not claim, however adequately established, to assert a law of more than probable connection. \({ }^{1}\) This second type may be termed Inductire Correlation. If, for instance, we base upon the data, that this and that and those swans are white, the conclusion that all swans are white, we are endeavouring to establish a universal induction. But if we base upon the data that this and those swans are white and that swan is black, the conclusion that most swans are white, or that the probability of a swan's being white is such and such, then we are establishing an inductive correlation.

Of these two types, the former-miversa! induction- presents both the simpler and the more fundammital problem. In this part of my treatise I shall confine myself to it ahmost entirely. In Part V., on the Foundations of Statistical himerence, I shall discuss, so far as I can, the logical basis of inductive correlation.
7. The fundamental connection between Inductive liethod and Probability deserves all the emphasis I can sive it. Many writers, it is true, have recognised that the conclusions which we reach by inductive argument are probable and inconclusive. Jevons, for instance, endeavoured to justify inductive procerses by means of the principles of inverse probability. And it is true also that much of the work of Laplace and his followers was

\footnotetext{
\({ }^{1}\) What Mill calls 'approximate generalisations.'
}
directal to the solution of assentially inductive problems. But it has been seldum appehmoded dearly, wither be these writers or by others. that the validity of evere induction, strictly interpreted, depends, mot on a matere of fact, hat on the existence of a mation of probability: An inductive aroument aftims. not that a certain matter of fact is so, but that relative to certain eridence theme is a probability in its favour. The validity of the induetion. relative to the original evidence, is not upset, therefore. if, as a fact, the truth turns out to be otherwise.

The clear apprehension of this truth profoundly modifies our attitude towards the solution of the inductive problem. The validity of the inductive mothod domes mot depend on the success of ite perdictions. Its repated failure in the past may. of couree. supply us with m.W evidence, the indurion of whel will medio. the force of subsequent inductions. But the force of the old induction retulier to the whd aridenese is untouched. The everdemere with which our experimen hats supplied us in the past may have proved misleading, but this is entirely irrelevant to the question of what conclusion we ought reasonably to have drawn from the evidence then before us. The validity and reasonable nature of inductive generalisation is, therefore, a question of logic and not of experience, of formal and not of material laws. The actual constitution of the phenomenal universe determines the character of our evidence ; but it camot determine what conclusions given evidence rationally supports.

\section*{CHAPTER XIX}

\section*{THE NATURE OF ARGUMENT BY ANALOGY}

All kinds of reasoning from causes or effects are founded on two particulars, viz. the constant conjunction of any two objects in all past experience, and the resemblance of a present object to any of them. Without some degree of resemblance, as well as union, 'tis impossible there can be any reasoning.Hume. \({ }^{1}\)
1. Hume rightly maintains that some degree of resemblance must always exist between the various instances upon which a generalisation is based. For they must have this, at least, in common, that they are instances of the proposition which generalises them. Some element of analogy must, therefore, lie at the base of every inductive argument. In this chapter I shall try to explain with precision the meaning of Analogy, and to analyse the reasons, for which, rightly or wrongly, we usually regard analogies as strong or weak, without considering at present whether it is possible to find a good reason for our instinctive principle that likeness breeds the expectation of likeness.
2. There are a few technical terms to be defined. We mean by a generalisation a statement that all of a certain definable class of propositions are true. It is convenient to specify this class in the following way. If \(f(x)\) is true for all those values of \(x\) for which \(\phi(x)\) is true, then we have a generalisation about \(\phi\) and \(f\) which we may write \(g(\phi, f)\). If, for example, we are dealing with the generalisation, " All swans are white," this is equivalent to the statement, " " \(x\) is white" is true for all those values of \(x\) for which ' \(x\) is a swan' is true." The proposition \(\phi(a) \cdot f(a)\) is an instance of the generalisation \(g(\phi, f)\).

By thus defining a generalisation in terms of propositional functions. it becomes possible to deal with all kinds of generalisa-
tions in a miform way: and also (10 bringe emenalisation inm convenient comnection with our definition of Analogy.

If some one thing is true about both of two objects, if, that is to saty, they both satisfy the same propesitional function, then to this extent there is an cmalugy between them. Erery generalisation \(g\) ob \(f\) ). therefore, asserts that one andoly is always acem paniod be another, namely, that between all injocts havime the analoy! \(\phi\) thene is alow the analoy \(f\). The set of propesitional functions, which are satisfied by both of the two objects, constitute the positice analogy. The analogies, which would be
 analogy; those which are relative to partial knowledge, the known positive analogy.

As the positive analogy measures the resemblances, so the negrative analone measures the differencesthe twen the two abjects. The set of functions, such that each is satisfied by one and not by the other of the objacts, constitutes the negative analogy. We have, as before, the distinction between the total negatice analogy and the known negative analogy.

This set of definitions is soon extended to the cases in which the number of instances exceeds two. The functions which are true of all of the instances constitute the positive analogy of the set of instances, and those which are true of some only, and are false of others, constitute the neqative analogy. It is clear that a function, which represents positive analogy for a group of instances taken out of the set, may be a negative analogy for the set as a whole. Analogies of this kind, which are positive for a sub-class of the instances, but negative for the whole class, we may term sub-unulogies. By this it is meant that there are resemblances which are common to some of the instances, but not to all.

A simple notation, in accordance with these definitions, will be useful. If there is a positive analogy \(\phi\) between a set of instances \(a_{1} \ldots a_{n}\), whether or not this is the total analogy between them. lif us whte this-
\[
A(\phi) .^{1}
\]

And if there is a negative analogy \(\phi^{\prime}\), let us write this-
\[
\bar{A}\left(\phi^{\prime}\right) .^{1}
\]

Thus A \((\phi)\) expresses the fact that there is a set of \(n_{1} \ldots{ }^{\prime}{ }^{\prime}\) characteristics \(\phi\) which are common to all the instances, and \(\bar{A}\left(\phi^{\prime}\right)\) that there is a set of characteristics \(\phi^{\prime}\) which is "..."" true of at least one of the instances and false of at least one.
3. In the typical argument from analogy we wish to generalise from one part to another of the total analogy which experience has shown to exist between certain selected instances. In all the cases where one characteristic \(\phi\) has been found to exist, another characteristic \(f\) has been found to be associated with it. We argue from this that any instance, which is known to share the first analogy \(\phi\), is likely to share also the second analogy \(f\). We have found in certain cases, that is to say, that both \(\phi\) and \(f\) are true of them ; and we wish to assert \(f\) as true of other cases in which we have only observed \(\phi\). We seek to establish the generalisation \(g(\phi, f)\), on the ground that \(\phi\) and \(f\) constitute between them an observed positive analogy in a given set of experiences.

But while the argument is of this character, the grounds, upon which we attribute more or less weight to it, are often rather complex ; and we must discuss them, therefore, in a systematic manner.
4. According to the view suggested in the last chapter, the value of such an argument depends partly upon the nature of the conclusion which we seek to draw, partly upon the evidence which supports it. If Hume had expected the sime degree of nourishment as well as the same taste and relish from all of the eggs, he would have drawn a conclusion of weaker probability. Let us consider, then, this dependence of the probability upon the scope of the generalisation \(g(\phi, f)\), -upon the comprehensiveness, that is to say, of the condition \(\phi\) and the conclusion \(f\) respectively.

The more comprehensive the condition \(\phi\) and the less comprehensive the conclusion \(f\), the greater \(\grave{a}\) priori probability do we attribute to the gencralisation \(g\). With every increase in \(\phi\) this probability increases, and with every increase in \(f\) it will diminish.

The condition \(b\left(b_{1} \phi_{2}\right)\) in men comprelnensive than the condition \(\oint_{1}\). relative te the meneral evidence he if \(\phi_{2}\) is a comdition indepembent of \(\phi_{1}\) relutive th h. \(\phi_{2}\) twing imbumbent of \(\phi_{1}\), if \(g\left(\phi_{1}, \phi_{2}\right) h \cdot 1\), i. if, relative to \(h\), the satisfaction of \(\phi_{2}\) is not inferrible from that of \(\phi_{1}\).

Similarly the conclusion \(f\left(: f_{1} f_{2}\right)\) is more comprehemsibe than the conclusion \(f_{1}\), relative th the eneral evidumer \(h\), if \(f_{2}\) is a conclusion independent of \(f_{1}\), relative to \(h\), i.e. if \(g\left(f_{1}, f_{2}\right) / h \neq 1\).

If \(\delta=\delta_{1} \phi_{2}\) and \(f f_{1} f_{2}\), where \(\phi_{1}\) and \(\phi_{2}\) are inderndent and

\[
\begin{aligned}
& \because\left(\phi_{1}, i\right) \neq\left(\phi_{1} \phi_{2},{ }^{\prime}\right) \cdot \because\left(\phi_{1} \phi_{2},!\right) \\
& =\left(b, I^{\prime}\right) \cdot h, \\
& \cdots(b, 1)=u(b, 1 / 2) \\
& \text { ( } \left.()_{1} \cdot(2) \cdot(\phi), i_{1}\right) / i \\
& \left.=(\phi .,)^{1}\right) \text {, }
\end{aligned}
\]
and
- 1 hat
\[
\cdots\left(\phi, 1_{1}\right) / h: \%(\phi, 1), \cdots\left(\phi_{1}, 1 / k .\right.
\]

This proves the statement made above. It will be noticed that we camot necessarily compare the à priori probabilities of two eremeralistions in respere of mome and less. unless the condition of the first is included in the condition of the second, and the conclusion of the second is included in that of the first.

We see, therefore, that some generalisations stand initially in a stronger position than others. In order to attain a given degree of probability, generalisations require, according to their scope, different amounts of favourable evidenee to support them.
5. Let us now pass from the character of the generalisation a priori to the evidmee by which we support it. Since. whenever the conclusion \(f\) is complex, i.e. resolvable into the form \(f_{1} f_{2}\) where g( \(\left.f_{1}, f_{0}\right) h=1\), we can "xpress the protmbitity of the generalisation \(g(\phi, f)\) as the product of the probabilities of the
 follows, that the condurion \(f\) is simple and mot caphthe of further analysis, without diminishing the generality of our argument.

We will begin with the simplest case, namely, that which arises in the following comditions. First, let us assume that our knowledge of the examined instancers is complete, so that we know of every statement, which is about the examined instaness, whether it is true or false of each. \({ }^{1}\) Second, let us assume that


all the instances which are known to satisfy the condition \(\phi\), are also known to satisfy the conclusion \(f\) of the generalisation. And third let us assume that there is nothing which is true of all the examined instances and yet not included either in \(\phi\) or in \(f\), i.e that the positive analogy between the instances is exactly co-extensive with the analogy \(\phi f\) which is covered by the generalisation.

Such evidence as this constitutes what we may term a perfect analogy. The argument in favour of the generalisation cannot be further improved by a knowledge of additional instances. Since the positive analogy between the instances is exactly coextensive with the analogy covered by the generalisation, and since our knowledge of the examined instances is complete, there is no need to take account of the negative analogy.

An analogy of this kind, however, is not likely to have much practical utility ; for if the analogy covered by the generalisation, covers the whole of the positive analogy between the instances it is difficult to see to what other instances the generalisation can be applicable. Any instance, about which everything is true which is true of all of a set of instances, must be identical with one of them. Indeed, an argument from perfect analogy can only have practical utility, if, as will be argued later on, there are some distinctions between instances which are irrel cout for the purposes of analogy, and if, in a perfect analogy, the positive analogy, of which we must take account, need cover only those distinctions which are relevant. In this case a generalisation based on perfect analogy might cover instances numerically distinct from those of the original set.

The law of the Uniformity of Nature appears to me to amount to an assertion that an analogy which is perfect, except that mere differencess of pusition in time and space are treated as irrelevant, is a valid basis for a generalisation, two total causes being regarded as the sume if they only differ in their positions in time or space. This, I think, is the whole of the importance which this law has for the theory of inductive argument. It involves the assertion of a deneralised judgment of irrelevance. namely, of the irrelevance of mere position in time and space to erneralisations which have no reference to particular positions in time and space. It is in respecet of such position in time or space that 'nature' is supposed 'uniform.' The significance of the law
and the mature of its justification, if any, are further disenssent in Chapter XXII.
6. Let us now pass to the type which is next in order of simplicity. We will relax the first comdition and no honer raseume that the whele wi the positive analugy between the instancos is ensered the the frerali-ation. though retaming the assumption that cur kmonleden of the examimed instances is complete. We know, that is to say, that there are some respects in which the examined instances are all alike, and yet which are not covered by the generalisation. If \(\phi_{1}\) is the part of the positive analogy
 then the peotability of this typ of armoment from analoey (an be writton-
\[
u(\phi, f) / I \quad\left(\phi \phi_{1} f\right) .
\]

The value of this probability turns on the comprehensiveness of \(\phi_{1}\). There are some characteristics \(\phi_{1}\) common to all the instances, which the generalisation treats as unessential, but the less comprehensive these are the better. \(\phi_{1}\) stands for the characteristics in which all the instances resemble one another outside those covered by the generalisation. To reduce these resemblances between the instances is the same thing as to increase the differences between them. And hence any increase in the Negative Analogy involves a reduction in the comprehensiveness of \(\phi_{1}\). When, however, our knowledge of the instancers is complete, it is not necessary to make separate mention of the negative analogy \(\bar{A} \quad\left(\phi^{\prime}\right)\) in the above formula. For \(\phi^{\prime}\) simply includes all these functions about the instances, which are not included in \(\phi \phi_{1} f\), and of which the contradictories are not included in them; so that in stating \(A \quad\left(\phi \phi_{1,} /\right)\), we state by implication \(\bar{A}\left(\phi^{\prime}\right)\) also.

The whole process of strengthening the argument in favour of the generalisation \(g(\phi, f)\) by the accumulation of further experience appears to me to consist in making the argument approximate as nearly as possible to the conditions of a perfoct analogy, by steadily reducing the comprehensiomess of those resemblances \(\phi_{1}\) between the instances which our generalisation

from experience, arises not out of their number as such, but out of their tendency to limit and recluce the comprehensiveness of \(\phi_{1}\), or, in other words, out of their tendency to increase the negative analogy \(\phi^{\prime}\), since \(\phi_{1} \phi^{\prime}\) comprise between them whatever is not covered by \(\phi f\). The more numerous the instances, the less comprehensive are their superfluous resemblances likely to be. But a single additional instance which greatly reduced \(\phi_{1}\) would increase the probability of the argument more than a large number of instances which affected \(\phi_{1}\) less.
7. The nature of the argument examined so far is, then, that the instances all have some characteristics in common which we have ignored in framing our generalisation ; but it is still assumed that our knowledge about the examined instances is complete. We will next dispense with this latter assumption, and deal with the case in which our knowledge of the characteristics of the examined instances themselves is or may be incomplete.

It is now necessary to take explicit account of the known negative analogy. For when the known positive analogy falls short of the total positive analogy, it is not possible to infer the negative analogy from it. Differences may be known between the instances which cannot be inferred from the known positive analogy. The probability of the argument must, therefore, be written-
\[
g\left(\phi, f^{\prime}\right) / \underset{n_{1} \ldots \omega_{n}}{\mathrm{~A}}\left(\phi \phi_{1} /\right) \underset{n_{1} \ldots n_{n}}{\overline{\mathrm{~A}}}\left(\phi^{\prime}\right),
\]
where \(\phi \phi_{1} f\) stands for the characteristics in which all \(n\) instances \(a_{1} \ldots a_{n}\) are known to be alike, and \(\phi^{\prime}\) stands for the characteristics in which they are known to differ.

This argument is strengthened by any additional instance or by any additional knowledge about the former instances which diminishes the known superfluous resemblances \(\phi_{1}\) or increases the negative analogy \(\phi^{\prime}\). The object of the accumulation of further experience is still the same as before, namely, to make the form of the argument approximate more and more closely to that of perfect analogy. Now, however, that our knowledge of the instances is no longer assumed to be complete, we must take account of the mere number \(n\) of the instances, as well as of our specific knowledge in regard to them; for the morn numerous the instances are, the greater the opportunity for the totul negative analogy to exceed the known negative analogy. But
the cmore complete our knowledre of the instances, the liss attention nowd we pay th their mere number, and the more imperfect our knowleder the greater the stress which must be laid upon the argument from number. This part of the argument will l.. discussed in detail in the following chapter on Pure Induction.
8. When our knowledre of the instances is incomplete, there may exist analowies which are known to be true of some of the instances and are not known to be false of any. These subanalogies sce \(\S 2\) ) are not so dangerous as the positive analogies \(\phi_{1}\), which are known to be true of all the instances, but the ir existence is, evidently, an clement of weakness, which we must cond avour to eliminate be the growth of knowlelqe and the multiplication of instancers. 1 sub-analory of this kind between the instances \(a_{r} \ldots a_{s}\) may be written A \(\left(\psi_{l}\right)\); and the formula, if it is to take account of all the relevant information, ought, thert fore, to be written-
\[
\left(\phi,,^{\prime}\right) \quad 1 \quad\left(\phi \phi_{1} j^{\prime}\right) \quad \bar{I}\left(\rho^{\prime}\right)!1 \quad \text { A }\left(\psi^{\prime}\right)!\text {, }
\]
where the terms of II! A \(\left(\psi_{i}\right)\) stand for the various sub,
analogies between sub-classes of the instances, which are not included in \(\phi \phi_{1} f\) or in \(\phi^{\prime}\).
9. There is now another complexity to be introduced. We must dispense with the assumption that the whole of the analong covered by the generalisation is known to exist in all the instances. For there may ber some instances within our experience, about. which our knowledere is incomplete, but which show part of the analogy required by the ereneralisation and mothine which contradicts it: and such instances afford some support to the Promeraliation. Buppose that क and \(f\) are prert of \(\phi\) and \(f\) respectively, then we may have a set of instaners \(i_{1} \ldots h\) which show the following analogies :
\[
\mathrm{A}\left(, \phi, \phi_{1}, f\right) \mathrm{A} \quad\left(, \phi^{\prime}\right) H: \mathrm{A}(, \psi)!,
\]

Where \(\phi_{1}\) is the analuge not coverad be theremeralisation, and so on, as before.

The formula, therefore, is now as follows:

In this expression " \(\phi\), , \(f\) are the whole or part of \(\phi, f\); the product II is composed of the positive and negative analogies for each of the sets of instances \(a_{1} \ldots a_{n}, b_{1} \ldots b_{m}\), etc.; and the product II contains the various sub-analogies of different subclasses of all the instances \(a_{1} \ldots a_{n}, b_{1} \ldots b_{m}\), etc., regarded as one set. \({ }^{1}\)
10. This completes our classification of the positive evidence which supports a generalisation ; but the probability may also be affected by a consideration of the negative evidence. We have taken account so far of that part of the evidence only which shows the whole or part of the analogy we require, and we have neglected those instances of which \(\phi\), the condition of the generalisation, or \(f\), its conclusion, or part of \(\phi\) or of \(f\) is known to be fulse. Suppose that there are instances of which \(\phi\) is true and \(f\) false, it is clear that the generalisation is ruined. But cases in which we know part of \(\phi\) to be true and \(f\) to be false, and are ignorant as to the truth or falsity of the rest of \(\phi\), weaken it to some extent. We must take account, therefore, of analogies
\[
\mathrm{A}\left(\ldots, \phi_{n} f\right),
\]
where \({ }_{\text {" }} \phi\), part of \(\phi\), is true of all the set, and \({ }_{"} f\), part of \(f\). is false of all the set, while the truth or falsity of some part of \(\phi\) and \(f\) is unknown. The negative evidence, however, can strengthen as well as weaken the evidence. We deem instances favourably relevant in which \(\phi\) and \(f\) are both false together. \({ }^{2}\)

Our final formula, therefore, must include terms, similar to those in the formula which concludes \(\S 9\), not only for sets of instances which show analogies " \(\phi_{u} f\), where " \(\phi\) and " \(f\) are parts of \(\phi\) and \(f\), but also for sets which show analogies \({ }_{a} \phi_{u} f\),
\({ }^{1}\) Even if we want to distinguish between the sub-analogies of the \(a\) set and the sub-analogies of the \(b\) set, this information can be gathered from the produet II.
\({ }^{2}\) I am disposed to think that we need not pay attention to instanose for which part of \(\phi\) is known to be false, and part of \(f\) to be true. But the \({ }^{q} q u e s t i o n ~ i s ~ a ~ l i t t l e ~ p e r p l e x i n g . ~\)
or analewies \(\bar{\phi} f\), where \(\phi\) and of are the whole or part of \(\phi\) and \(f\), and \(\phi f\) are the contratictorios of \(\phi\) and \(f\). \({ }^{1}\)

It should be added, perhaps, that the theoretical classification of tmont empirical arguments in daily use is complicatent be the account which we reasonably take of generalisations previously established. We often take account indirectly, therefore, of evidmew which supperts in som durew other meneralisations than that which we are concerned to establish or refute at the moment, hut the peobability of which is rele vant to the problem under investigation.
11. The argument will be rendered unnecessarily complex, without much benefit to its theoretical interest, if we deal with the most general case of all. What follows, therefore, will deal with the formula of the third degree of generality, namely-
in which no partial instances occur, i.e. no instances in which part only of the analogy, required by the generalisation, is known to exist. In this third degree of generality, it will be remembered, our knowledge of the characteristics of the instances is incomplete, there is more analogy between the instances than is covered by the generalisation, and there are some sub-analogies to be reckoned with. In the above formula the incompleteness of our knowledge is implicitly recognised in that \(\phi \phi_{1} f \phi^{\prime}\) are not between them entirely comprehensive. It is also supposed that all the evidence we have is positive, no knowledge is assumed, that is to say, of instances characterised by the con-


An argument, therefore, from experience, in which, on the basis of examined instances, we establish a deneralisation applicable beyond these instances, can be strengthemed, if we restrict our attention to the simpler type of cass, be the following means:
(1) By reducing the resemblances \(\phi_{1}\) known to be common to all the instances, but ignored as unessential by the generalisation.
(2) By increasing the differences \(\phi^{\prime}\) known to exist between the instances.

\footnotetext{


}
(3) By diminishing the sub-analogies or unessential resemblances \(\psi\), known to be common to some of the instances and not known to be false of any.

These results can generally be obtained in two ways, either by increasing the number of our instances or by increasing our knowledge of those we have.

The reasons why these methods seem to common sense to strengthen the argument are fairly obvious. The object of (1) is to avoid the possibility that \(\phi_{1}\) as well as \(\phi\) is a necessary condition of \(f\). The object of (2) is to avoid the possibility that there may be some resemblances additional to \(\phi\), common to all the instances, which have escaped our notice. The object of (3) is to get rid of indications that the total value of \(\phi_{1}\) may be greater than the known value. When \(\phi \phi_{1} f\) is the total positive analogy between the instances, so that the known value of \(\phi_{1}\) is its total value, it is (1) which is fundamental ; and we need take account of (2) and (3) only when our knowledge of the instances is incomplete. But when our knowledge of the instances is incomplete, so that \(\phi_{1}\) falls short of its total value and we camnot infer \(\phi^{\prime}\) from it, it is better to regard (2) as fundamental ; in any case every reduction of \(\phi_{1}\) must increase \(\phi^{\prime}\).
12. I have now attempted to analyse the various ways in which common practice seems to assume that considerations of Analogy can yield us presumptive evidence in favour of a generalisation.

It has been my object, in making a classification of empirical arguments, not so much to put my results in forms closely similar to those in which problems of generalisation commonly present themselves to scientific investigators, as to inquire whether ultimate uniformities of method can be found beneath the innumerable modes, superficially differing from another, in which we do in fact argue.

I have not yet attempted to justify this way of arguing. After turning aside to discuss in more detail the method of Pure Induction, I shall make this attempt ; or rather I shall try to see what sort of assumptions are capable of justifying cmpirical reasoning of this kind.

\section*{(HAPTER XX}

\section*{THE VALUE OF MULTIPLICATION OF IN゙STANCES, OR PVRE INI)UCTION}
1. It has oftom been thought that the essemee of inductive arenment lies in the multiplication of instances. "Where is that process of reasonime." Hume impuired, " which from om instane" draws a conclusion, so different from that which it infors from a humbed instances, that are no way different from that single instanew!" I re, eat that by emphasising the number of the instances Hume whscured the real whenet of the method. If it were strictly true that the hundred instances are no way different from the single instance. Hume would be right to wonder in what manner they can strengthen the argument. The object of increasine the number of instances arises out of the fact that we are nearly always aware of some difierener lwotwen the instances. and that wen where the known differenee is insignificant we may suspect, especially when our knowledge of the instances is very incomplete, that there may be more. Every new instance may diminish the unessential resemblances between the instances and
 For this reason, and for this reason only, new instances are valuable.

If our premisers comprise the body of memory and tradition which has been orivinally derived from direct experience, and the conclusion which we seek terestablish is the Newtemian theory of the sular systom. our arrument is one of P'ure Indurtion. if so lar as we suppert the Newtonian theory by pointime th the great mumber of consequmenes whic! it has in commum with the facts of experience. The predictions of the Nautical Almanark
 are verified many thousand times a day. But even here the
force of the argument largely depends, not on the mere number of these predictions, but on the knowledge that the circumstances in which they are fulfilled differ widely from one another in a vast number of important respects. The variety of the circumstances, in which the Newtonian generalisation is fulfilled, rather than the number of them, is what seems to impress our reasonable faculties.
2. I hold, then, that our object is always to increase the Negative Analogy, or, which is the same thing, to diminish the characteristics common to all the examined instances and yet not taken account of by our generalisation. Our method, however, may be one which certainly achieves this object, or it may be one which possibly achieves it. The former of these, which is obviously the more satisfactory, may consist either in increasing our definite knowledge respecting instances examined already, or in finding additional instances respecting which delinite knowledge is obtainable. The second of them consists in finding additional instances of the generalisation, about which, however, our definite knowledge may be meagre ; such further instances, if our knowledge about them were more complete, would either increase or leave unchanged the Negative Analogy ; in the former case they would strengthen the argument and in the latter case they would not weaken it ; and they must, therefore, be allowed some weight. The two methods are not entirely distinct, because new instances, about which we have some knowledge but not much, may be known to increase the Negative Analogy a little by the first method, and suspected of increasing it further by the second.

It is characteristic of advanced scientific method to depend on the former, and of the crude unregulated induction of ordinary experience to depend on the latter. It is when our definite knowledge about the instances is limited, that we must pay attention to their number rather than to the specific differences between them, and must fall back on what I term Pure Induction.

In this chapter I investigate the conditions and the manner in which the mere repetition of instances can add to the force of the argument. The chief value of the chapter, in my judgment, is negative, and consists in showing that a line of advance, which might have seemed promising, turns out to be a blind aller, and that we are thrown back on known Analogy. Pure

Induction will mot wive us any very substantial ascistance in getting to the bottom of the general inductive problem.
3. The prohlen of ermealisation \({ }^{1}\) hy Pare Induction can be stated in the following symbolic form :

Let hepmesent the -wneral ie frimi duta of the investigation: let \(g\) represent the generalisation which we seek to establish; let \(x_{1} x_{2} \ldots x_{n}\) represent instances of \(g\).

Then \(r_{1}\) ghe-1, \(x_{2}\) gh \(1 \ldots x\) ghe 1; given \(y\), that is to say, the truth of each of its instances follows. 'The protbem is to determine the probability \(g / h x_{1} x_{2} \ldots x_{n}\), i.e. the probability of the generalisation when \(u\) instances of it are given. Our
 will he lost if we intrexture the ascumption that there is nothing in our it priori duta which leads us to distinguish between the is prioni likelihome of ther different instances: We assumbe that is th saly, that there is nu reason it priori forexperting the oxecurrence of any one instance with greater reliance than any other, i.e.

ITrit.
\[
x_{1} \text { th } x_{2}{ }_{2}^{\prime} h \quad \ldots \quad x_{3} h
\]
al:4
\[
\text { "h 1 } 2 \ldots
\]

then
 is the i prori probability of the generalisation.

\footnotetext{
 tion of all the properitions which follow from it. For if \(h\) is any propesition.
 Pure Induction concister in finding as many instances of a pencrabisation as posable, it is, in the wident sense, the process of strengthening the probability of any proposition by adducing numerous instances of known truths which follow from it. 'The argument is one of Pure Induction, therefore, in so far as the probability of a condusion is based upon the number of independent cons-

}

It follows, therefore, that \(p_{n}>p_{n-1}\) so long as \(y_{n} \neq 1\).
Further,
\[
\begin{aligned}
& =y_{1} \cdot x_{1}^{x_{2}} \ldots n_{1}, \ldots 1 / h \\
& =y_{1, y_{n}-1 \cdots y_{1} .} \\
& \therefore I_{n \prime}^{\prime}=\frac{I_{0}}{y_{1} y_{2} \cdots y_{n}}=\frac{P_{n}}{I_{1} T_{2} \ldots I_{1}} \\
& =\quad p_{0} \\
& x_{1} x_{2} \ldots x_{n} y / h+x_{1} x_{2} \ldots x_{n} \pi^{\pi} / h \\
& =\frac{p_{0}}{g / h+x_{1} x_{2} \ldots x_{n} / \bar{y} h \cdot \bar{y} \mid h} \\
& =\frac{p_{0}}{p_{0}+r_{1} 1_{2} \ldots w_{n} / \bar{y} h\left(1-p_{0}\right)} .
\end{aligned}
\]

This approaches unity as a limit, if \(x_{1} x_{2} \ldots x_{n} / \bar{y} h .{ }_{p_{n}}^{1}\) approaches zero as a limit, when \(n\) increases.
4. We may now stop to consider how much this argument has proved. We have shown that if each of the instances necessarily follows from the generalisation, then each additional instance increases the probability of the generalisation, so long as the new instance could not have been predicted with certainty from a knowledge of the former instances. \({ }^{1}\) This condition is the same as that which came to light when we were discussing Analogy. If the new instance were identical with one of the former instances, a knowledge of the latter would enable us to predict it. If it differs or may differ in analogy, then the condition required above is satisfied.

The common notion, that each successive verification of a doubtful principle strengthens it, is formally proved, therefore, without any appeal to conceptions of law or of causality. But we have not proved that this probability approaches certainty as a limit, or even that our conclusion becomes more likely than not, as the number of verifications or instances is indefinitely increased.
5. What are the conditions which must be satisfied in order that the rate, at which the probability of the generalisation increases, may be such that it will approach certainty as a

\footnotetext{
\({ }^{1}\) Since \(p_{n} \cdot p_{n}\), so long as \(y_{n} \neq 1\).
}
limit when the number of independent instances of it are indefinitely increasad? Wie hase already shown, as a basis for this investimation, that \(p\) approaches the limit of certainty for a generalisation \(g\), if, as \(n\) increases, \(x_{1} x_{2} \ldots x_{n} / y h\) becomes small compared with \(p_{1}\).. ie. if the it priori probahility of so mans instances assumine the fals fhemet of the eremeralisation, is suall comprared with the exeneralisation's a priori probability. It. follows, therefore that the probatility of an induction tomels towards certanity as a limit, when the mumber of instances is increased, provided that
\[
12 \cdots-1^{2,1}-1+
\]
for all values of \(r\), and \(p_{0}>\eta\), where \(\epsilon\) and \(\eta\) are finite probabilitios, separated. that is to sat, from impossibility by a value of some finite amount, howerer small. These conditionsappar simple, but the meaning of a 'finite probability' requires a word of explanation. \({ }^{1}\)

I argued in Chapter III. that not all probabilities have an exact numerical value, and that, in the case of some, one can say no more about their relation to eertainty and impossibility than that they fall shom of the former and exemed the latter. There is one class of prohatilities, howerer, which I callod the nummerical class, the ratio of each of whose members to certainty can be expressed by sume number less than unity: and we can sometimes compare a nom-numerical probability in respect of more and less with one of these numerical probabilities. This enables us to
 tion to nom-numerical as well as to mumerical probabilities. I define a 'finite probability' as one which exceeds some numerical probability, the ration of which to certainty can be expressed be at finite number.2 The principal methond. in which a probability can be prowed finite hy a prowss of arsument arises either when
\({ }^{1}\) The proof of thase conditions, which is ohvious, is as follows:

\footnotetext{
 conditions, some finite value of \(n\) such that both ( 1,\()^{\prime \prime}\) and (b \({ }^{(1)}\) are less than any given finite guantity, however small.
 any positive finite number e however small, a positive inteser \(n\) can always bu found such that for all values of \(r\) greater tham \(n\) the difference between L and \(l^{\prime} r\) is less than \(e \cdot \gamma\), where \(\gamma\) is the measure of certainty.
}

its conclusion can be sliown to be one of a finite number of alternatives, which are between them exhaustive or, at any rate, have a finite probability, and to which the Principle of Indifference is applicable ; or (more usually), when its conclusion is more probable than some hypothesis which satisfies this first condition.
6. The conditions, which we have now established in order that the probability of a pure induction may tend towards certainty as the number of instances is increased, are (1) that \(x_{r} \mid x_{1} x_{2} \ldots x_{r-1} \bar{g} h\) falls short of certainty by a finite amount for all values of \(r\), and (2) that \(p_{0}\), the a priori probability of our generalisation, exceeds impossibility by a finite amount. It is easy to see that we can show by an exactly similar argument that the following more general conditions are equally satisfactory :
(1) That \(x_{r} / x_{1} x_{2} \ldots x_{r-1} \mathbf{1}^{j} h\) falls short of certainty by a finite amount for all values of \(r\) beyond a specified value \(s\).
(2) That \(p_{s}\), the probability of the generalisation relative to a knowledge of these first \(s\) instances, exceeds impossibility by a finite amount.

In other words Pure Induction can be uss fully employed to strengthen an argument if, after a certain number of instances have been examined, we have, from some other source, a finite probability in favour of the generalisation, and, assuming the generalisation is false, a finite uncertainty as to its conclusion being satisfied by the next hitherto unexamined instance which satisfies its premiss. To take an example, Pure Induction can be used to support the generalisation that the sun will rise every morning for the next million years, provided that with the experience we have actually had there are finite probabilities, however small, derived from some other source. first, in favour of the generalisation, and, second, in farour of the sun's not rising to-morrow assuming the generalisation to be false. Given these finite probabilities, obtained otherwise, however small, then the probability can be strengthened and can tend to inerease towards certainty by the mere multiplication of instances provided that these instances are so far distinct that they are not inferrible one from another.
7. Those supposed proofs of the Inductive Principle, which are based openly or implicitly on an argument in inverse probability, are all vitiated by unjustifiable assumptions relating to the magnitude of the a priori probability \(p_{0}\). Jevons, for
instance a anwedly assumes that we may. in the absence of sperial information? suppmar alay mesamined hypmothesis to he as like.t. as not. It is difficult to see how such a belief, if even its most immediat. impleations had hampropery appolmoded. conald have remained plausihbe to a mind of so sound a practical judgment as his. The arguments against it and the contradictions to which it leads have, been dealt with in Chapter IV. The demonstration of Laplace, which depends upon the Rule of Succession, will be discussed in Chapter XXX.
8. The prior probability, which must always be found, before the methent if pure induetion can he nse fully emplowed to support a suhstantial armment, is derised. I think, in most ordinary cases- with what justification it remains to discuss --from considerations of A Aalosey. But ther conditions of valid indurtion as they hase hawn mumetated atmes are quite independent of analocs and mixht ber applicabhe to other types of argument. In certain casta whe mieht fiel justifiod in assumine dimetly that the necessary conditions are satisfied.

Our belief, for instance, in the validity of a logical scheme is based partly upon inductive grounds on the number of conclusions, each seemingly true on its own account, which can be derived from the axioms - and partly on a degree of self-evidence in the axioms themselves sufficient to give them the initial probatility upen which imbution ran build. Wi. depend upen the initial presumption that, if a proposition appears to us to be true, this is by itself, in the absence of opposing evidence, some renson for its being as well as appearing true. We cannot deny that what appars true is sometimes false, but, unless we can assume some substantial relation of probability between the appearance and the reality of truth, the possibility of even probable knowledere is at an end.

The conception of our having some reason, though not a
 may prove important to the theory, of epistemology. The old metaphysics has beon greatly hindered by reason of its having always demanded demonstrative certainty. Much of the cogency of Hume's criticism arises out of the assumption of methods of certainty on the part of those systems against which it was directed. The earlier realists were hampered by their not per-

what they wanted in the end. And transeendental philosophy has partly arisen, I believe, through the belief that there is no knowledge on these matters short of certain knowledge, being combined with the belief that such certain knowledge of metaphysical questions is beyond the power of ordinary methods.

When we allow that probable knowledge is, nevertheless, real, a new method of argument can be introduced into metaphysical discussions. The demonstrative method can be laid on one side, and we may attempt to advance the argument by taking account of circumstances which seem to give some reason for preferring one alternative to another. Great progress may follow if the nature and reality of objects of perception, \({ }^{1}\) for instance, can be usefully investigated by methods not altogether dissimilar from those employed in science and with the prospect of obtaining as high a degree of certainty as that which belongs to some scientific conclusions; and it may conceivably be shown that a belief in the conclusions of science, enunciated in any reasonable manner however restricted, involves a preference for some metaphysical conclusions over others.
9. Apart from analysis, careful reflection would hardly lead us to expect that a conclusion which is based on no other than grounds of pure induction, defined as I have defined them as consisting of repetition of instances merely, could attain in this way to a high degree of probability. To this extent we ought all of us to agree with Hume. We have found that the suggestions of common sense are supperted by more precise methods. Moreover, we constantly distinguish between arguments, which we call inductive, upon other grounds than the number of instances upon which they are based; and under certain conditions we regard as crucial an insignificant number of experiments. The method of pure induction may be a useful means of strengthening a probability based on some other ground. In the case, however, of most scientific arguments, which would commonly be called inductive, the probability that we are right, when we make predictions on the hasis of past experience, depends not so much on the number of past experiences upon which we rely, as on the degree in which the circumstances of these experiences

\footnotetext{
\({ }^{1}\) A paper by Mr. G. E. Moore entitled, "The Nature and Reality of Objects of Pereeption." which was published in the Proceedings of the Aristoltian Society for fonf. serms to nut twaply for the first time a method somewhat resembling that which is described aboye,
}
rasemble the known cirenmatances in which the prediction is to take effect. Scientific method, indeed, is mainly devoted to discovering means of so heightening the known analogy that we may dispense as far as possible with the methods of pure imblution.

When, therefore, our previous knowledge is considerable and the analogy is good, the purely inductive part of the argument may take a very subsidiary place. But when our knowledge of the instances is slight, we may have to depend upon pure induction a grood deal. In an advanced science it is a last resort, -the least satisfactory of the methods. But sometimes it must be our first resort, the method upon which we must depend in the dawn of knowledge and in fundamental inquiries where we must presuppose nothing.

\section*{CHAPTER XXI}

\section*{THE NATURE OF INDUCTIVE ARGUMENT CONTINUED}
1. In the enunciation, given in the two preceding chapters, of the Principles of Analogy and Pure Induction there has been no reference to experience or causality or law. So far, the argument has been perfectly formal and might relate to a set of propositions of any type. But these methods are most commonly emploved in physical arguments where material objects or experiences are the terms of the generalisation. We must consider, therefore, whether there is any good ground, as some logicians seem to have supposed, for restricting them to this kind of inquiry.

I am inclined to think that, whether reasonably or not, we naturally apply them to all kinds of argument alike, including formal arguments as, for example, about numbers. When we are told that Fermat's formula for a prime, namely, \(: 2^{2^{x}}+1\) for all values of \(a\), has been verified in every case in which verification is not excessively laborious -namely, for \(a=1,2,3\), and 4, we feel that this is some reason for accepting it, or, at least, that it raises a sufficient presumption to justify a further examination of the formula. \({ }^{1}\) V'ot there can be no reference here to the uniformity of nature or physical causation. If inductive methods are limited to natural objects, there can no more be an appreciable ground for thinking that \(2^{2^{4}}+1\) is a true formula for primes, because empirical methods show that it violds primes up to a 1 , or even if they showed that it yielded primes for ceery number up to a million million, than there is to think that any formula which I may choose to write down

1 This formall har, in fart, heen dispored in recent times. c.g. \(2^{5}+1 \ldots\) \(4,294,967,297=641 \times 6,700,417\). Thus it is no longer so good an illustration as it would have been a hund"ed years ago.
at random is a true source of primes. To maintain that there is no appreciable gromed in such a case is paradoxical. If, on the other hand a partial verification does raise some just appreciable presumption in the formula's favour, then we must include numbers, at any rate, as well as maturial objects amomest the proper sulijects of the inductive methed. The conclusion of the previons chapter indicates, however, that, if arguments of this kimd have force it can only be in virtue of there being some finite a priori probability for the formula based on other than inductive grounds.

There are some illustrations in Jevons's Principles of science, \({ }^{1}\) which are relevant to this discolssion. We find it to be true of the following six numbers:
\[
5,15,35,45,65,95
\]
that they all end in five, and are all divisible by five without remainder. Would this fact. by itself, raise any kind of presumption that all numbers emding in five are divisible be fiwe withent remainder? Let us also consider the six numbers,
\[
7,17,37,47,67,97 .
\]

They all end in seven and also agree in being primes. Would this raise a presumption in fasour of the eremeralisation that all numbers atre prime, which cme in seren! We might be prajudiced in favour of the first aremment, beranse it would luad us to a true comelusion; but we ought nut to be prejudiced acainst the second because it would lead us tor a false one ; for the validity of empirical arguments as the fommation of a probability canmet be affected by the actual truth or falsity of their conclusions. If, on the evidence, the analogy is similar and cqual, and if the scope of the enmeralisation and its conclusion is similar, then the value of the two arguments must be equal also.

Whether or not the useof empirical argment appoars phansible to us in these particular examples, it is certainly true that mans mathematical thentems have actually been discovernd be such methods. Gemeralisations have been sumerested nearly as oftem, perhaps, in the lewical and mathematical sciences, as in the

\footnotetext{



}
physical, by the recognition of particular instances, even where formal proof has been forthcoming subsequently. Yet if the suggestions of analogy have no appreciable probability in the formal sciences, and should be permitted only in the material, it must be unreasonable for us to pursue them. If no finite probability exists that a formula, for which we have empirical verification, is in fact universally true, Newton was acting fortunately, but not reasonably, when he hit on the Binomial Theorem by methods of empiricism. \({ }^{1}\)
2. I am inclined to believe, therefore, that, if we trust the promptings of common sense, we have the same kind of ground for trusting analogy in mathematics that we have in physics, and that we ought to be able to apply any justification of the method, which suits the latter case, to the former also. This does not mean that the à priori probabilities, from some other source than induction, which the inductive method requires as its foundation, may not be sought and found differently in the two types of inquiry. A reason why it has been thought that analogy ought to be confined to natural laws may be, perhaps, that in most of those cases, in which we could support a mathematical theorem by a very strong analogy, the existence of a formal proof has done away with the necessity for the limping methods of empiricism ; and because in most mathematical investigations, while in our earliest thoughts we are not ashamed to consult analogy, our later work will be more profitably spent in searching for a formal proof than in establishing analogies which must, at the best, be relatively weak. As the modern scientist discards, as a rule, the method of pure induction, in favour of experimental analogy, where, if he takes account of his previous knowledge, one or two cases may prove immensely significant; so the modern mathematician prefers the resources of his analysis, which may yield him certainty, to the doubtful promises of empiricism.
3. The main reason, however, why it has often been held that we ought to limit inductive methods to the content of the particular material universe in which we live, is, most probably, the fact that we can casily imagine a universe so constructed that such methods would be useless. This suggests that analogy and induction, while they happen to be useful to us in this world,

\footnotetext{
\({ }^{1}\) Sce Jevons, loc. cit. p. 231.
}
camment he universal principhes of lowic, on the same footines. for instance, as the syllogism.

In one sense this opinion may be well founded. I do not deny or affirm at present that it may be necessary to confine inductive methods to arguments about certain kinds of objects or certain kinds of experiences. It may be true that in every useful arqument from analogy our premisses must contain fundamental assumptions, whatmed directly and mot inductively, which somme possible experiences might preclude. Moreover, the success of induction in the past can certainly affect its probable usefulness for the future. We may discover something about the nature of the universe -we may even discover it by means of induction itself the knowledge of which has the effect of destroving the further utility of induction. I shall argue later on that the confidence with which we ourselves use the method does in fact depend upon the nature of our past experience.

But this empirical attitude towards induction may, on the other hand, arise out of either one of two possible confusions. It may confuse, first, the reasonable character of arguments with their practical usefulness. The usefulness of induction depends, no doubt, upon the actual content of experience. If there were no repetition of detail in the universe, induction would have no utility. If there were only a single object in the universe, the laws of addition would have no utilits. But the processes of induction and addition would remain reasonable. It may confuse, secondly, the validity of attributing probability to the conclusion of an argument with the question of the actual truth of the conclusion. Induction tells us that, on the basis of certain evidence, a certain conclusion is reasonable, not that it is true. If the sun does not rise to-morrow, if (Queen Anne still lives, this will not prove that it was foolish or unreasomable of us to have believed the contrary.
4. It will be worth while to say a little more in this connection about the not infrequent failure to distinguish the rational from the true. The exeessive ridicule, which this mistake has visited on the supposed irrationality of barbarous and primitive peoples, affords some weod examples. "Reflection and enguiry should
 predecessors we are indehted for much of what we thought most our own, and that their errors were not wifful extravagances
or the ravings of insanity, but simply hypotheses, justifiable as such at the time when they were propounded, but which a fuller experience has proved to be inadequate. . . . Therefore, in reviewing the opinions and practices of ruder ages and races we shall do well to look with leniency upon their errors as inevitable slips made in the search for truth. . . ." The first introduction of iron ploughshares into Poland, he tells in another passage, having been followed by a succession of bad harvests, the farmers attributed the badness of the crops to the iron ploughshares, and discarded them for the old wooden ones. The method of reasoning of the farmers is not different from that of science, and may, surely, have had for them some appreciable probability in its favour. "It is a curious superstition," says a recent pioneer in Borneo, "this of the Dusuns, to attribute anything-whether good or bad, lucky or unlucky - that happens to them to something novel which has arrived in their country. For instance, my living in Kindram has caused the intensely hot weather we have experienced of late." \({ }^{1}\) What is this curious superstition but the Method of Difference?

The following passage from Jevons's Principles of Science well illustrates the tendency, to which he himself yielded, to depreciate the favourite analogies of one age, because the experience of their successors has confuted them. Between things which are the same in number, he points out, there is a certain resemblance, namely in number ; and in the infancy of science men could not be persuaded that there was not a deeper resemblance implied in that of number. "Seven days are mentioned in Genesis ; infants acquire their teeth at the end of seven months; they change them at the end of seven years ; seven feet was the limit of man's height ; every seventh year was a climacteric or critical year, at which a change of disposition took place. In natural science there were not only the seven planets, and the seven metals, but also the seven primitive colours, and the seven tones of music. So deep a hold did this doctrine take that we still have its results in many customs, not only in the seven days of the week, but the seven years' apprenticeship, puberty at fourteen years, the second climacteric, and legal majority at twenty-one years, the third climacteric:" Religious systems from Py thagoras to Comte have sought to derive strength from the virtue of seven.

\footnotetext{
\({ }^{1}\) Giolden liough, p. 174.
}
* And even in srimatilic matters the lufiest intellects have ocemsionally yilded, as when Xewtom was misled hy the analoey between the seven tones of music and the seven colours of his spectrum. . . . Even the genius of Huyghens did not prevent him from inferring that hut on- satrollite conld belone to saturn. because, with those of Jupiter and the earth, it completed the perfect number of six." But is it certain that Newton and Huyghens were only reasonable when their theories were true, and that their mistakes were the fruit of a disordered fancy? Or that the savages, from whom we have inherited the most fundamental imductions of our knowledere, were ahats superstitious when they believed what we now know to be preposterous ?

It is important to understand that the common sense of the race has been impressed by very weak analogies and has attributed to them an appreciable probability, and that a logical theory, which is to justify common sense, need not be afraid of including these marginal cases. Fern our belief in the real existence of other people, which we all hold to be well established, may require for its justification the combination of experience with a just appreciahle à priori possibility for Animism generally.' If we actually possess evidence which renders some conclusion absurd, it is very difficult for us to appreciate the relation of this conclusion to data which are different and less complete; but it is essential that we should realise arguments from analogy as relative to premisses, if we are th apprax h the Locical thoory of Induction without prejudice.
5. While we depreciate the former probability of beliefs which we no longer hold, we tend, I think, to exagrerate the presant duzere of cartainty of what we alll beliese. The pemediner paragraph is mot intended to deny that surames when areatly

\footnotetext{
 enlightened or civilised man is not there, and in the civilised man's child, if it be admitter that he has it at all, is but a faint survisal of it phase of tho primitive mind. And hy animism I do not moan the theory of a soul in nature, but the tendency or impulse or instinct, in which all myth orivinates, to animate all things: the projection of ourselves into nature; the sense and apprehension of an intellizence like our own, hut more powerful in all visible
 impulso or instinct,' refined by reason and enlarged by experience, may he required, in the shape of an intuitive a priori probability, if some of those universal condurions of common rense, which the mont seeptical to not kick away, are to be supported with rational furd.utions.
}
overestimate the value of their crude inductions, and are to this extent irrational. It is not easy to distinguish between a belief's being the most reasonable of those which it is open to us to believe, and its being more probable than not. In the same way we, perhaps, put an excessive confidence in those conclusions the existence of other people, for instance, the law of gravity, or to-morrow's sunrise- of which, in comparison with many other beliefs, we are very well assured. We may sometimes confuse the practical certainty, attaching to the class of beliefs upon which it is rational to act with the utmost confidence, with the more wholly objective certainty of logic. We might rashly assert, for instance, that to-morrow's sunrise is as likely to us as failure, and the special virtue of the number seven as unlikely, even to Pythagoras, as success, in an attempt to throw heads a hundred times in succession with an unbiassed coin. \({ }^{1}\)
6. As it has often been held upon various grounds, with reason or without, that the validity of Induction and Analogy depends in some way upon the character of the actual world, logicians have sought for material laws upon which these methods can be founded. The Laws of Universal Causation and the Uniformity of Nature, namely, that all events have some cause and that the same total cause always produces the same effect, are those which commonly do service. But these principles merely assert that there are some data from which events posterior to them in time could be inferred. They do not seem to yield us much assistance in solving the inductive problem proper, or in determining how we can infer with probability from partial data. It has been sugqested in the previous chapter that the Principle of the Uniformity of Nature amounts to an assertion that an argument from perfect analogy (defined as I have defined it) is valid when applied to events only differing in their positions in time or space." It has also been pointed out that ordinary inductive arguments appear to be strengthened by any evidence which makes them approximate more closely in character to a perfect analogy. But this, I think, is the whole extent to which this principle, even if its truth could be assumed, would help us.

\footnotetext{
\({ }^{1}\) Yet if every inhabitant of the world, Grimschl has calculated, were to toss a coin every second, day and night, this latter event would only occur once on the average in every twenty billion years.
\({ }^{2}\) Is this interpretation of the Principle of the Uniformity of Nature affected by the Doctrine of Relativity?
}

States of the miverse. identical in exery particular, may turem recur, and, even if identical states were to recur, we should not ktan it.

The hind of fomdanental :asimptien about the character of material laws, on which scientists appear commonly to act, seems to me to be much less simple than the bare principle of Uniformity: They appear to assume something much more like what mathematicians call the principle of the superposition of small effects, or, as I prefer to call it, in this comnection, the atomic character of natural law. The system of the material universe must consist, if this kind of assumption is warranted. of bodies which we may term (without any implication as to their size being convered thereby) leyal atoms, such that each of them exercises its own separate, independent, and invariable effect, a change of the total state being compounded of a number of separate changes each of which is solely due to a separate portion of the preceding state. We do not have an invariable relation between particular bodies, but nevertheless each has on the others its own separate and invariable effect, which does not change with changing circumstances, although, of course, the total effect may be changed to almost any extent if all the other accompanying causes are different. Each atom can, according to this theory, be treated as a separate cause and does not enter into different organic: combinations in each of which it is reculated by different laws.

Perhaps it has not always been realised that this atomic uniformity is in no way implied by the principle of the Uniformity of Nature. Yet there might well be quite different laws for wholes of different degrees of complexity, and laws of connection between complexes which could not be stated in terms of laws connecting individual parts. In this case natural law would be organie and not, as it is generally supposed, atomic. If every configuration of the Universe were subject to a separate and independent law, or if very small differences botween bodies in their shape or size, for instance,led to their obeying quite different laws, prediction would be impossible and the inductive method useless. Yet mature misht still be uniform, causation sovereign, and laws timeless and absolute.

The scientist wishes, in fact, to assume that the occurrence
of a phenomenon which has appeared as part of a more complex phenomenon, may be some reason for expecting it to be associated on another occasion with part of the same complex. l'et if different wholes were subject to difierent laws quê wholes and not simply on account of and in proportion to the differences of their parts, knowledge of a part could not lead, it would seem, even to presumptive or probable knowledge as to its association with other parts. Given, on the other hand, a number of leqally atomic units and the laws connecting them, it would be possible to deduce their effects pro tanto without an exhaustive knowledge of all the coexisting circumstances.

We do habitually assume, I think, that the size of the atomic unit is for mental events an individual consciousness, and for material events an object small in relation to our perceptions. These considerations do not show us a way by which we can justify Induction. But they help to elucidate the kind of assumptions which we do actually make, and may serve as an introduction to what follows.

\section*{(HAPTER XXII}
the Justification of these methods
1. The erencral line of thought to be followed in this chapter mas be indicated, briefly, at the outset.

I systom of facts or propesitions, as we ordinarily conceive it, may comprise an indefinite number of members. But the ultimate comstituents or indefimables of the sustem. which all the members of it are about, are less in number than these members themselves. Further, there are certain laws of necessary connection between the members, by which it is meant (I do not stop, to consider whe ther mene than this is meant) that the truth or falsity of every member can be infered from a knowledere of the laws of necessary comection together with a knowledge of the truth or falsity of some (but not all) of the members.

The ultimate constiturnts tugether with the laws of necessary connection make up what I shall term the indepembent variely of the system. The more numerous the ultimate comstitumes and the mecessary laws, the greater is the system's indepmentent variety. It is not necessary for my present purpuse, which is merely to brime before the readers mind the sort of comeeption which is in mine, that I should attompt a complete definition of what I mean by a system.

Now it is characteristic; of a systom, as distimmuishod from a collection of heteromemons and indepemdent facts or propositions, that the number of its premisses, or, in other wonds, the amount of independent varioty in it, shomld be liss than the number of its members. But it is mot an whionsly cessential characteristic of a systom that it premisses on its indermendent varioty should the actually finite. We must distimuish, therefore, botween systems which may ber termal finite and infinite respectively, the terms finite and infinite ow roime mot to
the number of members in the system but to the amount of independent variety in it.

The purpose of the discussion, which occupies the greater part of this chapter, is to maintain that, if the premisses of our argument permit us to assume that the facts or propositions, with which the argument is concerned, belong to a finite system, then probable knowledge can be validly olstained by means of an inductive argument. I now proceed to approach the question from a slightly different standpoint, the controlling idea, however, being that which is outlined above.
2. What is our actual course of procedure in an inductive argument? We have before us, let us suppose, a set of \(n\) instances which have \(r\) known qualities, \(a_{1} a_{2} \ldots a_{i}\) in common, these \(r\) qualities constituting the known positive analogy. From these qualities three (say) are picked out, namely, \(a_{1}, a_{2}, a_{3}\), and we inquire with what probability all objects having these three qualities have also certain other qualities which we have picked out, namely, \(a_{r-1}, a_{r}\). We wish to determine, that is to say, whether the qualities \(a_{r-1}, a_{1}\), are bound \(u p\) with the qualities \(a_{1}, a_{2}, a_{3}\). In thus approaching this question we seem to suppose that the qualities of an object are bound together in a limited number of groups, a sub-class of each group being an infallible symptom of the coexistence of certain other members of it also.

Three possibilities are open, any of which would prove destructive to our generalisation. It may be the case (1) that \({ }^{\alpha_{r-1}}\) or \({ }^{\prime}\), is independent of all the other qualities of the instances - they may not overlap, that is to say, with any other groups ; or (2) that \(a_{1} a_{2} f_{3}\) do not belong to the same groups as \(a_{r-1} a_{r}\); or (3) that \(a_{1} a_{2} \|_{3}\), while they belong to the same group as \(a_{r-1} 1_{1}{ }_{1}\), are not sufficient to specify this group uniquely they belong, that is to say, to other groups also which do not include \(a_{r-1}\) and \({ }^{\prime}\).. The precautions we take are directed towards reducing the likelihood. so far as we can, of each of these possibilities. We distrust the generalisation if the terms typified by \(a_{1,-11^{\prime \prime}}\), are numerous and comprehensive, because this increases the likelihood that some at least of them fall under heading (1), and also because it increases the likelihood of (3). We trust it if the terms typified by \("_{1} / l_{2} f_{3}\) are numerous and comprehensive, because this decreases the likelihood both of (2) and of (3). If
we find a new instance which agrees with the former instances in
 the possibility that it is \(\|_{4}\), alone or in combination, that is bound up with \(a_{r-1} l_{r}\). We desire to increase our knowledge of the properties, lest there be some positive analogy which is escaping us, and when our knowledge is incomplete we multiply instances, which we do not know to increase the negative analogy for certain, in the hope that they may do so.

If we sum up the various methods of Analogy, we find, I think, that they are all capable of arising out of anderlying assumption, that if we find two sets of qualities in coexistence there is a finite probability that they belong to the same group, and at tinite probabilit! also that the tirst are eperilies this ermup uniquely. Starting from this assumption, the object of the methods is to increase the finite probability and make it large. Whether or not anything of this sort is explicitly present to our minds when we reason scientifically, it seems clear to me that we do act exactly as we should act, if this were the assumption from which we set out.

In most cases, of course, the field is greatly simplified from the first by the use of our pre-existing knowledge. Of the properties before us we generally have good reason, derived from prior analogies, for supposing some to belong to the same group and others to belong to different groups. But this does not affect the theoretical problem confronting us.
3. What kind of ground could justify us in assuming the existence of these finite probabilities which we seem to require? If we are to obtain them, not directly, but by means of argument, we must somehow base them upon a finite number of exhaustive alternatives.

The following line of argument seems to me to represent, on the whole, the kind of assumption which is obscurely present to our minds. We suppose, I think, that the almost imnumerable apparent properties of any given object all arise out of a finite number of generator properties, which we may call \(\phi_{1} \phi_{2} \phi_{3} \ldots\) Some arise out of \(\phi_{1}\) alone, some out of \(\phi_{1}\) in conjunction with \(\phi_{2}\) and so on. The properties which arise out of \(\phi_{1}\) alone form one
 group, and so on. Nince the number of generator properties is finite, the number of groups also is finite. If a set of apparent
properties arise (say) out of three generator properties \(\phi_{1} \phi_{2} \phi_{3}\), then this set of properties may be said to specify the group \(\phi_{1} \phi_{2} \phi_{3}\). Since the total number of apparent properties is assumed to be greater than that of the generator properties, and since the number of groups is finite, it follows that, if two sets of apparent properties are taken, there is, in the absence of evidence to the contrary, a finite probability that the second set will belong to the group specified by the first set.

There is, however, the possibility of a plurality of generators. The first set of apparent properties may specify more than one group,-there is more than one group of generators, that is to say, which are competent to produce it; and some only of these groups may contain the second set of properties. Let us, for the moment, rule out this possibility.

When we argue from an analogy, and the instances have two groups of characters in common, namely \(\phi\) and \(f\), either \(f\) belongs to the group \(\phi\) or it arises out of generators partly distinct from those out of which \(\phi\) arises. For the reason already explained there is a finite probability that \(f\) and \(\phi\) belong to the same group. If this is the case, i.e. if the generalisation \(g(\phi f)\) is valid, then \(f\) will certainly be true of all other cases in which \(\phi\) is true ; if this is not the case, then \(f\) will not always be true when \(\phi\) is true. We have, therefore, the preliminary conditions necessary for the application of pure induction. If \(x_{r}\), etc., are the instances,
\[
\begin{aligned}
& g / h=p_{0}, \text { where } p_{0} \text { is finite, } \\
& x_{r} / g h=1, \text { etc. }
\end{aligned}
\]
and
\[
x_{r} / x_{1} x_{2} \ldots x_{r-1} \bar{y} h=1-\epsilon \text {, where } \epsilon \text { is finite. }
\]

And hence, by the argument of Chapter XX., the probability of a generalisation, based on such evidence as this, is capable, under suitable conditions, of tending towards certainty as a limit, when the number of instances is increased.

If \(\phi\) is complex and includes a number of characters which are not always found together, it must include a number of separate generator properties and specify a large group; hence the initial probability that \(f\) belongs to this group is relatively large. If, on the other hand, \(f\) is complex, there will be, for the same reasons mutatis mulundis, a relatively smaller initial probability than otherwise that \(f\) belongs to any other given group.

When the aroument is mainly by analogy, we endeavour to whtan widnowe which make the intial probability \(p\), relatively high: whon the analoy is wak and the argument depernds for its rimonth upen pure induction, \(p_{0}\) is small and \(p\), which is based upon numerwis instances, hepends for its magnitude upon their number. But an argument from induction must always inzohse some element of analogy, and, on the other hand, few armments from amalogy can afford to ignore altore ther the strengthening influence of pure induction.
4. Let us consider the manner in which the methods of analu-y inmease the initial like lifoned that two characters belone to the same eroup. The num rons characters of an object which are known to us may be represented by \(a_{1} a_{2} \ldots u^{\prime}\). Wie select two sets of these, \(a_{r}\) and \(a_{\text {s }}\), and seek to determine whether \(a_{\text {s }}\) always helonist th the erroup specitied be \(a\). Our pervious knowlenter will mable as in wemeral, to rule out many of the object:
 ahtionug this will but be presibl. in the most fundamental inynuriws. II. mate alow kown that (artann characters are abways associated with a or with \(a_{\text {. }}\). But there will be left a residuum of whose connection with \(a_{r}\) or \(a\) we are ignorant. These characturs, whome ruvance is in thout, may the represented by \(a_{1,1} \ldots a_{-1}\). If the analogy is perfect, these characters are Himmatad altone ther. Otherwis., the argument is wakened
 acters. For it may be the case that some of \(a_{r+1} \ldots a_{-1}\) are necessary as well as \(a_{r}\), in order to specify all the generators which are required to produce \(a\).
 characters \(a_{r, 1} \ldots a_{v-1}\) by direct judgments of irrelevance. Them afo certain properties of objects which we rule out trom the h...wimme as whilly or laresly inderendent and irrelevant to all. of th anne, wher properties. The principal judements of this kind, and those alone about which we seem to feel much contidence, are concerned with absolute position in time and space. this class of julluments of irrelevance beina summed up. I have suggested, in the Principle of the Uniformity of Nature. Wh. judere that mare pusition in time and apace cannot possibly aliont, as a determining callse, any other charateres: and this belief appors son strong and cortain, althoush it is hard tw see
how it can be based on experience, that the judgment by which we arrive at it seems perhaps to be direct. A further type of instance in which some philosophers seem to have trusted direct judgments of relevance in these matters arises out of the relation between mind and matter. They have believed that no mental event can possibly be a necessary condition for the occurrence of a material event.

The Principle of the Uniformity of Nature, as I interpret it, supplies the answer, if it is correct, to the criticism that the instances, on which generalisations are based, are all alike in being past, and that any generalisation, which is applicable to the future, must be based, for this reason, upon imperfect analogy. We judge directly that the resemblance between instances, which consists in their being past, is in itself irrelevant, and does not supply a valid ground for impugning a generalisation.

But these judgments of irrelevance are not free from difficulty, and we must be suspicious of using them. When I say that position is irrelevant, I do not mean to deny that a generalisation, the premiss of which specifies position, may be true, and that the same generalisation without this limitation might be false. But this is because the generalisation is incompletely stated; it happens that objects so specified have the required characters, and hence their position supplies a sufficient criterion. Position may be relevant as a sufficient condition but never as a necessary condition, and the inclusion of it can only affect the truth of a generalisation when we have left out some other essential condition. A generalisation which is true of one instance must be true of another which only differs from the former by reason of its position in time or space.
6. Excluding, therefore, the possibility of a plurality of generators, we can justify the method of perfect analogy, and other inductive methods in so far as they can be made to approximate to this, by mrans of the assumption that the objects in the field, over which our generalisations extend, do not have an infinite number of independent qualities ; that, in other words, their characteristics, however numerous, cohere together in groups of invariable connection, which are finite in number. This does not limit the number of entities which are only momericall!y distinct. In the language used at the begiming of this chapter, the use of inductive methods can be
justified if they are applied to what we have reascon to suppose a finite system. \({ }^{1}\)
7. Let us now take account of a possible plurality of generators. I mean hy this the possibility that a siven character can arise in more than one way, can belone to more than one di-tinct group, and can arise out of more than ome generator. \(\phi\) might, for instanee, he sometimes due to a generator \(a_{1}\) and \(a_{1}\) might invariably produce \(f\). But we could not eseneralise from \(\phi\) to \(f\), if \(\phi\) might be due in other cases to a different generator \(a_{2}\) which would not be competent to produce \(f\).

If we were dealine with inductive correlation, where we do not claim uniwersality for our conclusions, it would be sufficient for us to assume that the number of distinct menerators, to which a given property \(\phi\) can bedue, is always finite. Toobtain validity for universal eneralisations it seems necessary to make the more comprehensive and loss plansible assumption that a finite probability always exists that there is not, in any wiven case, a plurality of causes. With this assumption we have a valid argument from pure induction on the same lines, nearly, as before.
8. Wie have thus two distinct difficulties to deal with, and we require for the solution of each a separate assumption. The point may hee illustrated by an example in which only one of the difficulties is present. There are few arguments from analogy of which we are better assured than the existence of other people We ferl indend so well assured of their existence that it has heen thought sometimes that our knowledre of them must he in some way direct. But analome does not seem to me unequal to the pronf. Wie have numerous experiences in our own person of acts which are asseriated with states of consciousness, and we infer that similar acts in others are likely to be associated with similar states of comsciousness. But this argument from analoey is superior in one respect to nearly all other empirical arenments, and this superionity may possibly explain the great comfidence which we fred in it. Wer do seem in this case to have diect kmewhedere such as we have in no other case, that cour states of comsemmsmess are, semmetimes at least, causally connected with some of our acts. We do not, as in other cases,

1 Mr. C. D. Broad, in two articles " (On the Relation between Induction and Probability" (Mind, 1918 and 1920), has been following a mimilar line of thought.
merely observe invariable sequence or coexistence between consciousness and act; and we do believe it to be vastly improbable in the case of some at least of our own physical acts that they could have occurred without a mental act to support them. Thus, we seem to have a special assurance of a kind not usually available for believing that there is sometimes a necessary connection between the conclusion and the condition of the generalisation; we doubt it only from the possibility of a plurality of causes.

The objection to this argument on the ground that the analogy is always imperfect, in that all the observed connections of consciousness and act are alike in being mine, seems to me to be invalid on the same ground as that on which I have put on one side objections to future generalisations, which are based on the fact that the instances which support them are all alike in being past. If direct judgments of irrelevance are ever permissible, there seems some ground for admitting one here.
9. As a logical foundation for Analogy, therefore, we seem to need some such assumption as that the amount of variety in the universe is limited in such a way that there is no one object so complex that its qualities fall into an infinite number of independent groups (i.e. groups which might exist independently as well as in conjunction) ; or rather that none of the objects about which we generalise are as complex as this; or at least that, though some objects may be infinitely complex, we sometimes have a finite probability that an object about which we seek to generalise is not infinitely complex.

To meet a possible plurality of causes some further assumption is necessary. If we were content with Inductive Correlations and sought to prove merely that there was a probability in favour of any instance of the generalisation in question, without inquiring whether there was a probability in favour of every instance, it would be sufficient to suppose that, while there may be more than one sufficient cause of a character, there is not an infinite number of distinct causes competent to produce it. And this involves no new assumption; for if the aggregate varicty of the system is finite, the possible plurality of causes must also be finite. If, however, our generalisation is to be universal, so that it breaks down if there is a single exception to it, we must obtain, by some means or other, a finite probability that the set of characters,
which condition the ereneralisation, are not the possible effect of more than one distinet set of fumdamental properties. I do mot know upon what ground we could establish a linite probability to this effiet. The meressity for this semmingle arhitrary hypmthesis strongly sumests that our conclusions should be in the form of imductive comelations. rather than of universal emeralisations. Perhaps our enn ralisations should ahwas run: 'It is probable that any given \(\phi\) is \(f\), rather than, 'It is probable that all \(\phi\) are \(f\).' Certainly, what wi commonly seem to hold with convietion is the berlief that the sum will rise thememon, rather than the berlief that the sun will cheouss rise sol lone as the conditions explicitly known to us are fultilled. This will be matter for further disenssion in Part V .. when Inductive Correlation is specifically dealt with.
10. There is a vaurenss, it may be notiend, i., whe number of instances, which would be required on the abow assumptions to establish a given numerial degree of prombitilite, which corresponds to the vacueness in the degree of probability which we do actually attach to inductive conclusions. We assume that the neressary number of instances is linito. hut we do not know what the number is. We know that the probability of a well "stablishod imbuction is great, but, when we are asked to name its degree, we cannot. Common sense tells us that some inductive aremments are stromere than others, and that some are very strong. But how much stronger or how strong we cannot express. The probability of an induction is only numerically definite when we are able to make definite assumptions about the number of ind pomdent equiprobable intluemees at work. Otherwise, it is nom-mumerical, theweh bearing relations of ereater and liss to numerical probabilities ancordine to the approximate limits within which our assumption as to the possible number of these causes lies.
11. 'fin to this perint I hate suppersen, for the sake of simplicity, that it is mecessary to make our assumptions as to the limitation of independent varioty in an absolute form, to assimm, that is to say, the finitemess of the ssstem, to which the armument is applind, for certain. But we need not in fact go so far as this.

If our conelusion is C and our empirical evidence is E, then, in order to justify imdurtive m-thods, our premisses mast inchude, in addition to E, a er ural hypothesis II such that (H) the
à priori probability of our conclusion, has a finite value. The effect of E is to increase the probability of C above its initial à priori value, \(\mathrm{C} / \mathrm{HE}\) being greater than \(\mathrm{C} / \mathrm{H}\). But the method of strengthening \(\mathrm{C} / \mathrm{H}\) by the addition of evidence E is valid quite apart from the particular content of H . If, therefore, we have another general hypothesis \(\mathrm{H}^{\prime}\) and other evidence \(\mathrm{E}^{\prime}\), such that \(\mathrm{H} / \mathrm{H}^{\prime}\) has a finite value, we can, without being guilty of a circular argument, use evidence \(\mathrm{E}^{\prime}\) by the same method as before to strengthen the probability \(I / / H^{\prime}\). If we call \(I\), namely, the absolute assertion of the finiteness of the system under consideration, the inductive hypothesis, and the process of strengthening \(\mathrm{C} / \mathrm{H}\) by the addition E the inductive method, it is not circular to use the inductive method to strengthen the inductive hypothesis itself, relative to some more primitive and less far-reaching assumption. If, therefore, we have any reason \(\left(\mathrm{H}^{\prime}\right)\) for attributing à priori a finite probability to the Inductive Hypothesis (H), then the actual conformity of experience à posteriori with expectations based on the assumption of \(H\) can be utilised by the inductive method to attribute an enhanced value to the probability of H . To this extent, therefore, we can support the Inductive Hypothesis by experience. In dealing with any particular question we can take the Inductive Hypothesis, not at its di priori value, but at the value to which experience in general has raised it. What we require \(\grave{a}\) priori, therefore, is not the certainty of the Inductive Hypothesis, but a finite probability in its favour. \({ }^{1}\)

Our assumption, in its most limited form, then, amounts to this, that we have a finite à priori probability in favour of the Inductive Hypothesis as to there being some limitation of independent varicty (to express shortly what I have already explained in detail) in the objects of our generalisation. Our experience might have been such as to diminish this probability à posteriori. It has, in fact, been such as to increase it. It is because there has been so much repetition and uniformity in our experience that we place great confidence in it. To this extent the popular opinion that Induction depends upon experience for its validity is justified and does not involve a circular argument.

\footnotetext{
1 I have implecitly assumed in the above argument that if II' supports \(H\), it strengthens an "rrument which 11 would strengthen. This is not necessarily the case for the reasons wiven on pr .68 and 147 . In these passages the necessary conditions for the above are elucidated. I am, therefore, assuming that in the case now in question these conditions actually are fulfilled.
}
12. I think that this assumption is adequate to its purpose and would justify our ordinary ine thods of procedure in inductive argument. It was surfested in the previous chapter that our theory of Analogy ourhit to ber as applicable to mathomatical as to material inemralisations, if it is to justify commom sense. The above assumptions of the limitation of independent variely sufficently satisfy this condition. There is mothine i:t these assumptions which gives them a peruliar refernce to material whenets. Wi. Lelievi. in fact, that ail the properties of numbers can te deriond from a limited number of laws, and that the same s.t of laws povem- all numbers. To apply empirical metheds to such thines as numbers rembers it necessary: it is true, to maki an assumption abont the nature of numbers. But it is the same kind of assumption as we have to maker about material ohjects, and has just about as much, or as little, plausibility. There is no new difficulty.

The assumpion, also, that the system of Nature is tinite is in aceordane with the analysis of the underlying assumption of scientists, given at the close of the previous chapter. The hypothesis of atomis uniformity, as I have called it, while mot formally equivalent to the hypothesis of the limitation of imbl... pemdent variety, amounts to very much the same thime. If the. fundamental laws of comection chanmed altone ther with variations, for instanee, in the shape or size of boties, or if the laws governing the behasmur of a complex had mo relation whaterer (1) the laws erneminer the behaviour of its parts when belongine to other complexes, there cond bardly be a limitation of indwpendent variely in the sense in which this has beendetimed. And. on the wther hamd, a limitation of imbemement variety sisems ureessarily to carty with it some deren of atomice unifomity. The underlying conception as to the character of the syatem of Nature is in each case the same.
13. Wi have now reachent the last and most difticult stime of the discussion. The lumical part of our impuiry is compluy. and it has left us, as it is its business to leane us, with a gur-tion of "pistemolors. such is the promiss of assumption whith our logical processes need to work upon. What right have we to maker it? It is no sufficient answer in philnsuphy to phest that the assumption is after all a very little one.

1 do not belien that any conclusion or perfertly satinfactury
answer to this question can be given, so long as our knowledge of the subject of epistemology is in so disordered and undeveloped a condition as it is in at present. No proper answer has yet been given to the inquiry-of what sorts of things are we capable of direct knowledge? The logician, therefore, is in a weak position, when he leaves his own subject and attempts to solve a particular instance of this general problem. He needs guidance as to what kind of reason we could have for such an assumption as the use of inductive argument appears to require.

On the one hand, the assumption may be absolutely à priori in the sense that it would be equally applicable to all possible objects. On the other hand, it may be seen to be applicable to some classes of objects only. In this case it can only arise out of some degree of particular knowledge as to the nature of the objects in question, and is to this extent dependent on experience. But if it is experience which in this sense enables us to know the assumption as true of certain amongst the objects of experience, it must enable us to know it in some manner which we may term direct and not as the result of an inference.

Now an assumption, that all systems of fact are finite (in the sense in which I have defined this term), cannot, it seems perfectly plain, be regarded as having absolute, universal validity in the sense that such an assumption is self-evidently applicable to every kind of object and to all possible experiences. It is not, therefore, in quite the same position as a self-evident logical axiom, and docs not appeal to the mind in the same way. The most which can be maintained is that this assumption is true of some systems of fact, and, further, that there are some objects about which, as soon as we understand their nature, the mind is able to apprehend directly that the assumption in question is true.

In Chapter II. § 7, I wrote: "By some mental process of which it is difficult to give an account, we are able to pass from direct acquaintance with things to a knowledge of propositions about the things of which we have sensations or understand the meaning." Knowledge, so obtained, 1 termed direct knowledge. From a sensation of yellow and from an understanding of the meaning of 'yellow' and of 'colour,' we could, I suggested, have direct knowledge of the fact or proposition 'yellow is a colour ;' we might also know that colour cannot exist without extension, or that two colours cannot be perceived at the same
time in the same place. Other philosophers might use thrms differently and express themselves otherwise; but the sub)etance. of what 1 was there trying to saty is not very disputable. But when we come to the guestion as to what kinds of propesitions we can come to know in this manner, we enter upm an unexphored field where no certain opinion is discoverable.

In the case of logical terms, it seems to be generally ammend that if we umberstand their meaning we can know directly pre positions about them which go far beyond a mer expresion of this meanine: - propositions of the kind which some philusophers have termed synthetic. In the case of mom-logical on empirical entities, it seems sometimes to be assumed that our direct knowledge must be confined to what may be resardad as an expression or description of the meaning or sensation apprehended by us. If this view is correct the Inductive Hypothesis is not the kind of thine about which we can have direct knowledge as a result of our acquaintance with objects.

I suguest, however, that this view is incorrect, and that we are capable of direct knowledge about empirical entities which groes beyond a mere expression of our understanding or sthsation of them. It may be useful to give the reader two examples, more familiar than the Inductive Hypothesis, where, as it appears to me, such knowledge is commonly assumed. The first is that of the causal irrelevance of mere pesition in time and space. commonl? called the Uniformity of Nature. We do believe, and yet hater no adequate imductive reason whatever for bolieving, that mem position in time and space cannot make any differonce. This belicf arises directly, I think, out of our acequaintance with the objects of experience and our undurstanding of the conceptof 'time' and 'space.' The second is that of the Law of Causation. We believe that every object in time has is necessary ' comection \({ }^{1}\) with some set of objocts at a previons tim... This belief also, I think, arises in the same way. It is to boc noticed that neither of these beliefs chearly arises, in spite of the directness which may be clamed for therm, out of any one simgly expericnce. In a way analogous to these, the validity of assamine the Inductive Hypothesis, as applied to a particular class of objects, appears to me to be justified.

Our justification for using imductive methods in an argument

\footnotetext{
\({ }^{1}\) I do not propese to define the meaninf ot thas.
}
about numbers arises out of our perceiving directly, when we understand the meaning of a number, that they are of the required character. \({ }^{1}\) And when we perceive the nature of our phenomenal experiences, we have a direct assurance that in their case also the assumption is legitimate. We are capable, that is to say, of direct synthetic knowledge about the nature of the objects of our experience. On the other hand, there may be some kinds of objects, about which we have no such assurance and to which inductive methods are not reasonably applicable. It may be the case that some metaphysical questions are of this character and that those philosophers have been right who have refused to apply empirical methods to them.
14. I do not pretend that I have given any perfectly adequate reason for accepting the theory I have expounded, or any such theory. The Inductive Hypothesis stands in a peculiar position in that it seems to be neither a self-evident logical axiom nor an object of direct acquaintance ; and yet it is just as difficult, as though the inductive hypothesis were cither of these, to remove from the organon of thought the inductive method which can only be based on it or on something like it.

As long as the theory of knowledge is so imperfectly understood as now, and leaves us so uncertain about the grounds of many of our firmest convictions, it would be absurd to confess to a special scepticism about this one. I do not think that the foregoing argument has disclosed a reason for such scepticism. We need not lay aside the belief that this conviction gets its invincible certainty from some valid principle darkly present to our minds, even though it still eludes the peering eyes of philosophy.
\({ }^{1}\) Since numbers are logical entities, it may be thought less unorthodox to make such an assumption in their case.

\section*{('HAPTER XNHII}

\section*{SOME HISTORICAL NOTES ON INIOTCTION}
1. The number of books. which deal with inductive \({ }^{1}\) theorey is extramerdinarily small. It is usual to associate the subject with the names of Batem. Hume and Mill. In spite of the mendern tendeney to deperciate the first and the last of these, they are the principal names, I think. with which the history of induction ought to be asocriated. The next place is held by Laplace and Jevons. Amomst contmperary lowicians there is an almost complete absence of constructive theory, and they content themselves for the mest part with the easy task of eritici-ime Mill, or with the more difficult one of following him.

That the inductive theories of Bacon and of Mill are full of errors and "sen of ahsurdities, is, of course, a commmondace of eriticism. But when we ignore details, it beomes char that the: were really attomptime to disentangle the usimential isomes. Wi. depreciate them partly. pertaps, as a reaction from the view once held that they hellued the promeres of sementific dimensery. For it is mot plansihde to -uppose that Nowtonowed anythme facom. or Darwin tw. Mill. But with the lewical prohlem the ir mind were truly orcupied, and in the history of logical theory they should always be important.

It is true, nevertheless, that the advancement of science was the main wheet which Bacon himself, though not Mill, beliesod that his philesophy would prometr. The Cireat Instanrution was intended to promuleate an actual method of diemsery matieds different from any which had been provinusly known.2 It did

\footnotetext{
\({ }^{1}\) See note at the end of this chapter on "The Use of the Term Induction."

 the mariner's compass, until the discovery of which no widn sea could be crossed (see Spedding and Fillis, vol. i. p. 24).
}
not do this, and against such pretensions Macaulay's well-known essay was not unjustly directed. Mill, however, expressly disclaimed in his preface any other object than to classify and generalise the practices "conformed to by accurate thinkers in their scientific inquiries." Whereas Bacon offered rules and demonstrations, hitherto unknown, with which any man could solve all the problems of science by taking pains, Mill admitted that " in the existing state of the cultivation of the sciences, there would be a very strong presumption against any one who should imagine that he had effected a revolution in the theory of the investigation of truth, or added any fundamentally new process to the practice of it."
2. The theories of both seem to me to have been injured, though in different degrees, by a failure to keep quite distinct the three objects: (1) of helping the scientist, (2) of explaining and analysing his practice, and (3) of justifying it. Bacon was really interested in the second as well as in the first, and was led to some of his methods by reflecting upon what distinguished good arguments from bad in actual investigations. To logicians his methods were as new as he claimed, but they had their origin, nevertheless, in the commonest inferences of science and daily life. But his main preoccupation was with the first, which did injury to his treatment of the third. He himself became aware as the work progressed that, in his anxiety to provide an infallible mode of discovery, he had put forth more than he would ever be able to justify. \({ }^{1}\) His own mind grew doubtful, and the most critical parts of the description of the new method were never written. No one who has reflected much upon Induction need find it difficult to understand the progress and development of Bacon's thoughts. To the philosopher who first distinguished some of the complexities of empirical proof in a generalised, and not merely a particular, form, the prospects of systematising these methods must have seemed extraordinarily hopeful. The first investigator could not have anticipated that Induction, in spite of its apparent certainty, would prove so elusive to analysis.

Mill also was led, in a not dissimilar way, to attempt a too

\footnotetext{
1 This view is taken in the edition of James Spedding and Leslie Ellis. 'Iheir introductions to Bacon's philosophical works seem to me to be very greatly superior to the accounts to be found elsewhere. They make intelligible, what seems, according to other commentaries, fanciful and without sense or reason.
}
simple treatment, amd, in seeking for rase and cortaines, to treat far tow lighty the prohlem of justifying what her had claimed. Mill shirks, almost openly, the difficulties : and scarcely attempts to disguise from himself or his readers that he gromods induction upon a circular argument.
3. Fome of the most characteristic errors both of Bacon and of Mill arise. I think, out of a misapprethension, which it has beren a principal objoct of this book to correet. Both believed, without hesitation it semes. that imburtion is capable of establishime a comelusion which is ahsolutely cortain, and that an aroument is invalid if the cromeratisation, which it supports, admits of excoptions in fact. " Absolute cortainty," says Leslie Ellis, \({ }^{1}{ }^{\text {" }}\) is one of the distimenishing characters of the Bacomian induction." It was, in this respect, mainly that it improved upon the older induction per enumerutionem simplicem. "The induction which the lonicians spacak of," Bacon aroues in the Adencement of Learming, " is utterly vicious and incompetent. . . . For to conclude upon an cmumeration of particulars, without instance. contradictory, is no conclusion but a conjecture." The conclusions of the new method, unlike those of the old, are not liable to bee upset by further experience. In the attempt to justify these claims and to ohtain demonstrative methonds, it was necessary to introduce assumptions for which there was no warrant.

Precisely similar claims were made by Mill, although theme are passunes in which he abates them, \({ }^{2}\) for his own rules of procedure. In induction has Im, validity, according to him as accordine to Bacon, unless it is absolutely erertain. The followinge pasare \({ }^{3}\) is significant of the spirit in which the subject was approached by him: "Let us compare a few cases of incorrect inductions with others which are acknowledered to be legitimate. Some, we know, which were believed for centuries to be correct, *were nevertheless incorrect. That all swans are white, onnout have bren a good induction, since the comclusion hus thrnel out ctroncous. The expertience, however, on which the conclusion rested was emmine." Mill has mot justly apprehemded the relativity of all inductive arguments to the evidence, nor the rfment of uncertainty which is prisment, mome

\footnotetext{
1 11p. id. val. 1. f. : : :

s lik. iii. 'hap. iii. 3 (the italii s are :anい.).
}
or less, in all the generalisations which they support. \({ }^{1}\) Hill's methods would yield certainty, if they were correct, just as Bacon's would. It is the necessity, to which Mill had subjected himself, of obtaining certainty that occasions their want of reality. Bacon and Mill both assume that experiment can shape and analyse the evidence in a manner and to an extent which is not in fact possible. In the aims and expectations with which they attempt to solve the inductive problem, there is on fundamental points an unexpectedly close resemblance beween them.
4. Turning from these general criticisms to points of greater detail, we find that the line of thought pursued by Mill was essentially the same as that which had been pursued by Bacon, and, also, that the argument of the preceding chapters is, in spite of some real differences, a development of the same fundamental ideas which underlie, as it seems to me, the theories of Mill and Bacon alike.

We have seen that all empirical arguments require an initial probability derived from analogy, and that this initial probability may be raised towards certainty by means of pure induction or the multiplication of instances. In some arguments we depend mainly upon analogy, and the initial probability obtained by means of it (with the assistance, as a rule, of previous knowledge) is so large that numerous instances are not required. In other arguments pure induction predominates. As science advances and the body of pre-existing knowledge is increased, we depend increasingly upon analogy ; and only at the earlier stages of our investigations is it necessary to rely, for the greater part of our support, upon the multiplication of instances. Bacon's great achievement, in the history of logical theory, lay in his being the first logician to recognise the importance of methodical analogy to scientific argument and the dependence upon it of most wellestablished conclusions. The Novum Organum is mainly concerned with explaining methodical ways of increasing what I have termed the Positive and Negative Analogies, and of avoiding false Analogies. The use of exclusions and rejections, to which
\({ }^{1}\) This misapprehension may be connected with Mill's complete failure to grasp with any kind of thoroughness the nature and importance of the theory of probability. The treatment of this topic in the System of Logic is exceedingly bad. His understanding of the subject was, indeed, markedly inferior to the best thought of his own time.

Bacon attached supreme importance, and which he held to constitute the essential superiority uf his method over thoser which preceded it, entirely consists in the determination of what characters (or natures as her would call them) lielong to the positive and negative analogies respertioly. The first two tables with which the investigation theins are tirst. the table essentiae et pruesentine, which contains all known instances in which the given nature is preant, and second, the table declinatiomis sie absentive in prorime, which contains instances comperponding in each case to thone of tle first table, but in which, not withstandines this comerespmener. the evisen mature is ahsent. \({ }^{1}\) The doctrine of prerogative instances is concerned no less plainly with the methomical dutermination of Amalogy. And the doctrine of idels is expounded for the aroidance of folse analogios, standines. he says, in the same relation to the interpretation of Nature, as the doetrine of fallacies to ordinary loric. \({ }^{2}\) Baconis error lay in supposine that, hecause these methods were new to logic, they were therefore bew to practice. He exageverated also their precision and their cortainty: and he und restimated the impertance of pure indurtion. But there was, at bottom, mothing about his rules impracticable or fantastic, or indeed unusual.
5. Almon the whole of the precodine paragraph is equally applicable to Mill. He atred with Bacon in deprectiating the part phayed in scimtifie inquiry by pure induction, and in emphasising the importane of analugy to all systmatice investigators. But he saw further than Bacon in allowing for the Plurality of Causes, and in admitting that an element of pure induction was therefore made necessary. "The Plurality of Causes," he says, \({ }^{3}\) " is the only reason why mere number of instances is of any importance in inductive inquirs. The tembeney of unscimitife inguirers is (1) mely tow much on mumber, without. analysine the instances. . . Most prophe hold the ir conclusions with a derwe of as arame propertioned th the mere mass of the experinnce on which they appear to rest: mot considurine that. hy the addition of instances to instances, all of the same kind, that is, difierine from one another only in points alroady recomnised as immaterial, nothine whaterer is added to the whidene of

\footnotetext{


- Mant an imy a
}
the conclusion. A single instance eliminating some antecedent which existed in all the other cases, is of more value than the greatest multitude of instances which are reckoned by their number alone." Mill did not see, however, that our knowledge of the instances is seldom complete, and that new instances, which are not known to differ from the former in material respects, may add, nevertheless, to the negative analogy, and that the multiplication of them may, for this reason, strengthen the evidence. It is easy to see that his methods of Agreement and Difference closely resemble Bacon's, and aim, like Bacon's, at the determination of the Positive and Negative Analogies. By allowing for Plurality of Causes Mill advanced beyond Bacon. But he was pursuing the same line of thought which alike led to Bacon's rules and has been developed in the chapters of this book. Like Bacon, however, he exaggerated the precision with which his canons of inquiry could be used in practice.
6. No more need be said respecting method and analysis. But in both writers the exposition of method is closely intermingled with attempts to justify it. There is nothing in Bacon which at all corresponds to Mill's appeals to Causation or to the Uniformity of Nature, and, when they seek for the ground of induction, there is much that is peculiar to each writer. It is my purpose, however, to consider in this place the details common to both, which seem to me to be important and which exemplify the only line of investigation which seems likely to be fruitful ; and I shall pursue no further, therefore, their numerous points of difference.

The attempt, which I have made to justiiy the initial probability which Analogy seems to supply, primarily depends upon a certain limitation of independent variety and upon the derivation of all the properties of any given object from a limited number of primary characters. In the same way I have supposed that the number of primary characters which are capable of producing a given property is also limited. And I have argued that it is not easy to see how a finite probability is to be obtained unless we have in each case some such limitation in the number of the ultimate alternatives.

It was in a mamer which bears fundamental resemblances to this that Bacon endeavoured to demonstrate the cogency of his method. He considers, he says, " the simple forms or differ-
ence of things which are few in number, and the dogrees and co-ordinations whereof make all this varicty." And in Colerins Terminus he argues "that every particular that worketh any effect is a thing compoumded more or less of diverse single natures, more manifest and more obscure, and that it appeareth mot to which of the natures the effect is to be aseribed." \({ }^{1}\) It is indeed essential to the method of exclusions that the matter to which it is applied should be somethew resolvable into a finite number of elements. But this assumption is not preculiar, I think, to Bacon's methoxl, and is involved, in some form or other, in every argument from Analogy. In making it Bacon was initiating, perhaps whscurely: the modern conception of a tinite number of laws of nature out of the combinations of which the almost boundless variety of experience ultimately arises. Bacon's error was double and lay in supposing, first, that these distinct elements lie upon the surface and consist in visible characters, and second, that their natures are, or casily (an be, known to us, although the part of the Instuurution, in which the manner of conceriving simple natures was to be explained, he never wrote. These beliefs falsely simplified the problem as he saw it, and led him to exaggerate the ease certainty, and fruitfulness of the new method. But the view that it is possible to reduce all the phenomena of the universe to combinations of a limited number of simple elements-which is, according to Ellis. \({ }^{2}\) the central point of Bacon's whole system-was a real contribution to philosophy.
7. The assumption that every event can be analysed into a limited number of ultimate elements, is never, su far as I am aware, explicitly ayowed by Mill. But he makes it in almost every chapter, and it underlies, throughout, his mode of procedure. His methods and arguments would fail immediately, if wir were to suppose that phenomena of infinite complexity, due to an infinite number of independent elements, were in question, ur if an infinite plurality of causes had to be allowed for.

In distinguishing, therefore, analogy from pure imduction, and in justifying it by the assumption of a limilel complexity in the problems which we investigate, I am, I think, pursumes. with mumerous differeners, the line of thought which Bacon tirst

\footnotetext{

\({ }^{2}\) Vol. i. p. 28.
}
pursued and which Mill popularised. The method of treatment is dissimilar, but the subject-matter and the underlying beliefs are, in each case, the same.
8. Between Bacon and Mill came Hume. Hume's sceptical criticisms are usually associated with causality ; but argument by induction -inference from past particulars to future generalisations - was the real object of his attack. Hume showed, not that inductive methods were false, but that their validity had never been established and that all possible lines of proof seemed equally unpromising. The full force of Hume's attack and the nature of the difficulties which it brought to light were never appreciated by Mill, and he makes no adequate attempt to deal with them. Hume's statement of the case against induction has never been improved upon; and the successive attempts of philosophers, led by Kant, to discover a transcendental solution have prevented them from meeting the hostile arguments on their own ground and from finding a solution along lines which might, conceivably, have satisfied Hume himself.
9. It would not be just here to pass by entirely the name of the great Leibniz, who, wiser in correspondence and fragmentary projects than in completed discourses, has left to us sufficient indications that his private reflections on this subject were much in adrance of his contemporaries'. He distinguished three degrees of conviction amongst opinions, logical certainty (or, as we should say, propositions known to be formally true), physical certainty which is only logical probability, of which a well-established induction, as that man is a biped, is the type, and physical probability (or, as we should say, an inductive correlation), as for example that the south is a rainy quarter.' He condemned generalisations based on mere repetition of instances, which he declared to be without logical value, and he insisted on the importance of Analogy as the basis of a valid induction." He regarded a hypothesis as more probable in proportion to its simplicity and its power, that is to say, to the number of the phenomena it would explain and the fewness of the assumptions it involved. In particular a power of accurate prediction and of explaining phenomena or experiments pre-

\footnotetext{
1 ('outurat, ()pussenles et frumments initits de Leibmi:, p. 232.
\({ }^{2}\) Couturat, La Logique de Leibniz d'après des documents inédits, pp. \(\therefore 292,247\).
}
viousty untriod is a ju-t ground of samure contidence of which he cites as a nearle perfect "xample the key to a eryptereram.'
10. Whewell and Joroms furnished logicians with a storehomse of "xamples derived from the practice of srimotists Jovons, partly anticipated hy Laplace, madn an impertant advance when he emphasised the close relation between Induction and Probability. Combinins insight and orror, he. spoilt brilliant sugesti me herratic and atrocions armuments. Ilis application of Inverse Prombibity to the inductive problem is crude and fallacious, but the idea which underlies it is substantially good. He, too, made explicit the element of Analory, which Mill, thourh her constantly emplowed it, hat scldom called by its right name. There are few books, so superficial in argument yet sumerating en much truth as J.woms's Principles of Science.
11. Mondern text-books wan Laric all contain their chapters on Induction. but contribute little th the suljeect. Their recomis tion of Dill's inadequace rembers their expesition, which, in spit. of criticisms. is enemerally ahme his limes. mervelus and confused. Where Mill is clear and offers a solution, they, confusedlycriticising, must withmal ome. The best of them. siewart and Vemn, contain criticism and disenssion which is intereatines. hut constructive theory i hace ime. Hitherto Hum has heen master only to be refuted in the manner of Diogenes or Dr. Johnson.

\footnotetext{
\({ }^{1}\) Letter to Conring, 19th March 1678.
}

\section*{NOTES ON PART III}

\section*{(i.) On the Use of the Term Induction}
1. Induction is in origin a translation of the Aristotelian \(\mathfrak{\epsilon} \pi \omega \gamma \omega \gamma\) í. This term was used by Aristotle in two quite distinct senses-first, and principally, for the process by which the observation of particular instances, in which an abstract notion is exemplified, enables us to realise and comprehend the abstraction itself ; secondly, for the type of argument in which we generalise after the complete enumeration and assertion of all the particulars which the generalisation embraces. From this second sense it was sometimes extended to cases in which we gencralise after an incomplete enumeration. In post-Aristotelian writers the induction per emmerationem simplicem approximates to induction in Aristotle's second sense, as the number of instances is increased. To Bacon, therefore, "the induction of which the logicians speak " meant a method of argument by multiplication of instances. He himself deliberately extended the use of the term so as to cover all the systematic processes of empirical generalisation. But he also used it, in a manner closely comresponding to Aristotle 's fivet use, for the process of forming scientific conceptions and correct notions of "simple natures." \({ }^{1}\)
2. The modern use of the term is derived from Bacon's. Mill defines it as "the operation of discovering and proving general propositions." His philosophical system required that he should define it as widely as this; but the term has really been used, both by him and by other logicians, in a narrower sense, so as to cover those methods of proving general propositions, which we call empiri(alal, and so as to exclude gemeralisations, such as those of mathematies, which have been proved formally. Jevons was led, partly by the linguistic resemblance, partly because in the one case we proceed from the particular to the peneral and in the other from the general to the particular, to define Induction as the inverse process of Deduction. In contemporary logic Mill's use prevails; but there
\({ }^{1}\) See Ellis's edition of Bacon's Work's, vol. i. p. 37. On the first occasion on which Iuduction is mentioned in the Novum Organum, it is used in this secondary sense.
is, at ther satme time, a because Bacon and Mill never quite freed themselves from it - of argument by mere multiplication of instances. I have thought it best, therefore, to use the term pure induction to describe arguments which are based upon the mumler of instances, and to use imduction itself for all those types of arguments which combine in one form or another, pure induction with analogy:

\section*{}
1. Throughout the preceding argument, as well as in Part II. I have been able to avoid the metaphysical difficulties which survound the true meaning of couse. It was not necessary that I should inquire whether I meant by causal connection an invariable connection in fact merely, or whet her some more intimate relation was involved. It has also been convenient to speak of causal relations between objects which do not strictly stand in the position of cause and effect, and even to speak of a probuble couse. where there is no implication of necessity and where the antecedents will sometimes lead to particular consequents and sometimes will not. In makine this use of the term, I have followed a practice not uncommon amomgst writers on probability, who constantly use the term conser. where hypothesis might seem more appropriate. \({ }^{1}\)

One is led, almost inevitably, to use "cause "more widely than 'sufficient cause ' or than 'necessary cause,' because, the neressary causation of particulars by particulars being rarely apparent to us, the strict sense of the term has little utility. Those antecedent circumstances, which we are usually content to accept as causes, ate only so in strictness under a favourable conjunction of imnumerable other influmence.
2. As our knowledge is partial, there is constantly, in our use of the term cause, some reference implied or expressed to a limitent body of knowledge. It is clear that, whether or not, as Cournot? maintains, there are such things as independent series in the order of causation, there is often a sense in which we may hold that there is a closer intimacy betwern some series than between others. This intimacy is relative, I think, to particular information, which is actualls known to us, or which is within our reach. It will ber useful, therefore, to give precise definitions of these wider semses in which it is oftem comvenient to use the expression conese





We must first distinguish betwem assertions of law and assertions of fact, or, in the terminology of Von Kries, \({ }^{1}\) between nomologic and ontologic knowledge. It may be convenient in dealing with some questions to frame this distinction with reference to the special circumstances. But the distinction generally applicable is between propositions which contain no reference to particular moments of time, and existential propositions which cannot be stated without reference to specific points in the time series. The Principle of the Uniformity of Nature amounts to the assertion that natural laws are all, in this sense, timeless. We may, therefore, divide our data into two portions \(k\) and \(l\), such that \(k\) denotes our formal and nomologic evidence, consisting of propositions whose predication does not involve a particular time reference, and \(l\) denotes the existential or ontologic propositions.
3. Let us now suppose that we are investigating two existential propositions \(a\) and \(b\), which refer two events A and B to particular moments of time, and that A is referred to moments which are all prior to those at which B occurred. What various meanings can we give to the assertion that A and B are causally connected?
(i.) If \(b / a k=1, \mathrm{~A}\) is a sufficient cause of B . In this case A is a cause of B in the strictest sense. \(b\) can be inferred from \(a\), and no additional knowledge consistent with \(k\) can invalidate this.
(ii.) If \(b / i k=0, A\) is a necessary cause of \(B\).
(iii.) If \(k\) includes all the laws of the existent universe, then A is not a sufficient cause of B unless \(b / a k-1\). The Law of Causation, therefore, which states that every existent has to some other previous existent the relation of effect to sufficient cause, is equivalent to the proposition that, if \(k\) is the body of natural law, then, if \(b\) is true, there is always another true proposition \(a\), which asserts existences prior to B, such that \(b / a k=1\). No use has been made so far of our existential knowledge \(l\), which is irrelevant to the definitions preceding.
(iv.) If \(b / a k l=1\) and \(b / k l \neq 1, \mathrm{~A}\) is a sufficient cause of B under conditions \(l\).
(v.) If \(b / a k l=0\) and \(b / k l \neq 0, \mathrm{~A}\) is a necessary cause of B under conditions \(l\).
(vi.) If there is any existential proposition \(h\) such that \(b / a h k=1\) and \(b / h k \neq 1, \mathrm{~A}\) is, relative to \(k\), a possible sufficient cause of B.
(vii.) If there is an existential proposition \(h\) such that \(b / i h k=0\) and \(b / h k \neq 0, \mathrm{~A}\) is, relative to \(k\), a possible necessary cause of B .
(viii.) If \(b / a h k l=1, b / h k \neq 1\), and \(h / a k l \neq 0, \mathrm{~A}\) is, relative to \(k\), a possible sufficient cause of B under conditions \(l\).
(ix.) If \(b / \sim h k l ~ 0, b / h / l+0, h / \omega k l \neq 0\), and \(h / a k l \neq 0, \mathrm{~A}\) is, relative to \(k\), a possible necessary cause of B under conditions \(l\).

\footnotetext{
\({ }^{1}\) Die Principien der Wahrscheinlichkeitsrechnung, p. 86.
}

Thus an event is a pussible memsary cause of another, relative to siven nombluyic clata, if cincmatances can arise, mot inconsistent with our rxistential data, in which the first event will be indiopersathe if the second is to occur.
(x.) Two events are causally independent if no part of either is, relative to our nomologic data, a possible cause of any part of the it her und of the comditins of our existential knowlentre. The ereater the scope of our "aistential hmowheder, the greater is the likelihoul of our being able to pronounce events causally dependent or independent.
4. These definitions preserve the distinction between 'causally independent' and 'independent for probability,'-the distinction
 where \(a\) and \(b\) may be any propositions whatever and are not limited as they were in the causal definitions, we have 'dependence for probahility,' and \(a\) is a causu comnoscendi for \(b\), relative to data kl . If \(a\) and \(b\) are causally dependent, according to definition (x.), \(b\) is as possible causa essendi, relative to data kl .

But, after all, the essential relation is that of 'independence for probability.' We wish to know whether knowledge of one fact throws light of any kind upon the likelihood of another. The theory of causality is only important because it is thought that by means of its assumptions light can be thrown by the experience of one phenomenon upon the expectation of another.

\section*{PARTIV}

\section*{
}

\section*{('HAPTER XXIV}

THE MEANINGS OF OBJECTIVE CHANCE, ANI) OF RANDOMNESS
1. Many important differences of opinion in the treatment of Probability have been due to confusion or varueness as to what is meant by Randomness and by Objective Chance, as distinguished trom what, for the perpeses of this dhemer, nay bue termed Subjective P'robability. It is agreed that there is a sort
 is relative, in some manner, to the mind of the subjoct; but it is
 is not thus dependent, or less completely so, though precisely what this conception stands for is not plain. The relation of Randommess to the other coneepts-is ako ohecure. The prohtmom of charing up these distinctions is of impertance if wr ate to criticise certain schmots of opinion intelliqutly, as weil as to the treatment of the foundations of statistical Inerener which is t" be attempted in Part V.

There are at last throw distinct isours to be kept apart. There is the antithesis betwern kumblede and immener. botwem events, that is to say, which we have some reason to expect, and events which we have no reason to expect, which gives rise to the theory of subjectite probability and subjectix chate... : amil. connected with this, the distinction between 'random' selection
 whertice chane. which ane as set Herure, hot which are com. monly held to arise out of the antithesis between 'canse' and 'chance,' between events, that is to say, which are causally connected and events which are not causally connected. And there
 blind causes' and 'final causes,' where we oppose a 'chanee '
event to one, part of whose cause is a volition following on a conscious desire for the event. \({ }^{1}\)
2. The method of this treatise has been to regard subjective probability as fundamental and to treat all other relevant conceptions as derivative from this. That there is such a thing as probability in this sense has been admitted by all sensible philosophers since the middle of the eighteenth century at least. \({ }^{2}\) But there is also, many writers have supposed, something else which may be fitly described as objective probability ; and there is, besides, a long tradition in favour of the view that it is this (whatever it may be) which is logically and philosophically important, subjective probability being a vague and mainly psychological conception about which there is very little to be said.

The distinction exists already in Hume: "Probability is of two kinds, either when the object is really in itself uncertain, and to be determined by chance; or when, though the object be already certain, yet 'tis uncertain to our judgment, which finds a number of proofs on each side of the question." \({ }^{3}\) But the distinction is not elucidated, and one can only infer from other passages that Hume did not intend to imply in this passage the existence of objective chance in a sense contradictory to a determinist theory of the Universe. In Condorcet all is confused ; and in Laplace nearly all. In the nineteenth century the distinction begins to grow explicit in the writings of Cournot. "Les explications que j'ai données . . .," he writes in the preface to his Exposition," sur le double sens du mot de probabilité, qui tantôt sir rapporte à une certaine mesure de nos conmaissances, et tantôt à une mesure de la possibilité des choses, indépendamment de la connaissance que nous en avons: ces explications, dis-je, me semblent propres à resoudre les difficultés qui ont rendu jusqu’ici suspecte à de bons esprits toute la théorie de la probabilité mathomatique." It will be worth while to pause for a moment to consider the ideas of Cournot.

\footnotetext{
\({ }^{1}\) This is discussed in Chapter XXV. \& 4.
\({ }^{2}\) D'Alembert, collecting (largely from Hume, many passages being translated almost verbatim) in the Encyclopédie méthodique the most up-to-date commonplaces of the subject, found it natural to write: "Il n'y a point de hasard à proprement parler; mais il y a son équivalent: l'ignorance, où nous sommes des vraies causes des événemens, a sur notre esprit l'influence qu'on suppose au hasard." Compare also the sentences from Spinoza quoted on p. 117 above.
\({ }^{3}\) A Treatise of Human Nature, Book ii. part iii. section ix.
}
3. (ournot. While admittine that there is -uch at thine as mh.
 is merely the offspriner of ignorance, saying that in this case "le calcul des chances" is merely "un caleul des illusions." The chance, upon which " le calcul des chances" is based, is something different, and depends, according to him. on the combination or eonvergence of phenomena belonging to independent series. By "independent series "he means series of phenomena which develop) as parallel or sucesssive series without any causal interdependence or link of solidarity whatever. \({ }^{1}\) No one, he says by way of example, seriously believes that in striking the ground with his foot he puts out the navigator in the Antipodes, or disturbs the system of Jupiter's satellites. Separate trains of events, that is to saly, have been set groing hy distinct initial acts of creation, so to speak. \({ }^{2}\) Every event is causally connected with previous evonts belonging to its own series, but it cannot be modified by contact with events belonging to a different series. A'chance' event is a complex due to the concurrence in time or place of events belomgine to cansally independent series.

This theory, as it stands, is eridently unsatisfactory. Even if there are series of phenomena which are independent in ('ournot's sense, it is not clear how we can know which they are, or how we can set up a calculus which presumes an aequaintance with them. Just as it is likely that we are all cousins if we go back far enough, so there may be, after all, remote relationships between ourselves and Jupiter. A remote commetion or a reaction quantitatively small is a matter of degree and not by any means the same thing as absolute independenee. Nevertheless Cournot has contributed something, I think, to the stock of our ideas. He has

\footnotetext{

 est cotle de la combinaison entre plusients siries de catuses on de fates qui se


 renferme "futlyue chase de rial et de pesitif, savoir un concours de deux out







}
hinted at, even if he has not disentangled, one of the elements in a common conception of chance ; and of the notion, which he seems to have in his mind, we must in due course take account. \({ }^{1}\)
4. In the writings of Condorent, I have said above, all is confused. But in Bertrand's criticism of him a relevant distinction, though not elucidated, is brought before the mind. "The motives for believing," wrote Condorcet, " that, from ten million white balls mixed with one black, it will not be the black ball which I shall draw at the first attempt is of the same kind as the motive for believing that the sun will not fail to rise to-morrow." "The assimilation of the two cases," Bertrand writes in criticism of the above, \({ }^{2}\) " is not legitimate : one of the probabilities is objective, the other subjective. The probability of drawing the black ball at the first attempt is \(10,000,000\), neither more nor less. Whoever evaluates it otherwise makes a mistake. The probability that the sun will rise varies from one mind to another. A scientist might hold on the basis of a false theory, without being utterly irrational, that the sun will soon be extinguished; he would be within his rights, just as Condorcet is within his ; both would exceed their rights in accusing of error those who think differently." Before commenting on this distinction, let us have before us also some interesting passages by Poincaré.
5. We certainly do not use the term ' chance,' Poincaré points out, as the ancients used it, in opposition to determinism. For us therefore the natural interpretation of 'chance' is subjective, -"Chance is only the measure of our ignorance. Fortuitous phenomena are, by definition, those. of the laws of which we are

\footnotetext{
\({ }^{1}\) Cournot's work on Probability has been highly praised by authorities as diverse and distinguished as Boole and Von Kries, and has been made the foundation of a school by some recent French philosophers (see the special number of the Revue de métaphysique et de morale, devoted to Cournot and published in 1905, and the bibliography at the end of the present volume passim). The best account with which I am acquainted, of Cournot's theory of probability, is to be found in A. Darbon's Le Concept du hasard. Cournot's philosophy of the subject is developed, not so much in his Exposition de la théorie des chances, as in later works, especially in his Essai sur les fondements de nos connaissances. Cournot never touched any subject without contributing something to it, but, on the whole, his work on Probability is, in my opinion, disappointing. No doubt his Exposition is superior to other French text-books of the period, of which there is so large a variety, and his work, both here and elsewhere, is not without illuminating ideas: but the philosophical treatment is so confused and indefinite that it is difficult to make much of it beyond the one specific point treated above.
\({ }^{2}\) Calcul des probabilités, p. xix.
}
ignorant." But Poincaré immediately adals: "Is this derinition very satisfactory ? Whon the first Chakdaran shepherds follow ed with their eyes the movements of the stars, they did not yot know the laws of astronomy, hut would they have dreaned of saviner that the stars move he chance? If a modern physicist is studying a now phemomemon, and if he discovers its law om Tuestay, would he have said on Monday that the phenomenon Wa* fortuitmas ? " 1

There is also another type of case in which "chance must be something more than the name we give to our ignorance." Amone the phemmenta, of the canse of which we are ignorant, there are some, su-h as thmse dualt with by the manager of a lif. insurane company, ahout which the calculus of probatilities can give real information. Surely it cannot be thanks to our ignorance, Poincare ure-s, that we are able to arriwe at valuable comelusions. If it were, it would be necessary to answer an inquirer thus : " You ass: me to mediet the phomememathat will ber produrad. If I had the misiomune to know the laws of thes phemomena, I could not sucemen wempt by inextricable calculations. and I shmald have to give up the attempt to answer you; but since I am fortunate whugh tw lew immant of them, I will give you an answer at once. And, what is more extraordinary still, my answer will be right." The ignorance of the manager of the life insurance company as to the prospects of life of his individual policyholders does not mexent his brine able to pay dividemeds to his shareholders.

Both these distinctions seem to be real ones, and Poincaré proceds to examine further instances in which we seem to
 are not due to 'chance.' He takes the case of a cone balanced upon its tip ; we know for certain that it will fall, but not on which side-chance will determine. "A very small canse which
 fail to see, and then we say that that effect is due to chance." The wather, and the distrithation of the mimer flame son the Zodiac, are analogous instances. And what we term 'games of chance 'afford, it has always been reconmisand, an almon t. perfect

\footnotetext{

 the English translation of which I have made use above, at the cont of doing

}
example. "It may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes imposible and we have the fortuitous phenomenom." "' The greatest chance is the birth of a great man. It is only by chance that the menting oceurs of two genital cells of different sex that contain precisely, each on its side, the mysterious elements, the mutual reaction of which is destined to produce genius. . . . How little it would have taken to make the spermatozoid which carried them deviate from its course. It would have been enough to deffect it a hundredth part of an inch, and Napoleon would not have been born and the destinies of a continent changed. No example can give a better comprehension of the true character of chance."

Poincaré calls attention next to another class of events, which we commonly assign to * chance, 'the distinguishing characteristic of which seems to be that their causes are very numerous and complex, - the motions of molecules of gas, the distribution of drops of rain, the shoffling of a pack of cards, or the errors of observation. Thirdly there is the type, usually comected with one of the first two, and specially emphasised. as we have seen above, by cournot, in which something comes about through the concurrence of erents which we regard as belongine to distinct causal trains, - a man is walking along the street and is killed by the fall of a tile.
6. When we attribute such events, as those illustrated by Poincaré, to chance, we certainly do not man merely to assert that we do not know how they arose or tiat we had no special reason for anticipating them it pirri. So far from this being the case, we mean to make a definite assertion as to the kind of way in which they arose ;-though exactly what we mean to assert about them it is extremely difficult to say.

Now a careful examination of all the cases in which various writers claim to detect the presence of objoctive chance "confirms the view that 'subjective chance. which is concerned with knowledre and ignorance is fundamental, and that so-called 'objective chance,' however important it may turn out to be from the practical or scientifie point of view, is really a special kind of 'subjective chance 'and a derivative type of the latter. For none of the adherents of 'objective chance' wish to question
 of this objective chance of theirs seems always to depend on the possibility that a particular kind of knowledge either is ours or is within our powers and capacity. Let me try to distinguish as exactly as I can the criterion of objective chance.
7. When we say that an event has happened by chance, we do not mean that previous to its occurrence the event was, on the available evidence, very improbable: this may or may not have been the case. We say, for example, that if a coin falls heads
 The term 'by chance' has reference rather to the state of our information about the concurrence of the event considered and the event premised. The fall of the coin is a chance event if our knowledge of the circumstances of the throw is irrelecant to our expectation of the posible alternative results. If the number of alternatives is very large, then the occurrence of the event is not only subject to chance but is also very improbable. In general two events may be said to have a chance connection, in the subjective sense, when knowledqe of the first is irrelevant to our expectation of the second, and produces no additional presumption for or against it ; when, that is to say, the probabilities of the propositions asserting them are independent in the sense defined in Chapter XII. § 8.

The above definition deals with chance in the widest sense. What is the differentin of the narrower group of cases to which it is desired to apply the term 'objective chance '? The occurrence of an event may be said to be subject to objective chance. I think, when it is not only a chance went in the above sense, but when we also have groed reason to suppose that the addition of further knowledge of a given kind, if it were procurable. would not affect its chance character. We must consider, that is onsay, the probability which is relative not to actual knowledse but to the uhole of a certuin liand of knowledge. Wie may be able to infer from our evidence that, even with certain kinds of additions to our knowledge, the connections between the events would still be subject to chance in the sense just defined, and we may be able to infer this without actually having the addi tional information in question. If, however complete our knowledge of certain kinds of things might be, ther" would still

of which we are investigating, then we may say there is an , bjective sense in which the actual conjunction of these propositions is due to chance.
8. This is, I think, the richt line of inquiry. It remains to decide, what kinds of information must be irrelevant to the connection, in ordwr that the presence of objective chance may be established.

When we attribute a coincidence to objective chance, we mean not only that we do noit actrally know a law of connection, but, speaking roughly; that there is no law of connection to be known. And when we say that the occurrence of one alternative rather than another is due to chance, we mean not only that we know no principle by which to choose between the alternatives, but also that no such principle is knowable. This use of the term closely corresponds to what Venn means by the term ' casual': "We call a coincidence casual, I apprehend, when we mean to imply that no knowledge of one of the two elements, which we can suppose to be practically attainable, would enable us to expect the other." \({ }^{1}\)

To make this more precise, we must revive our distinction, \({ }^{2}\) between nomoloric knowledge and ontologic knowledge, between knowledge of laws and knowledge of facts or existence. Given eritain facto \(f(r)\) about \(a\) and certain laws of connection, L , we can infer certainly or probably other facts \(\phi(a)\) about \(a\). If a complete knowludge of laws of connection together with \(f(a)\) yields no appreciable probability for preferring \(\phi(a)\) to other alternatives, then I suggest that an actual connection between \(\phi\) and \(f\) in a particular instance may be said \(t\), be due to chance in a sense which usage justifies us in calling objective. We do not, in fact, when we speak of objective chance, always use it in sol strict a sense as this, but this. is, I think, the underlying conception to which current usaqe approximates. Current usage diveres from this sense mainly for two reasons. We speak of objective chance if in the above conditions our grounds for preferesce, though appreciable, are very small ; and we are not insistent to assert the rule of chance if a comparatively slight addition to our ontologic knowledge would render the probability or the yrounde for preference appreciable.

\footnotetext{
\({ }^{1}\) Logic of Chance, p. 245.
}
\({ }^{2}\) See Part III. Note (ii.) \(\leqslant 2\), p. 275.

To sum up the above, an event is due to objective chance if in order to predict it, or to prefer it to alternatives. at present equi proimble. wity any hiel degrem of probability. it womb hos necessary to know a great many more facts of existence about it than we actually do know, and if the addition of a wide knowledge of general principles would be little use.

It must be added that we make a distinction botween facts of existence which are highly variablo from case to case and those which are constant or mearly constant over a certain field of observation or experience. Within the limits of this field we regard the permanent facts of existrnce as beins. from the standpoint of chance, in nearly the same position as laws. A comection is not dur to chance, therefore, if a knowledge of the permanent facts of existence could load to their prediction.

To sum up again therefore, -if within a given field of observation or experience a knowledqu of those facts of existence which are permanent or invariable within that field, tnether with a knowledge of all the relevant fundamental causal laws or general principles, and of a fex other facts of existence, would not permit us, given \(f(a)\), to attribute an appreciable probability to \(\phi(a)\) (or an appreciable probability to the alternative \(\phi_{1}(q)\) rather than \(\left.\phi_{2}(a)\right)\); then the conjunction of \(\phi(a)\) (or of \(\phi_{1}(a)\) rather than \(\phi_{2}(a)\) with \(\left.f(n)\right)\) is due to objective chance.
9. If we return to the examples of Poincaré, the above definition appars to common antivnctorily with the mathen common sense. It is when an exact knowledge of fact, as distinguished from priuciple, is requed tor c.e.en approximut. prodiotion tha: the expression 'objective chance' seems applicable. But neither our definition nor usage is precise as to the amount of knowloder of hact which must low tequired for prodimions. in order that, in the absence of it, the event may be regarded as subject to objective chance.

It may be added that the expression 'chance' can be used with reference to meneral statements as well as to particular facts. We say, for "xample, that it is a matter of chance if a man dies on his birthday, meaning that, as a general principle and in the absence of spercial information bearing on a particular case, there is me presmuptim whateget in favour of las dyym no las bertidey rather than mony uther diay. If as a grumal rule thom pome.e.ti brations on such a day such as would be not unlikely to acelelerate
death, we should say that a men's dying on his birthday was not altorether a matter of chance. If we knew no such general rule but did not know enough about birthdays to be assured that there was no such rule, we could not call the chance 'objective'; we could only speak of it thus, if on the evidence before us there was a strong presumption against the existence of any such general rule.
10. The philosophical and scientific importance of objective chance as defined above cannot be made plain, until Part V., on the Foundations of Statistical Inference, has been reached. There it will appear in more than one connection, but chiefly in connection with the application of Bernoulli's formula. In cases where the use of this formula is valid, important inferences can be drawn; and it will be shown that, when the conditions for objective chance are approximately satisfied, it is probable that the conditions for the application of Bernoulli's formula will be approximately satisfied also.
11. The term random has been used, it is well recognised, in several distinct senses. Venn \({ }^{1}\) and other adherents of the 'frequency' theory have given to it a precise meaning, but one which has avowedly very little relation to popular usage. A random sample, says Peirce, \({ }^{2}\) is one " taken according to a precept or method, which, being applied over and over again indefinitely, would in the long run result in the drawing of any one set of instances as often as any other set of the same number." The same fundamental idea has been expressed with greater precision by Professor Edgeworth in connection with his investigations into the law of error. \({ }^{3}\) It is a fatal objection, in my opinion, to this mode of defining randomness, that in general we can only know whether or not we have a random sample when our knowledre is nearly complete. Its divergence from ordinary usage is well illustrated by the fact that there would be perfect randomness in the distribution of stars in the heavens, as Venn explicitly points out, if ther were disposed in an exact and symmetrical pattern. \({ }^{4}\)

\footnotetext{
\({ }^{1}\) Logic of Chance, chap. v., "The Conception Randomness and its Scientific Treatment."

2 "A Theory of Probable Inference" (published in Johns Hopkins Studies in Logic), p. 152.
\({ }^{3}\) "Law of Error," Camb. Phil. Trans., 1904, p. 128.
\({ }^{4}\) But it may be added that this seems inconsistent with Venn's conception of randomness as that of aggregate order and individual irregularity; nor is it concordant with Venn's typically random diagram (p. 118). His usage, therefore, is sometimes nearer than his definition to the popular usage.
}

I do not believe, therefore, that this kind of definition is a useful one. The term must be defined with reforence to prothability, not to what will happen "in the lomer run "; thoush there may be two senses of it, eorresponding to subjective and objective probability respectively.

The most important phrase in which the term is uand is that of 'a random setection ' or 'taken at random.' When we aptly this term to a particular member of a series or collection of objects, we may mean one of two thinus. We may mean that our Imowledes of the uethot of choosing the partiontar nowner is such that is preirit the member themen is as likele (1) bue an one memler of the series as any other. We may also mean. not that we have no knowledge as to which particular member is in question, but that such knowledge as we have respecting the particular mewher. as distincuished from other membure of the series, is irrelevant to the question as to whether or not this member has the characteristic under examination. In the first case the particular momber is a random member of the series for all characteristics; in the second case it is a random member fur some waly. Is the sencond case is the wore qeneval. we land hemer take that for the purpose of definime random selection.'

The point will be brought out further if we discuss the more dificult use of the term. What exactly dow moan hes the statement: "Any number, taken at random, is cqually. likely to be odd or even" ? Aceordine to the frequmen thanery, this simply means that there are as many whe numbers as theme are even. Takine it in a sonse corresponding to subjeotion chano (and to the explanations given ahove). I propese as a definition the following: \(a\) is taken at random from the chases for the purposes of the propesitional function s \((x)\). \(b(x)\), relative to evidence \(h\), if ' \(x\) is \(a\) ' is irrelevant to the probability \(\phi_{(x)} \mathrm{F}(x)\). h. Thus 'the number of the inhabitants of France is odd" is remative to my knowledere, a ramdom instane. of the propositional function ' \(x\) is an odd number,' since ' \(a\) is the number if the inhabitants of France' is ierele vant to the prot ability of ' \(a\) is odd.' \({ }^{1}\) Thus to say that a number taken at random is as likely to be odd as even, means that there is a

\footnotetext{
 is it atwle for ti... momoter of mithatitants of Frame.
}
probability \(\frac{1}{2}\) that any instance taken at random of the generalisatioin 'all numbers are odd ' (or of the corresponding gencralisation ' all numbers are even') is true ; an instance being taken at random in respect of evenness or oddness, if our knowledge about it satisfics the conditions defined above. Whether or not a given instance is taken at random, depends, therefore, upon what generalisation is in question.
12. We may or may not have reason to believe that, if we take a seri:s of random selections, the proportionate number of occurrences of one particular type of result will very probably lie within certain limits. For reasons to be explained in Chapter XXIX., random selection relative to such information may conveniently be termed 'randorn selection under Bernoullian conditions.' It is this kind of random selection which is scientifically and statistically important. But, as this corresponds to "objective chance,' it is convenient to have a wider definition of 'random selection' unqualified, corresponding to 'subjective chance '; and it is this wider definition which is given above.

The term opposite to 'random selection' in ordinary usage is ' biassed selection.' When I use this phrase without qualification I shall use it as the opposite of 'random selection' in the wider unqualified sense.

\section*{CHAPTER XXV}

\section*{SOME PROBLEMS ARISING OUT OF THE DISCUSSION OF CHANCE}
 made to attribut, certain astronomical phenomena to a specific cause, rather than to objective chance in some such sense as has been defined in the preceding chapter.

The first of these is concerned with the inclinations to the ecliptic of the orbits of the planets of the solar system. This problem has a long history, but it will be sufticient to take \(\mathrm{De}_{\mathrm{e}}\) Morgan's statement of it. \({ }^{1}\) If we suppose that each of the orbits mistht have ctay inclination, We ohtain a vass namber of combimtims of which omly a small number are such that their sum is as small or smaller than the sum of those of the actual system. But the very existmoe of oursolves and our wold can bo showes to imply that one of this small number has been selected, and De Morgan derives from this an enormous presumption that " there was a necessary cause in the formation of the solar system for the inclinations being what they are."

The answer to this was pointed out by D'Alembert 2 in criticis-


 that all the planets mowe in the same sanse as the earth. He conduder : " ()n



 Laphace had in his turn borrawed the example, also without acknowledemont.






ing Daniel Bernoulli. De Morgan could have reached a similar result whatever the configuration might have happened to be. Any arbitrary disposition over the celestial sphere is vastly improbable \(\grave{a}\) priori, that is to say in the absence of known laws tending to favour particular arrangements. It does not follow from this, as De Morgan argues, that any actual disposition possesses à posteriori a peculiar significance.
2. The second of these problems is known as Michell's problem of binary stars. Nichell's Memoir was published in the Philosophical Transactions for 1767. \({ }^{1}\) It deals with the question as to whether stars which are optically double, \(i . e\). which are so situated as to appear close together to an observer on the earth-are also physically so " either by an original act of the Creator, or in consequence of some general law, such perhaps as gravity." He argues that if the stars "were scattered by mere chance as it might happen . . . it is manifest . . . that every star being as likely to be in any one situation as another, the probability that any one particular star should happen to be within a certain distance (as, for example, one degree) of any other given star would be represented . . . by a fraction whose numerator would be to its denominator as a circle of one degree radius to a cirele whose radius is the diameter of a great circle . . . that is, about 1 in 13131." From this begimning he derives an inmesse presumption against the scattering of the several contiguous stars that inay be observed "by mere chance as it might happen." And he groes on to argue that, if there are causal laws directly tending to produce the observed proximities, we may rasonably suppose that the proximities are actual, and not merely optical and aiprar" int. The fact that Michell's induction was confirmed by the later investigations of Herschell adds interest to the speculation. But apart from this the argument is evidently

\footnotetext{
que ins Manètes pouroient no avoir pas une certaine disponition deteminie à volonté. . . ."

D'Alembert is employing the instance for his own purposes, in order to build up an ad hominem argument in favour of his theory concerning 'runs' against I. Bernoulli (see also p. 317).
\({ }^{1}\) See also Todhunter's History, pp. 332-4 ; Venn, Logic of Chance, p. 260 ; Furbes, "On the Alleged Evidence for a Physical Connexion between Stars forming, Binary or Multiple Groups, deduced from the Doctrine of Chances," Phil. Mag., 1850, and Boole, "On the Theory of Probabilities and in particular on Micheli's Problem of the Distribution of the Fixed Stars," Phil. May., 1s.j1.
}
subther than in the first wample. Michell arews that them are more stars optically contiguons, than would her lithly it thom were no speceial cause artine towards this cond and further that, if such is cause is in operation, it must low real, and now meroly optical, contiguity that results from it.

Let us analyse the argument more closely. By " mere chance as it might happon" Michell cannot be supposed to mean "uncaused." He is thinking of objective chance in the sense in which I have defined this in the precediny chapter. We speak of a chance occurrence when it is brought about by the coinedence of forcos and circumstances so mumemes and complex that knowloden sufficont for its prediction is at a kind altomenthes out of our reach. Michell uses the term vaguely but means, I think. somethiner of this kimd: An event is dur tor mare chumere when it can only occur if a large number of independent \({ }^{1}\) conditions are fulfilled simultaneously. The alternatives which
 due to the interaction of a vast variety of stellar laws and positions or are they the result of a few fundamental tendencies, whien misht be the subjoct of knowleden and which womld hat us to expect such stars in relative profusion ?

The existence of numerous binary stars may give a real inductive argument in favour of their arising out of the inter-
 it is not possible to arrive at such precise results as Michell's. If there is some finite probability à priori that binary stars. when they arise, do arise in this way, then, since the frequent. coincidence of a given set of independent causes relatively few in number is more likely than that of a set relatively numerous, the whombution of himarys stars will rais. this probatility if pasleriori to an extent which depends upon the relative profusion in which such stars appear. If, in short, the first of the two alternatives proposed above is assumed, there is no greater presumption for a distribution, covering a part of the heavens, in which binary stars appear, than for any other distribution ; if the second is assumed, there is a greater presumption. The observation of numerous distributions in which binary stars appear increuses, therefore, by the inverse principle, any a priori probability which may exist in fayour of the soomel hymetheris.

\footnotetext{
1 … : : \& Ninte (ii.) t.a Part III.
}

But more than this the argument cannot justify. That Michell's argument is, as it stands, no more valid than De Morgan's, becomes plain when we notice that he would still have a high probability for his conclusion even if only one binary star had been observed. The valuable part of the argument must clearly turn upon the observation of numerous binary stars.

Let us now turn to Michell's second step. He argues that, if binary stars arise out of the interaction of a small number of independent forces, they must be physically and not merely optically double. The force of this argument seems to depend upon our possessing previous knowledge as to the nature of the principal natural laws, and upon an assumption, arising out of this, that there are not likely to be forces tending to arrange stars, in reality at great distances from one another, so as to appear double from this particular planet. But Michell, in arguing thus, was neglecting the possibility that the optical connection between the stars might be due to the observer and his means of observation. It was not impossible that there should be a law, comnected with the transmission of light for example, which would cause stars to appear to an observer to be much nearer together than they really are.

While, therefore, a relative profusion of binary stars constitutes evidence favourably relevant to Michell's conclusion, the argument is more complex and much less conclusive than he seems to have, supposed. This is a criticism which is applicable to many such arguments. The simplicity of the evidence, which arises out of the lack of much relevant information, is liable, unless we are careful, to lead us into deceptive calculations and into assertions of high numerical probabilities, upon which we should never venture in cases where the evidence is full and complicated, but where, in fact, the conclusion is established far more strongly. The enormously high probability in favour of his conclusion, to which Michell's calculations led him, should itself have caused him to suspect the accuracy of the reasoning by which he reached it.
3. some mon" reeent problems of this type seem, however, so far as I am acruainted with them, to follow safer lines of argumont. The: most important are concerned with the existence of star driits. It serms to me not at all impossible to possess daln on which a valid argument can be constructed from the
ohservation of optically apparmen star drifts to the probabilits. of a real uniformity of menion amomest certain s.tio of stare relatively to others.

Amplher problem, sommewnt analemus th the precolines. hats been recently discussed by Professor Karl Pearson. \({ }^{1}\) The title misht fome a lint miswading pertaps motil the oxplanation has been reached of the sense in which the term 'random' is used in it. But Professor Pearson uses the term in a perfectly precise sense. He defines a random distribution as one in which spherical shells of equal volume about the sun as centre contain the same number of stars. \({ }^{2}\) He argues that the observed facts remder probable the thiloming disjumation: Bithor the diemibution of stars is not random in the sense defined above, or there is a correlation between their distance and their brilliancy, such as might be produced, for example, by the absorption of light in its transmission through space, or the space within which they all lie is limited in volume and not spherical in form. \({ }^{3}\) But it is
 as Michell's. For there is no reason to suppose that a nonrandom distribution is more likely than a random distribution to depend upon the interaction of a smabll number of ind pendent forces, and there might even exist a presumption the other way. This arbitrary interpertation of randommess does not help us to the solution of any interesting problem.
4. The discussion of fincl causes and of the argument from
 theology. But the logical prohbom is phain and can be determined upon formal and abstract considerations. The areument is in all cases simply this-an event has oerurred and has been observed which would be very improbable a priori if we did not know that it had actually happened; on the other hand, the event is of such a character that it might have been not unreasonably predicted if we had assumed the existence of a conscions agent whose motives are of a certain kind and whose powers are sufficient.

\footnotetext{


\({ }^{2}\) It in therefore, independent of direction, and the distribution is ramdom
 tion is. therefore, exceedingly arbitrary.
 contre."
}

Symbolically: Let \(h\) be our original data, \(a\) the occurrence of the event, \(b\) the existence of the supposed conscious agent. Then \(a / h\) is assumed very small in comparison with \(a / b h\); and we require \(b / a h\), the probability, that is to say, of \(b\) after \(a\) is known. The inverse principle of probability already demon-
strated shows that \(b / a h=a / b \hbar{ }_{a}^{b / h}, \overrightarrow{a / h}\), and \(b / a h\) is therefore not determinate in terms of \(a / b h\) and \(a / h\) alone. Thus we cannot measure the probability of the conscious agent's existence after the event, unless we can measure its probability before the event. And it is our ignorance of this, as a rule, that we are endeavouring to remedy. The argument tells us that the existence of the hypothetical agent is more likely after the event than before it ; but, as in the case of the general inductive problem dealt with in Part III., unless there is an appreciable probability first, there cannot be an appreciable probability afterwards. No conclusion, therefore, which is worth having, can be based on the argument from design alone; like induction, this type of argument cain only strengthen the probability of conclusions, for which there is something to be said on other grounds. We cannot say, for example, that the human eye is due to design more probably than not, unless we have some reason, apart from the natrie of its construction, for suspecting conscious workmanship. But the necessary à priori probability, derived from some other souree, may sometimes be fortheoming. The man who upon a desert island picks up a watch, or who sees the symbol John Smith traced upon the sand, can use with reason the argument from design. For he has other grounds for supposing that beings, capable of designing such objects, do exist, and that their prescince on the island, now or formerly, is appreciably possible.
5. The most important problems at the present day, in which arguments of this kind are employed, are those which arise in connection with psychical research. \({ }^{1}\) The analysis of the 'cross-

\footnotetext{
1 The probability that a remarkable success in naming playing cards is due to psyehic agency, was discussed by Professor Edeeworth in Mctretike. This was, I think, the first application of probabilities to these questions. See also Proceedings of the Society for Psychical Research, Parts VIII. and X. ; Professor Edgeworth's article on Pisychucal Research and Statistical Method. Stat. Journ. vol. Ixxxii. (1919) p. 222; and Experiments in Psychical Research at Leland Stanford Junior University, by J. Coover.
}
correspondunces.' which have played so largo a part in recent discussions. prosents many peints of diffisulty which are int dissimilar to those which arise in other scientific inquiries of great complexity in which our intial knowlodye is small. In impertant part of the lengial problem, therefore, is to di-timguish the pecolimrity of psedices problems and to discover what special evidene they demand beyond what is required when we dral with other questions. There is a certain tendency, I think, arising out of the belief that psychical problems are in some way peculiar, to raise seeptical deuters arainat them, which are myally valid against all wi mitio: froms. Withentent rine into any questions of detail, let us endeavour to separate those difficulties which sem perention to pretheal researeh from these which, howewr grat, are nen difirrat from the dillimulties which confront students of heredity, for instance, and which are not less likely than these to yield ultimately to the patience and the insight of investigators.

For this purpose it is necessary to recur, briefly, to the analysis of Part III. It was argued there that the methods of empirical proof, by which we strengthen the probability of our conclusions, are not at all dissimilar, when we apply them to the discovery of formal truth, and when we apply them to the discovery of the laws which relate material objects, and that they may possibly prove useful evern in the case of metaphysics; but that the initial probability which we strengthen by these means is differently obtained in each class of problem. In logic it arises out of the prstulate that upment sidervidome inverts what stmes self-evident with some degree of probability; and in physical sciener, out of the postulate that there is a limitation to the amount of independent vari ty amongst the qualities of material objects. But both in logic and in physical science we may wish to consider hypotheses which it is not possible to invest with any à priori probability and which we entertain solely on account of the known truth of many of their consequences. An axiom which has no self-evidence, but which it seems necessary to combine with other axioms which are seli-e ident in order to deduce the generally aceptid body of formal truth, stands in this category. A scientific entity, such as the ether or the electron.
 peratulati. for purperes of ixplanation, stamds in it alow. If the
analysis of Part III. is correct, we can never attribute a finite probability \({ }^{1}\) to the truth of such axioms or to the existence of such scientific entities, however many of their consequences we find to be true. They may be convenient hypotheses, because, if we confine ourselves to certain classes of thicir consequences, we are not likely to be led into error ; but they stand, nevertheless, in a position altogether different from that of such generalisations as we have reason to invest with an initial probability.

Let us now apply these distinctions to the problems of psychical research. In the case of sone of them we can obtain the initial probability, I think, by the same kind of postulates as in physical seime, and our conclusions no d not be op: n to a greater degree of doubt than these. In the case of others we caunot ; and these must remain, unless some method is opeti to us peculiar to psychical ressarch, as tentatio unproved hypotheses in the same category as the ether.

The best example of the first class is afforded by telepathy. We know that the consciousnesses which, if our hypothesis is correct, act upon one ainother, do exist; and I see no logical difference betwenn the problem of establishing a law of telepathy and that of (wtablishing the law of gravitation. There is at present a practical difference on account of the much narrower scope of our knowledge, in the case of telepathy, of cognate matters. We can, therefore, be much luss certain; but there seems no reason why we should necessarily remain less certain after more evidence has been accumulated. It is important to remember that, in the case of telepathy, we are merfly discovering a relation between objects which we already know to exist.

The best example of the other class is afforded by attempts to attribute psychic phonomena to. the agency of 'spitits' other than human beings. Such arguments are weakend at present by the fact that no phenommare known, so far as I am aware, which cannot be explained, though improbably in some cases, in other ways. Bute even if phemomema were to be observed of

1 I am'assuming that there is no argument, arising either from self-evidence or analogy, in addition to the argument arising from the truth of their consequences, in favour of the truth of such axioms or the existence of such objects ; but I daresay that this may not cortainly be the case. The reader may be reminded also that, when I deny a finite probability this is not the same thing as to affirm that the probability is infinitely small. I mean simply that it is not greater than some numerically measurablo probability.
which in known agemer monh afiond even an impentahbe explanation, the hypothesis ot 'spirits 'would still li.. in ther sam. logical limbo as the hypothesis of the 'ether,' in which they might be supposed not inappropriately to move.

Such an hypothesis as the existence of 'spirits' could only become substantial if some peculiar method of knowledge were within our pmowe which would yifid us the initial probrabilits which is demanded. That such a method exists, it is not infrequently claimed. If we can directly perceive these 'spirits,' as many of those who are described in James's Varieties of Religions.s Erperience think ther ran, ther prohberm is, lowicalls: altogether changed. We have, in fact, very much the same kind of reason, though it may be with less probability, that we have for believing in the existence of other people. The preceding prastapl: applins only to attempts at prosine the existence of 'spirits' from such evidence as is discussed by the Society for Psychical Research.

In between these two extremes comes a class of cases, with regard to which it is extremely difficult to come to a decisionthat of attembts that tribut. peselice phenomena to the conscious. agency of the dead. I wish to discuss here, not the nature of the existing evidence. but the question wher ther it is pmasible for any evidence to be convincing. In this case the object whose existence we are endeavouring to demonstrate resembles in muny respects objects whicl: w. know to exist. Ther quastion of epistmolery, which is berne us, is this: Is it mecessary, in order that we may have an initial probahility, that the objere, of our hypothesis should resemble in every relevant partioular some ome objoet which we knmw to exist, or is it sutficiont that we should know instances of all its suppused qualities, thunch u.w.e. in combination? It is char that some qualitios may he irrelawans -position in time and space, for example-and that 'every relevant partionlar " need notituelude these. |But can the initial probubility exist if our hypothesis assumes qualitios, which haw
 have mo knowlater of consciousaess existine apart from a living body, can imfiect evidenee of whatewer charactor atfond us any probability of such a thing? Could any widenco. for examplo. persuade us that a trea. hat the cmotion of amusement, aven if it haughed ropmatediy when we made joken? Set the maloges
which we demand seems to be a matter of degree ; for it does not seem unreasonable to attribute consciousness to dogs, although this constitutes a combination of qualities unlike in many respects to any which we know to exist.

This discussion, however, is wandering from the subject of probability to that of epistemology, and it will not be solved until we possess a more comprehensive account of this latter subject than we have at present. I wish only to distinguish between those cases in which we obtain the initial probability in the same manner as in physical science from those in which we must get it, if at all, in some other way. The distinctions I have made are sufficiently summarised by a recapitulation of the following comparisons: We compared the proo of telepathy to the proof of gravitation, the proof of non-human 'spirits' to the proof of the ether, and, much less closely, the proof of the consciousness of the dead to the proof of the consciousness of trees, or, perhaps, of dogs.

Before passing to the next of the rather miscellaneous topics of this chapter, it may be worth while to add that we should be very chary of applying to problems of psychical research the calculus of probabilities. The alternatives seldom satisfy the conditions for the application of the Principle of Indifference, and the initial probabilities are not capable of being measured numerically. If, therefore, we endeavour to calculate the probability that some phenomenon is due to 'abnormal' causes, our mathematics will be apt to lead us into unjustifiable conclusions.
6. Uninstructed common sense seems to be specially unreliable in dealing with what are termed 'remarkable occurrences.' Unless a 'remarkable occurrence' is simply one which produces on us a particular psychological effect, that of surprise, we can only define it as an event which before its occurrence is very improbable on the available evidence. But it will often occurwhenever, in fact, our data leave open the possibility of a large number of alternatives and show no preference for any of them -that every possibility is exceedingly improbable à priori. It follows, therefore, that what actually occurs does not derive any peculiar significance merely irom the fact of its being 'remarkable' in the above sense. Something further is required before we can build with suceess. Jet Miehell's argument and the argu-
ment from design derive a grood deal of their plawsitility. 1 thimk, from the 'remarkable " character of the actual comstitution whe ther of the heavens or of the utiverse, in foremphbere of the fact that it is impossible to propmund amy comstitution which would if it existed be other than 'remarkable.' It is supposeed that a remarkable occurmence is speriully in meed of an explarm. tion, and that any suffierent explanation has a hich probathilits. in its favour. That an explanation is particularly required, possesses a measure of truth ; for it is likely that our original data were much lacking in completmioss, and the wentrence of the extraordinary event brings to light this deficiency. But that wre are not justifiod in adepting with comfidne. any oufficion explanation, has been shown already.

Such arguments, however, get a part of their plausibility from a quit different source. There is a gen-tal supmaition that fome kimds of oecurremes are there lik.iy than others to boe sume prithe of an explanation by us ; and, therefore, any explanation which deals with such eases falls in prepared soil. Results which,
 duce fall intor this cate mory. li.mats which would be probahle, supposing a direct and predominant cans.! d ferdence botwon the efomments when conemmitnce is remarkel, belong to it atore There is. in fact, a surt of amment from analoney as th whether certain sorts of phenomena are or are not likely to be due to 'chance.' This may explain, for example, why the particular concurrence of atoms that go to compose the human eye, why a serims of correct gums.s in mamme playing conds. why apmolal symmetry or special asymmetry amongst the stars, seem to require explanation in no ordinary degree. Prior to an explanation these particular concurrences or series or distributions are no more improbable than any other. But the causes of such conjunctions as these are more likely to be discoverable by the human mind than are the causis of mimers. atd the attempt to explain them deserves, therefore, to be more carefully considered.
 cases in which we beli.w. the canses th he known th us, hak, perhapse, some weresh. But the direct upplimetion of the Calculus of Probabilities can do no more in these wam tham ancewnt matere for investigation. Ther fact that a man has made a long series of correct guesses in cases where he is cut off from tho modinary
channels of communication, is a fact worthy of investigation, because it is more likely to be susceptible of a simple causal explanation, which may have many applications, thasi a case in which false and true guesses follow one another with no apparent regularity.
7. In the case of empirical laws, such as Bode's law, which have no more than a very slight connection with the general body of scientific knowledge, it is sometimes thought that the law is more probable if it is proposed before the examination of some or all of the available instances than if it is proposed after their examination. Supposing, for example, that Bode's law is accurately true for seven planets, it is held that the law would be more probable if it was suggested alter the examination of six and was confirmed by the sulsequent discovery of the seventh, than it would be if it had not been propounded until after all seven had been observed. The arguments in favour of such a conclusion are well put by Peirce: \({ }^{1}\) "Ali the qualities of objects may be conceived to result from variations of a number of continuous variables; hence any lot of objects possesses some character in common, not possessed by any other." Hence if the common character is not predesignate we can conclude nothing. Cases must not be used to prove a generalisation which has only been suggested by the cases themselves. He takes the first five poets from a biographical dictionary with their ages at death:
\begin{tabular}{lll|lll} 
Aagard & 48 & Abunowas &. & 18 \\
Abeille & - & 76 & Accords & \(\cdot\) & 45
\end{tabular}
.: These five ages have the following characters in common:
" 1 . The difference of the two digits composing the number, divided by three, leaves a remainder of one.
" 2. The first digit raised to the power indicated by the second, and then divided by three, leaves a remainder of one.
\(\cdots 3\). The sum of the prime factors of each age, including one as a prime factor, is divisible by three."
He compares a generalisation regarding the ages of poets based

\footnotetext{
\({ }^{1}\) C. S. Peirce, A Theory of I'robable Inference, pp. 162-167; published in Johns Hopkins Studies in Logic, 1883.
}
on this evidence to Dr. Lyon Playfair's argument about the specific gravities of the three allotropic forms of carbon

approximately, the atomice weight of earbon beine 12. 1) Playlair thinks that the abowe renders it probable that the -pererine gravitios of the allotropic forms of other elements. would. it w. knew them, be found to equal the difierent ronts of their atomic weight.

The weakness of these arguments, howner, has a difinemt explanation. These inductions are were improhahb, herausw they are out of relation to the rest of our knowlodge and are hased on a very small number of instances. The appormit ahsurdity moreover, of the inductive law of Ponts" Ines is inmeand be the fact that we take aceount of the knowleden we ant nally persoss that the ares of poms are not in fact emmected by any such law. If we knew mothing whatever about fuens anes exem what is stated above, the induction would be as valid as ans wher which is based on a very weak analogy and a very small mumber in instances and is unsupported by indirect evidence.

Thepeculiar virtue of prediction or praderamation i-altw... ther imaginary. The number of instances examined atal the ammey between them are the essential points, and the yurstion as ow whe ther a particular hypothesis happens to ber promumad home. or after their examination is quite irrelevant. If all our inductions had to be thought of before we axamimed the cams th which we apply them, we should, doubthess, make fewer inductions; but there is no reastun the think that the fiew we shonh make would be any better than the many from which we should be preduded. The plausibility of the argument is derised thom a differont source. If an hypothesis is propmend it primit, this eommonly moans that there is some promen for it arisime ont of our previons knowlodge, apart from the purely inductive anmmi. and if such is the case the hypothesis is clearly stromer than ome which repmess on inductive grounds only. But if it is a mem. gurss, the lucky fact of its preceding some or all of the cases which verify it ablds nothing whatever to its value. It is the union of
prior knowledge, with the inductive grounds which arise out of the immediate instances, that lends weight to an hypothesis, and not the occasion on which the hypothesis is first proposed. It is sometimes said, to give another example, that the daily fulfilment of the predictions of the Nuutical Alinanack constitutes the most cogent proof of the laws of dynamics. But here the essence of the verification lies in the variety of cases which can be brought accurately under our notice by means of the Almanack, and in the fact that they liave all been obtained on a uniform principle, not in the fact that the verification is preceded by a prediction.

The same point arises not uncommonly in statistical inquiries. If a theory is first proposed and is then confirmed by the examination of statistics, we are inclined to attach more weight to it than to a theory which is constructed in order to suit the statistics. But the fact that the theory which precedes the statistics is more likely than the other to be supported by general considerations --for it has not, presumably, been adopted for no reason at allconstitutes the only valid ground for this preference. If it does not receive more support than the other from general considerations, then the circumstances of its origin are no argument in its favour. The opposite view, which the unreliability of some statisticians has brought into existence,- that it is a positive advantage to approach statistical evidence without preconceptions based on general grounds, because the temptation to ' cook' the evidence will prove otherwise to be irresistible,--has no logical basis and need only be considered when the impartiality of an investigator is in doubt.

\section*{CHAPTER XXVI}

\section*{The Application of probability to conduct}
1. Given as our basis what knowledge we actually have, the prohable. I haw said, is that whith it is manal for us to lneliove. This is not a definition. For it is not rational for us to believe that the probable is true; it is only rational to have a probable belief in it or to believe it in preference to alternative beliefs. To

 must have reference to action and must be a loose way of expressing the propriety of acting on one hypothesis rather than on another. We might put it, therefore, that the probable is the hypothesis on which it is rational for us to act. It is, however, not su simple as this. for the chmims wasm that of two hyputherse it may be rational to act on the less probable if it leads to the greater good. We cannot say more at present than that the probability of a hypothesis is one of the things to be determined and taken account of before acting on it.
2. I do not know of passarees in the ancient philosophers which explicitly point out the dependence of the duty of pursuing goods on the reasonable or probable expectation of attaining them relative to the aments knowledar. This mans only drat analysis had not disentangled the various elements in rational action, not that common sense neqlected them. Herodotus puts the perint quite phainls. "There is mothere mon protitable for a man," he says, "than to take good counsel with himself; for even if the event turns out contrary to one's hope, still one's decision was risht, wem thomgh fortume han made it of menefter : whereas if a man acts contrary to frood counsel, atthough by luck he gets what her had me rieht tar expert, his deceision was men any the less foolish." \({ }^{1}\)

\footnotetext{
\({ }^{1}\) Herod. vil. II.
}
3. The first contact of theories of probability with modern ethics appears in the Jesuit doctrine of probabilism. According to this doctrine one is justified in doing an action for which there is any probability, however small, of its results being the best possible. Thus, if any priest is willing to permit an action, that fact affords some probability in its favour, and one will not be damned for performing it, however many other priests denounce it. \({ }^{1}\) It may be suspected, however, that the object of this doctrine was not so much duty as safety. The priest who permitted you so to act assumed thereby the responsibility. The correct application of probability to conduct naturally escaped the authors of a juridical ethics, which was more interested in the fixing of responsibility for definite acts, and in the various specificd means by which responsibility might be disposed of, than in the greatest possible sum-total of resultant good.

A more correct doctrine was brought to light by the efforts of the philosophers of the Port Royal to expose the fallacies of probabilism. "In order to judge," they say, " of what we ought to do in order to obtain a good and to avoid an evil, it is necessary to consider not only the good and evil in themselves, but also the probability of their happening and not happening, and to regard geometrically the proportion which all these things have, taken together." \({ }^{2}\) Locke perceived the same point, although not so clearly. \({ }^{3}\) By Leibniz this theory is advanced more explicitly; in such judgments, he says, "as in other estimates disparate and heterogencous and, so to speak, of more than one dimension, the greatness of that which is discussed is in reason composed of both estimates (i.e. of goodness and of probability), and is like a rectangle, in which there are two considerations, viz. that of length and that of breadth. . . . Thus we should
\({ }^{1}\) Compare with this doctrine the following curious passage from Jeremy Taylor:-"We being the persons that are to be persuaded, we must see that we be persuaded reasonably. And it is nnreasonable to assent to a lesser evidence when a greater and clearer is propounded: but of that every man for himself is to take cognisance, if he be able to judge; if he be not, he is not bound under the tie of necessity to know anything of it. That that is necessary shall be certainly conveyed to, him: (God, that best can, will certainly take care for that ; for if he does not, it becomes to be not necessary; or if it should still remain necessary, and he be damned for not knowing it, and yet to know it be not in his power, then who can help it! There can be no further care in this business."
\({ }_{2}\) The Port Royal Logic (1662), Eng. Trans. p. 367.
\({ }^{3}\) Essay concerning Human Understanding, book ii. chap. xxi. § 66.
still nomed the art of thinking and that of estimatiner probabilitios, besides the kmewloder of the value of eromels and cvils, in under properly to employ the art of consequences." \({ }^{1}\)

In his preface to the Analogy Butler insists on " the absolute and formal whligation" umber which wom a low pmotatitity. if it is the erentest may lay us: "Tous prohalitity is the wery guide of life."
4. With the development of a utilitarian ethies laredy concerned with the summing up of consequences, the place of proh)ability in ethical theory has become much more explicit. But although the general outlines of the problem are now elear, there are some elements of confusion not yet dispersed. I will deal with some of them.

In his Principia Ethica (p. 152) 1)r. Moore argues that "the first difficulty in the why of establishime a pmotahility that onfo cours of action will erise a hetter thtal mault than anmener, lies in the fact that we have to take account of the effects of both throughout an infinite future. . . . We can certainly only pretend to calculate the atiocts of actions within what may ho callat an 'immediate future. . . . We must, therefore, certainly have some reason to believe that no consequences of our action in a further future will grenerally be such as to reverse the balance of good that is probable in the future which we can foresee. This large postulate: must be made, if we are ever to assert that the results of one action will be even probably better than those of another. Our utter ignorance of the far future gives us no justification for saying that it is even probably right to choose the greater good within the region over which a probable forecast may extend."

This argument seems to me to be invalid and to depend on a wrong philosophical interpretation of probability. Mr. Moore's reasoning endeavours to show that there is not even a probubility by showing that there is not a certainly. We must not, of coursen. have reason to believe that remote consequences will generully be such as to reverse the balance of immediate goond. But we need not be certain that the opposite is the case. If good is additive, if we hase reusun to think that if itwo actions mee pro. duens mone grond than the other in the now future, and is or- have no moans of diserimimatne thetween the ir results in the distont

\footnotetext{

}
future, then by what seems a legitimate application of the Principle of Indifference we may suppose that there is a probability in favour of the former action. Mr. Moore's argument must be derived from the empirical or frequency theory of probability, according to which we must know for certain what will happen generully (whatever that may mean) before we can assert a probability.

The results of our endeavours are very uncertain, but we have a genuine probability, even when the evidence upon which it is founded is slight. The matter is truly stated by Bishop Butler : "From our short views it is greatly uncertain whether this endeavour will, in particular instances, produce an overbalance of happiness upon the whole ; since so many and distant things must come into the account. And that which makes it our duty is that there is some appearance that it will, and no positive appearance to balance this, on the contrary side. . . ." \({ }^{1}\)

The difficulties which exist are not chiefly due, I think, to our ignorance of the remote future. The possibility of our knowing that one thing rather than another is our duty depends upon the assumption that a greater goodness in any part makes, in the absence of evidence to the contrary, a greater goodness in the whole more probable than would the lesser goodness of the part. We assume that the goodness of a part is farourably relevant to the qoodness of the whole. Without this assumption we have no reason, not even a probable one, for preferring one action to any other on the whole. If we suppose that goodness is always organic, whether the whole is composed of simultaneous or successive parts, such an assumption is not easily justified. The case is parallel to the question, whe ther physical law is organic or atomic, discussed in Chapter XXI. § 6.

Nevertheless we can admit that goodness is partly organic and still allow ourselves to draw probable conclusions. For the alternatives, that cither the goodness of the whole universe throughout time is orqanic or the goodness of the universe is the arithmetic sum of the groodnesses of infinitely numerous and infinitely divided parts, are not exhaustive. We may suppose that the goodness of conscious persons is organic for each distinct

\footnotetext{
1 This passage is from the Analogy. The Bishop adds: " . . . and also that such benevolent endeavour is a cultivation of that most excellent of all virtuous principles, the active principle of benevolence."
}
and individual personality. Or we may suppose that when conscions untes are in comments relatimship, then the whole which we must treat as orsanie includes both units. These are only examples. We must suppose, in general, that the units whose fomduess we must rewand as organic and indivisible are not always larer than these the grondness of which we can perceive and judge directly.
5. The difficulties, however, which are most fundamental from the standpuint of the student of probability, are of a different kind. . Aormal ethical themery at the present day, if theme can be said to be any such, makes two assumptions: first, that degrees of erominess are num rically measurable and arithmetically addition, and socond, that degress of probability also are numerically measurable. This theory goes on to maintain that what we cucht to add thereher. when, in order to decide betwen two cenurses of action. we sum up, the results of cach, are the ' mathematical expectations' of the several results. 'Mathematical "xpertation" is a twethical expression originally derised from the scimetitic study of cramblime and games of ohanere and stamds for the prodact of the prosible eain with the probability of attainime it. \({ }^{1}\) In order to obtain, therefore, a measure of what ought to be our proverne in revarl to varimus alt rnative courses of act iom. we must sum for each course of action a series of terms made up of the amounts of grood which may attach to each of its
 ability.

The first assumption. that quantitios of emendness are duly subject to the laws of arithmetic, appears to me to be open to a certain amount of doubt. But it would take me too far from my proper subject fo discuss it hame and 1 shall allows for the purposes of further argument, that in some sense and to some extmet this as-umption can be justitiad. The s. cand assumption. however, that degrees of probability are wholly subject to the


\footnotetext{
\({ }^{1}\) Proority in the conception of mathematical expectation can, I think, bo
 p. 248). In a letter to Placeius, lis7 (Dutens, vi. i. 34 and ('onturat, op. cil. p. 246) Leibniz proposed an application of the same principle to juris prudence, by virtue of which, if two litigants lay claim to a sum of money. and if the claim of the one is twice as probsble as that of the other, the sum should tee divided between them in that propertion. The dectrine, serms sensible, but I am not aware that it has ever boen acted on.
}
been advocated in Part I. of this treatise. Lastly, if both these points be waived, the doctrine that the ' mathematical expectations ' of alternative courses of action are the proper measures of our degrees of preference is open to doubt on two grounds--first, because it ignores what I have termed in Part I. the 'weights' of the arguments, namely, the amount of evidence upon which each probability is founded; and second, because it ignores the element of 'risk' and assumes that an even chance of heaven or hell is precisely as much to be desired as the certain attainment of a state of mediocrity. Putting on one side the first of these grounds of doubt, I will treat each of the others in turn.
6. In Chapter III. of Part I. I have argued that only in a strictly limited class of cases are degrees of probability numerically measurable. It follows from this that the 'mathematical expectations' of goods or advantages are not always numerically measurable ; and hence, that even if a meaning can be given to the sum of a series of non-numerical ' mathematical expectations,' not every pair of such sums are numerically comparable in respect of more and less. Thus even if we know the degree of advantage which might be obtained from each of a series of alternative courses of actions and know also the probability in each case of obtaining the advantage in question, it is not always possible by a mere process of arithnetic to determine which of the alternatives ought to be chosen. If, therefore, the question of right action is under all circumstances a determinate problem, it must be in virtue of an intuitive judgment directed to the situation as a whole, and not in virtue of an arithmetical deduction derived from a series of separate judgments directed to the individual alternatives each treated in isolation.

We must accept the conclusion that, if one good is greater than another, but the probability of attaining the first less than that of attaining the second, the question of which it is our duty to pursue may be indeterminate, unless we suppose it to be within our power to make direct quantitative judgments of probability and goodness jointly. It may be remarked, further, that the difficulty exists, whether the numerical indeterminateness of the probability is intrinsic or whether its numerical value is, as it is according to the Frequeney Theory and most other theories, simply unknown.
7. The second difficulty, to which attention is called above,
is the neglect of the 'weishts' of arguments in the comeeption of 'mathematical expectation.' In Chapter VI. of I'art I. the signiticance of 'weinht' has bern disensem. In the present comnection the question comes to this - if two probalilitios are equal in degree, ought wi, in chonsing cur emurse of action, th prefer that one which is based on a greater body of knowledice!

The question appears to me to be hifthly perphexime and it is difficult to say much that is useful about it. But the degree of completeness of the information upon which a pombility is based does seem to be relevant, as well as the actual magnitude of the probability, in making practical decisions. Bernoulli's maxim, \({ }^{1}\) that in reckonine a probahility we must tak into amomet all the information which we have, even when reinforced by Locke's maxim that we must get all the information we can, \({ }^{2}\) does not seem completely to meet the case. If, for one alternative, the available information is meresarily small, that dones mit som to be a consideration which ought to be left out of account altogether.
8. The last difficulty concerns the question whether, the
 of different courses of action accurately measures what our preferences ought to be-whether, that is to say, the undesirability of a given course of action increases in direct proportion to any increase in the uncertainty of its attaining its object, or whether some allowance ought to be made for 'risk,' its undusirability increasing more than in proportion to its uncertainty.

In fact the meaning of the judgment, that we ought to act in such a way as to produce most probably the greatest sum of goodness, is not perfectly plain. Does this mean that we ought so to act as to make the sum of the mondurssos of sach of the possible consequences of our action multiplied by its prob)ability a maximum? Those who rely on the conception of mathematical expectation' must hold that this is an indisput-



\footnotetext{






}
sur la règle qénérale, qui prescrit de prendre pour valeur d'un événement incertain, la probabilité de cet événement multipliée par la valeur de l'événement en lui-même," \({ }^{1}\) where he argues from Bernoulli's theorem that such a rule will lead to satisfactory results if a very large number of trials be made. As, however, it will be shown in Chapter XXIX. of Part V. that Bernoulli's theorem is not applicable in by any means every case, this argument is inadequate as a general justification.

In the history of the subject, nevertheless, the theory of 'mathematical expectation' has been very seldom disputed. As D'Alembert has been almost alone in casting scrious doubts upon it (though he only brought himself into disrepute by doing so), it will be worth while to quote the main passage in which he declares his scepticism: " Il me sembloit " (in reading Bernoulli's Ars Conjectundi) " que cette matière avoit besoin d'être traitée d'une manière plus claire ; je voyois bien que l'espérance étoit plus grande, \(1^{\prime \prime}\) que la somme espérée étoit plus grande, \(2^{\circ}\) que la probabilité de gagner l'étoit aussi. Mais je ne voyois pas avec la même évidence, et je ne le vois pas encore, \(1^{\circ}\) que la probabilité soit estimée exactement par les méthodes usitées ; 2" que quand elle le seroit, l'espérance doive être proportionnelle à cette probabilité simple, plutôt qu`à une puissance ou même à une fonction de cette probabilité ; \(3^{\prime \prime}\) que quand il y a plusieurs combinaisons qui domnent différens avantages ou différens risques (qu'on recrarde comme des avantages négatifs) il faille se contenter d'ajouter simplement ensemble toutes les espérances pour avoir l'espérance totale." \({ }^{2}\)

In extreme cases it seems difficult to deny some force to D'Alembert's objection ; and it was with reference to extreme cases that he himself raised it. Is it certain that a larger good, which is extremely improbable, is precisely equivalent ethically to a smaller good which is proportionately more probable? We may doubt whether the moral value of speculative and cautious action respectively can be weighed against one another in a simple arithnetical way, just as we have already doubted whether a good whose probability can only be determined on a slight basis of evidence can be compared by means merely of the

\footnotetext{
\({ }^{1}\) Mist. de l'Acad., Paris, 1781.
\({ }^{2}\) Opuscules mathématiques, vol. iv., 1768 (extraits do lettres), pp. 284, 285. See also p. 88 of the same volume.
}
magnitude of this probahility with another groed whese like lihood is based on completer knowledge.

There seems, at any rate, a rood deal to bee said for the comclusion that, other thimess heiner equal, that murse of action is preferable which involves leat risk, and about the results of which we have the most complef kmewtedere. In marginal cases, therefore, the conefficients of weight and risk as well as that of probatility are reflevant to our comelusion. It semems natural to suppesi that they should exem some influence in other casess also, the only diflienter in this bering the lack of any principle for the calculation of thee dearee of their influence. A high weight and the ahsone of risk inerase pro tanto the desirability of the action to which they reftre, hut wer canmot measure the amount of the increase.

The 'risk' may be defined in some such way as follows. If A is the ammunt of mond which may result, \(p\) its probability \((p \circ q-1)\), and E the value of the "mathematical expectation, so that E: ph, then the 'risk' is I , where \(\mathrm{R}-\mu(\mathrm{A}-\mathrm{E})\) \(p(1-p) \mathrm{A}=p q \mathrm{~A}=q \mathrm{E}\). This may be put in another way: E measures the net immediate saterition which should bee made in the hope of whainine A: \(q\) is the probatility that this sacritice will be made in vain ; so that qlis is the 'risk.' 1 The ordinary theory suppesies that the whical value of an expectation is a function of E only and is entirely independent of R .

We could, if we liked, define a conventional coefficient \(c\) of
 \((1, q)(1 \cdot w)\)
'weight,' which is equal to unity when \(p=1\) and \(w=1\), and to zero when \(p=0\) or \(w=0\), and has an intermediate value in other cases. \({ }^{2}\) But if doubts as to the sufficiency of the coneention of 'mathematioal expectation' In. su-tamet, it is mut likely that the solution will lie, as 1). Xhembert sumests, ame as has been exemplified above, in the discovery of some mos

\footnotetext{


 and by as suftionent number of such remanances the risk can be complotely


 \(p A=p^{\prime} A^{\prime}, w \quad w^{\prime}\), and \(\eta-q^{\prime}\), we canmut in general compare c \(A\) and \(r^{\prime} A^{\prime}\).
}
complicated function of the probability wherewith to compound the proposed good. The judgment of goodness and the judgment of probability both involve somewhere an element of direct apprehension, and both are quantitative. We have raised a doubt as to whether the magnitude of the 'oughtness' of an action can be in all cases directly determined by simply multiplying together the magnitudes obtained in the two direct judg. ments ; and a new direct judgment may be required, respecting the magnitude of the 'oughtness' of an action under given circumstances, which need not bear any simples and neeessary relation to the two former.

The hope, which sustained many investigator's in the course of the nineteenth century, of gradually bringing the moral sciences under the sway of mathematical reasoning, steadily recedes-if we mean, as they meant, by mathematics the introduction of precise numerical methods. The old assumptions, that all quantity is numerical and that all quantitative characteristics are additive, can be no longer sustained. Mathematical reasoning now appears as an aid in its symbolic rather than in its numerical character. I, at any rate, have not the same lively hope as Condorcet, or even as Edgeworth, "éclairer les Sciences morales et politiques par le flambera de l'Algèbre." In the present case, even if we are able to range goods in order of magnitude, and also their probabilities in order of marnitude, yet it does not follow that we can range the products composed of each good and its corresponding probability in this order.
9. Discussions of the doctrine of Mathematical Expectation, apart from its directly cthical bearing, have chicfly centred round the classic Petersburg Paradox, \({ }^{1}\) which has been treated by almost all the more notable writers, and has been explained by them in a great variety of ways. The Petersburg Paradox arises out of a game in which Peter engages to pay Paul one shilling if a head appears at the first toss of a coin, two shillings if it does not appear until the second, and, in general, \(2^{-1}\) shillings if no head appears until the \(r^{\text {th }}\) toss. What is the value of Paul's expectation, and what sum must he hand over to Peter before the game commences, if the conditions are to be fair?

\footnotetext{
\({ }^{1}\) For the history of this paradox see Todhunter. The name is due, he says, to its having first appeared in a memoir by Daniel Bernoulli in the Commentarii of the Petersburg Academy.
}

The mathematical answer is \(\mathcal{( 1}\left(\frac{1}{2}\right)^{\prime} 2^{r-1}\), if the number of tosses
 is removed. That is to say, Paul should pay " shillings in the first case, and an infinit" stmon in the sumpl. Nollines. it is said, could be mon paradoxical, and no sinn l'anl would entrate on these terms even with an honest Peter.

Hang of the solutions which hate beren otfored will oceour at once to the reader. The comditions wif the game impl! enntrat diotions saty Poisson and (omblorent: Poter has umbertaken
 is deferred even to the 100 th toss, he will owe a mass of silver greatur in bulk thatn the sun. But this is mo atnswor. Potrer has promised mum and a hadin his sulveney wil! atrain our imamination ; but it is imaginable. And in any case, as Bertrand points
 sand or molecules of hydrogen.

D'Alembert's principal explanations are, first, that true expectation is not necessarily the product of probability and
 very long runs are not unly very improbable, but do not occur at all.
 Bemonlli. atmi furns wat the tat that lum one hut a miser remats
 to their amount; as Buffon says, "L'avare est comme le
 numérique." Daniel Bernoulli deduced a formula from the assumption that the importance of an increment is inversely proportional to the size of the fortune to which it is added. 'Thus, if \(x\) is the 'physical' fortune and \(y\) the 'moral ' fortune,
\[
\text { ly } \int_{x}^{d x}
\]
or \(y=k \log _{a} \frac{i}{a}\), where \(k\) and \(a\) are constants.
On the hasis of this formutit of tiormoutlifis a consmberable
theory has been built up both by Bernoulli \({ }^{1}\) himself and by Laplace. \({ }^{2}\) It leads easily to the further formula-
\[
x=\left(a+x_{1}\right) p_{1}\left(a+x_{2}\right) p_{2} \ldots,
\]
where \(a\) is the initial 'physical' fortune, \(p_{1}\), etc., the probabilities of obtaining increments \(x_{1}\), etc., to \(a\), and \(x\) the 'physical' fortune whose present possession would yield the same 'moral' fortune as does the expectation of the various increments \(x_{1}\), etc. By means of this formula Bernoulli shows that a man whose fortune is \(£ 1000\) may reasonably pay a \(£ 6\) stake in order to play the Petersburg game with £1 units. Bernoulli also mentions two solutions proposed by (ramer. In the first all sums greater than \(2^{24}(16,777,116)\) are regarded as 'morally' equal; this leads to \(£ 13\) as the fair stake. According to the other formula the pleasure derivable from a sum of money varies as the square root of the sum ; this leads to \(£ 2: 9\) s. as the fair stake. But little object is served by following out these arbitrary hypotheses.

As a solution of the Petersburg problem this line of thought is only partially successful: if increases of 'physical' fortune beyond a certain finite limit can be regarded as 'morally' negligible, Peter's claim for an infinite initial stake from Paul is, it is true, no longer equitable, but with any reasonable law of diminution for successive increments Paul's stake will still remain paradoxically large. Daniel Bernoulli's suggestion is, however, of considerable historical interest as being the first explicit attempt to take account of the important conception known to modern economists as the diminishing marginal utility of money, -a conception on which many important arguments are founded relating to taxation and the ideal distribution of wealth.

Each of the above solutions probably contains a part of the psychological explanation. We are unwilling to be Paul, partly because we do not believe Peter will pay us if we have good fortune in the tossing, partly because we do not know what we should do with so much money or sand or hydrogen if we won it, partly because we do not believe we ever should win it, and partly because we do not think it would be a rational act to risk

\footnotetext{
1 "Specimen Theorian Novat de Mensura Montis," ('omm. Acud. P'rop. vol. v. for 1730 and 1731, pp. 175-192 (published 1738). See Todhunter, pp. 213 et seq.

² Théorie analytíque, chap. x. "De l'espérance morale," pp. 432-445.
}
an infinite sum or "ren it very laren finite sum for an infinitely larger one, whose attainment is infinitely unlikely.

When we have made the proper hypotheses and have eliminated these elements of psycholorical doubt, the themetie dispersal of what eloment of paradox remains must be brousht about. I think, by a development of the theory of risk. It is primarily the great risk of the wager which devers us. Exem in the came where the number of tosses is in me case forexemed a finite number. the risk \(R\), as already defined, may be very ervat, and the relative risk \(\frac{\mathrm{R}}{\mathrm{E}}\) will be almost unity. Where there is no limit to the number of tosses, the risk is infinite. A relative risk, which approaches unity, may, it has been alroady surensted, be a factor which must be taken into account in ethical calculation.
10. In establishing the doctrine, that all private gambling must be with certainty a hasimerne pmexisly contraty arguments are employed to those which do service in the Petersbury problem. The argument that " you must lose if only you go on long enough " is well known. It is suceinctly put by Laurent : \({ }^{1}\) Two phayers \(A\) and \(B\) hawn "and b trames respectively. \(f(a)\) is the chance that A will be ruined. Thus \(f(a)=\frac{b}{a+b)^{\prime}}\) so that the poorer a gambler is, relatively to his opponent, the more likely he is to be ruined. But further, if \(b-x, f(a)=1\), i.e. ruin is certain. The intintely rich gambler is the publice. It is agrainst the public that the protusciomal gambler phays and his ruin is therefore certain.

Might not Poisson and Condorcet reply, The conditions of the came imply contratiotion, for no sambler phatss as this arenmont supposes, for ever!3 . It ther and of ams fimite quantity of play, the player, even if her is met ther puthie. mu!! timish with wimines of any finitesize. The gambler is in a worse pmation it his capital is smaller thim his "ppoments at prker. for instance. or on the Stock Exchange. This is clear. But our desire for moral improvement outstrips our hage it we teli him that her must lowe. Bensides it is paraduxical the say that eworybondy

\footnotetext{
1 Cishal do. fo.. ciontit - P. 12!
\({ }^{2}\) This would possibly follow from the theorom of lhaniel Bernoulli. The


}
individually must lose and that everybody collectively must win. For every individual gambler who loses there is an individual gambler or syndicate of gamblers who win. The true moral is this, that poor men should not gamble and that millionaires should do nothing else. But millionaires gain nothing by gambling with one another, and until the poor man departs from the path of prudence the millionaire does not find his opportunity. If it be replied that in fact most millionaires are men originally poor who departed from the path of prudence, it must be admitted that the poor man is not doomed with certainty. Thus the philosopher must draw what comfort he can from the conclusion with which his theory furnishes him, that millionaires are often fortunate fools who have thriven on unfortunate ones. \({ }^{1}\)
11. In conclusion we may discuss a little further the conception of ' moral ' risk, raised in \(\S 8\) and at the end of § 9. Bernoulli's formula crystallises the undoubted truth that the value of a sum of money to a man varies according to the amount he already possesses. But does the value of an amount of goodness also vary in this way? May it not be true that the addition of a given good to a man who already enjoys much good is less good than its bestowal on a man who has little? If this is the case, it follows that a smaller but relatively certain good is better than a greater but proportionately more uncertain good.

In order to assert this, we have only to accept a particular theory of organic goodness, applications of which are common enough in the mouths of political philosophers. It is at the root of all principles of equality, which do not arise out of an assumed diminishing marginal utility of money. It is behind the numerous arguments that an equal distribution of benefits is better than a very unequal distribution. If this is the case, it follows that, the sum of the goods of all parts of a community taken together being fixed, the organic good of the whole is greater the more equally the benefits are divided amongst the individuals. If the doctrine is to be accepted, moral risks, like financial risks, must not be undertaken unless they promise a profit actuarially.

\footnotetext{
\({ }^{1}\) From the social point of view, however, this moral against gambling may be drawn-that those whostart with the largest initial fortunes are most likely to win, and that a sriven increment to the wealth of these benefits them, on the assumption of a diminishing marcinal utility of money, less than it injures those from whom it is taken.
}

There is a ervat dwal which cond the said comomemine such a doctrine. Sut it would load tow far from what is r levant to the study of Probahility: Onf or two instamens of its us. hownerer, moy b. taken from the literature of I'robatility. In his essay, *Sur l'application du catent olve pmonhilités à l"imenlation de la petite vérole," \({ }^{1}\) D'Alembert points out that the community would gain on the average if, by sacrificing the lives of one in five of its citizens, it could ensure the health of the rest, but he argues

 an argument which depends essentially on the same point. Suppose that the members of a certain class cause an average detriment MI to society, and that the mischiefs done by the sereral individuats dition mom or las. from 11 be amomets when average is \(D\), so that \(D\) is the average amount of the individual deviations, all regarded as positive, from MI then, Galton argued, the smaller D) is. the stroner is the justitication tor :ahmer surts drastice measures acainst the proparation of the class as would be consomant to the for limes. if it wero hown that eath individual member caused a detriment M. The use of such arguments seems to involve a qualification of the simple ethical doctrine that right action should make the sum of the benefits of the several individual consequences, each multiplied by its probability, a maximum.

On the other hand, the opposite view is taken in the Port Royal Lorgie and be: Buther, when hery argue that verethines cusht to be sacrificed for the hope of heaven, even if its attainment be thought infinitely improbable, since "the smallest degree of facility for the: att.inment of salvation is of hiogher value than all the blessings of the world put twen ther."2 Ther argument is, that we coutht on follew a course of combluct which mese with the slight...t probability lead to an immito eroud, matil it is lowically. disproved that such a result of our action is impossible. The


\footnotetext{



 things of the world. This is why the smallest degree of facility for the attainment of malvation is of higher value than all the hlosinge of the world put toceether.
}
he believed it, but hecause it offered insurance against a disaster whose future occurrence, however improbable, he could not certainly disprove, may not have considered, however, whether the product of an infinitesimal probability and an infinite good might not lead to a finite or infinitesimal result. In any case the argument does not enable us to choose between different courses of conduct, unless we have reason to suppose that one path is more likely than another to lead to infinite good.
12. In estimating the risk, 'moral' or 'physical,' it must be remembered that we cannot necessarily apply to individual cases results drawn from the observation of a long series resembling them in some particular. I am thinking of such arguments as Buffon's when he names \(\frac{1}{10,000}\) as the limit, beyond which probability is negligible, on the ground that, heing the chance that a man of fifty-six taken at random will die within a day, it is practically disregarded by a man of fifty-six tho knows his health to be goorl. "If a public lottery," Gibbon truly pointed out, " were drawn for the choice of an immediate victim, and if our name were inscribed on one of the ten thousand tickets, should we be perfectly easy?"

Bernoulli's second axiom, \({ }^{1}\) that in reckoning a probability we must tak crerything into account, is easily forgotten in these cases of statistical probabilities. The statistical result is so attractive in its definiteness that it leads us to forget the more vague though more important considerations which may be, in a given particular case, within our knowledge. To a stranger the probability that: I shall send a letter to the post unstamped may he derived !rom the statisties of the Post Office: for me those figures would have but the slightest baring upon the question.
13. It has been pointed out already that no knowledre of probabilities, less in degree than certainty, helps us to know what conclusions are true. and that there is no direct relation between the truth of a proposition and its probability. Probability bewins and ends with probability. That a scientific investigation pursued on account of its probability will generally lead to truth, rather than falsehood, is at the best only probable. The proposition that a course of action suided by the most probable considerations will en merally lead to success, is not certainly true and has nothing to recommend it but its probability.

The importance of probabilit can only be derised from the judernemt that it is momenel to berudend by it in action: and a practical dependence on it can only he justified bo a jandgnent that in action we ought to act to take some account of it. It is for this reason that probability is to us the "guide of life," since to us, as Locke says, " in the greatest part of our concernment, God has afforded only the Twilight, as I may so sal, of I'rohability, suitable, I presume, to that state of Mediocrity and Probationership Ho has been pleased to place us in here."
\[
\text { PAliT } 1
\]

\section*{ INFERENCE}

\section*{CIIAPTER XXVII}

\section*{THE NATURE OF STATISTICAL INFERENCE:}
1. The Theory of Statistics, as it is now understood, \({ }^{1}\) cam be divided into two parts which are for many purposes better kept distinct. The first function of the theory is purely descriptive. It d vises mumerical and diaurammat in monds ho which mertain saliont characteristics of harge aroups of phomementa (an h.. hriofly described ; and it provides formular by the aid of which we can measure or summarise the variations in some particular character which we have observed over a long series of events or instances. The second function of the theory is inductive. It senks to extemed its description of certain characteristics of ohserved events to the corresponding characteristics of other events which have not been observed. This part of the subject may be called the Theory of Statistical Inference; and it is this which is closely bound up with the theory of probatility:
2. The union of these two distinct theories in as single science is natural. If, as is generally the case, the development of some inductive conclusion which shall go beyond the actually observed instances is our ultimate object, we naturally chonse those modes of description, while we are enguged in our preliminary investigation, which are most capable of extension beyond the particular instances which they primarily describe. But this union is also the occasion of a creat deal of confusion. 'The statistician, who is mainly interested in the technical methods of his science, is less concerned to diseover the precise conditions in which a description can be legitimately extended by induction. He slips som what easily from one to the other, and laving found is complete and satisfactory mode of description he

\footnotetext{


}
may take less pains over the transitional argument, which is to permit lim to use this description for the purposes of generalisation.

One or two examples will show how easy it is to slip from description into generalisation. Suppose that we have a series of similar objects one of the characteristics of which is under observation ;-a number of persons, for example, whose age at death has been recorded. We note the proportion who die at each age, and plot a diagram which displays these facts graphically. We then determine by some method of curve fitting a mathematical frequency curve which pass sith close approximation through the points of our diagram. If we are given the equation to this curve, the number of persons who are comprised in the statistical scrics, and the degree of approximation (whether to the nearest year or month) with which the actual age has been recorded, we have a very complete and succinct account of one particular characteristic of what may constitute a very large mass of individual records. In providing this comprehensive description the statistician has fulfilled his first function. But in determining the accuracy with which this frequency curve can be employed to determine the probability of death at a given age in the population at large, he must pay attention to a new class of considerations and must display a different kind of capacity. He must take account of whaterer extraneous knowledge may be available regarding the sample of the population which came under observation, and of the mode and conditions of the observations themselves. Much of this may be of a vague kind, and most of it will be necessarily incapable of exact, numerical, or statistical treatment. He is faced, in fact, with the normal problems of inductive science, one of the data, which must be taken into account, being given in a convenient and manageable form by the methods of descriptive statistics.

Or suppose, again, that we are given. over a series of years, the marriage rate and the output of the harvest in a certain area of population. We wish to determine whether there is any apparent degree of correspond nene between the variations of the two within this field of observation. It is technically difficult to measure such degree of comespondence as may appear to exist between the variations in two series, the terms of which are in some mamer associated in couples, - by coincidence, in this ease,
of time and place. By the mothod of correlation tables and correlation coeflicients the desmptipe statistician is able to effect this ohject, and to prosent the inductive seientist with a hiohly significant part of his data in a compact and instructice form. But the statistician has not, in calculatine these conefficimits o: ohserved cortelation, cevered the whole eromed of which the in ductive selmist must take cognisanee. He has reenord d the results of the observations in circumstances where they camot be recorded so chearly without the aid of technical methods: hut the precise natur. of the conditinns in which the ohservations took place and the numerous other considerations of onn so:t in another, of which we must take account when we wish to generalise, are mot usually suse ptible of num rical or statistical expression.

The truth of this is obvious ; yet, not unnaturally, the more complicated and technical the preliminary at ation ical insestigations become, the more prone inquirers are to mistake the statistical description for an induction generalisation. \({ }^{1}\) This tendency, which has exi-w ed in som du-ure. as, I think, the whol histury of the subject shows, from the eighteenth century down to tha presint then, has been further menurared be the 1 mamblay in ordinary use. For several statistical coefficients are given the same name when they are used for purcly descriptive purposes,
 or the precision of an induction. The term 'probable error,' for example, is used both for the purpose of supplementing and improving a statistical description, and for the purpose of indicating the precision of some generalisation. The term 'correlation' itself is used both to describe an observed characteristic of particular phenomena and in the enunciation of an inductive law which relates to phemomena in general.
3. I have been at pains to enforee this contrast between statistical deseription and statistical induction, becamse the
 nearly all statistical treatises are mainly eoneerned with the former. Ny object will be to analyse, so far as I can. the logical


 of nature is absolutely certain."
basis of statistical modes of argument. This involves a double task. To mark down those which are invalid amongst arguments having the support of authority is relatively easy. The other branch of our investigation, namely, to analyse the ground of validity in the case of those arguments the force of which all of us do in fact admit, presents the same kind of fundamental difficulties as we met with in the case of Induction.
4. The arguments with which we have to deal fall into three main classes :
(i.) Given the probability relative to certain evidence of each of a series of events, what are the probabilities, relative to the same evidence, of various proportionate frequencies of occurrence for the events over the whole series? Or more briefly, how often may we expect an event to happen over a series of occasions, given its probability on each occasion ?
(ii.) Given the frequency with which an event has occurred on a series of occasions, with what probability may we expect it on a further occasion?
(iii.) Given the frequency with which an event has oceured on a series of occasions, with what frequency may we probably expect it on a further series of occasions ?

In the first type of argument we seek to infer an unknown statistical frequency from an à priori probability. In the second type we are engaged on the inverse operation, and seck to base the calculation of a probability on an observed statistical frequency. In the third type we seek to pass from an observed statistical frequency, not merely to the probability of an individual occurrence, but to the probable values of nther unknown statistical frequencies.

Each of these types of argument can be further complicated by being applied not simply to the occurrence of a simple event but to the concurrence under given conditions of two or more events. When this two or more dimensional classification replaces the one dimensional, the theory becomes what is sometimes termed Correlation, as distiuguished from simple istatistical Frequency.
5. In Chapter XXVIII. I touch briefly on the observed phenomena which have given rise to the so-called Law of Great Numbers, and the discovery of which first set statistical
investigation suines. In thapter XXIX. the first typ of arenment, as elassitied abowe is analyand, and the comditions wheh are required for its validity are stated. The crucial problem of attacking the second and third types of argument is the subje t of my concluding chapters.

\section*{CHAPTER XXVIII}

\section*{THE LAW OF GREAT NUMBERS}

Natura quidem suas habet consuetudines, natas ex reditu causarum, sed non
 ergo de mortibus quotcunque experimenta feceris, non ideo naturae rerum limites posuisti, ut pro futuro variare non possit.-Leibniz in a letter to Bernoulli, December 3, 1703.
1. Ir has always been known that, while some sets of events invariably happen together, other sets generally happen together. That experienen shows one thing, while not always a sign of another, to be a usual or probable sign of it, must have been one of the earliest and most primitive forms of knowledge. If a dog is generally siven scraps at table, that is sufficient for him to judge it reasonable to be there. But this kind of knowledge was slow to be made precise. Nuinerous experiments must be carefully recorded before we can know at all accurately how usual the association is. It would take a dog a long time to find out that he was given scraps except on fast days, and that there was the same number of these in every year.

The necessary kind of knowledge began to be accumulated during the seventeenth and eighteenth centuries by the carly statisticians. Halley and others began to construct mortality tables; the proportion of the births of each sex were tabulated; and so forth. These investigations brought to light a new fact which had not been suspected previously--namely, that in certain cases of partial association the degree of association, i.e. the proportion of instances in which it existed, shows a very surprising regularity, and that this regularity becomes more marked the greater the number of the instances under consideration. It was found, for example, not merely that boys and girls are born on the whole in about equal proportions, but that the proportion,
which is not one of completo eqquality, tends exery where, when the number of werorded instances hecomes laren, th approximat. towards a certain definite figure.

During the dishemth century matters wern not pmathed much further that this, hat in certain cases, wif when momperatively
 in degree as the instances became more numerous. Bernoulli, however, took the first step towards giving it a theoretical hasis by showing that, if the it piomi proh ahinity is know thomehom. then (sulij ot themtain mombitions which he himes.lf did not make
 is to be expected. Süssmilch (Die götliche Ordnung in den
 theological interest in these remularities. Such ideas had become sufficiently familiar for Gibbon to characterise the results of probability as "so true in general, so fallacious in particular." Kant found in them (as many later writers have done) some bearing on the problem of Free Will. \({ }^{1}\)

But with the nineteenth century came bolder theoretical methods and a wider knowledge of facts. After proving his extension of Bernoulli's Theorem, \({ }^{2}\) Poisson applied it to the observed facts, and gave to the principle underlying these

 selle qu'on peut appeler la loi des grands nombres. . . . De ces exemples de toutes natures, il résulte que la loi universelle des grands nombres est defjà pur mons un fait cénéralet incomustahtw, résultant d"xpérionces qui ne s. dómentont jamais." This is the lamguate of exagereration: it is alsurestromels vactu. But
 a discussion of this passage and for the connection between K :unt und Subsmilch,

\({ }^{2}\) See p. 345.
 pp. 655-8560) has maintained that Poisson intended to state his principle in is less general way than that in which it has been generally taken, and that he wats misunderstood by Quetelet and others. If we attend only to Poisson's con
 there, it is possible to make out a good case for thinking that ho intended his law to extend only to cases where certain strict conditions were fulfilled. But this is mot the spirit of his more popular writings or of the passage quoted above. At any rato, it is tho fashion, in which Poisson influonced his contempuraries, that is historically interesting; and this is certainly not representeal by Von

it is exciting ; it seems to open up a whole new field to scientific investigation ; and it has had a great influence on subsequent thought. Poisson seems to claim that, in the whole field of chance and variable occurrence, there really exists, amidst the apparent disorder, a discoverable system. Constant causes are always at work and assert themselves in the long run, so that each class of event does eventually occur in a definite proportion of cases. It is not clear how far P'oisson's result is due to à priori reasoning, and how far it is a natural law based on experience ; but it is represented as displaying a certain harmony between natural law and the \(a\) priori reasoning of probabilities.

Poisson's conception was mainly popularised through the writings of Quetelet. In 1823 Quetelet visited Paris on an astronomical errand, where he was introduced to Laplace and came into touch with " la grande école française." "Ma jeunesse et mon zile," he wrote in later years, " ne tardèrent pas à me mettre en rapport avece les hommes les plus distingués de cette époque ; qu'on me permette de citer Fourier, Poisson, Lacroix, spécialement comus, comme Laplace, par leurs excellents écrits sur la théorie mathématique des probabilités. . . . U'est donc au milieu des savants, statisticiens, et économistes de ce temps que j'ai commencé mes travaux." 1 Shortly afterwards began his long series of papers. extending down to 1873, on the application of Probability to social statistics. He wrote a text-book on Probability in the form of letters for the instruction of the Prince Consort.

Before accepting in 1815) at the age of nineteen (with a view to a livelihood) a professorship of mathe matics, (Quetelet had studicd as an art student and written poetry ; a year later an opera, of which he was part-author, was produced at Ghent. The character of his scientific work is in keeping with these beginnings. There is scarcely any permanent, accurate contribution to knowledge which can be associated with his name. But suggestions, projects, far-reaching ideas he could both conceive and express, and he has a very fair claini, I think, to be regarded as the parent of modern statistical method.

Quetelet very much increased the number of instances of the

\footnotetext{
1 For the details of the life of Quetelet and for a very full discussion of his writings with special reference to Probability, see Lottin's Quetelet, statisticien et sociologue.
}

Law of tirat Numbers and alda hmutht into prominenew as slightly :arion. type in it. of which a characteristic example is the law of height, accordiner to which the heights of any considerable sample taken from any population tend to group, themselves according to a certain well-known curve. His instances were chiefly drawn from social statistics, and many of them were of a
 the number of suicides, "l'offrayante exactitude avec laquelle les crimes se reproduisent," and so forth. (Quetelet writes with an almost religious awe of these musterious laws, and certainly makes the mistake of treating them as being as adequate and complete in themselves as the laws of physics, and as little needing any further analysis or explanation. \({ }^{1}\) Quetelet's sensational language may have wiven a considerable impertus to the mallation on smal statisties. lout it also involved statistics in a slight element of suspicion in the minds of some who, like Comte, requarded the application of the mathematical calculus of probability to social science as "purement chimérique et, par conséquent, tout à fait viciense." The suspicion of quackery has not yet disappeared. Quetelet belonus, it must be admitted, to the lone line of brilliant writers, not yet extinct, who have prevented Probability from becoming, in the scientific salon, perfectly respectable. Thore is still about it for scientists a smack of astrology, of alchemy.

The progress of the conception since the time of Quetelet has
 respectability have been taken. Instances have been multiplied and the conditions necessary for the existence of statistical stability have been to some extent analysed. While the most fruitiul :upplientions of these: methods have stibl hown perhaps, as at first, in social statistics and in errors of observation, a number of uses for them have been discovered in quite recent times in the other sciences; and the principles of Mendelism
 biology.

\footnotetext{







}
2. The existence of numerous instances of the Law of Great Numbers, or of something of the kind, is absolutely essential for the importance of Statistical Induction. Apart from this the more precise parts of statistics, the collection of facts for the prediction of future frequencies and associations, would be nearly useless. But the 'Law of Great Numbers' is not at all a good name for the principle which underlies Statistical Induction. Thle' 'Stability of Statistical Frequencies' would be a much better name for it. The former suggests, as perhaps l'oisson intended to suggest, but what is certainly false, that every class of event shows statistical regularity of occurrence il only one takes a sufficient number of instances of it. It also encourages the method of procedure, by which it is thought legitimate to take any observed degree of frequency or association, which is shown in a fairly numerous set of statistics, and to assume with insufficient investigation that, because the statistics are numerous, the observed degree of frequency is therefore stable. Observation shows that some statistical frequencies are, within narrower or wider limits, stable. But stable irequencies are not very common, and cannot be assumed lightly.

The gradual discovery, that there are certain classes of phenomena, in which, though it is impossible to predict what will happen in each individual case, there is nevertheless a regularity of occurrence if the phenomena be considered torether in successive sets, gives the clue to the abstract inquiry upon which we are about to embark.

\title{
(HAPTER XNIX
}

\author{
THE USE OF A PRIORI PROBABILITIES FOR THE PREDICYION OF STATASICAL FREQUENCY-THE THEOREME OF BERXOU IA, POISSON, AND TCHEBYCHEFF
}

\begin{abstract}




\end{abstract}
1. Bernoulti's Theorem is generally regarded as the central themerem of atatistical probability. It contodes the thet antumpt
 of individual probabilitios, and it in a sufficient eruit of the twemy years which Bornoulli alleges that he epont in machiner ini- moult, if out of it the conception first arose of general laws amongst masses of phenomena, in spite of the uncertainty of each prarticular casc': But, as we shall sew, the themem is mily valid suligen to stricter qualifications, than haw always hem rememburat. and in conditions which are the exeeption, not the rule.

The problem, to be discussed in this chapter, is as follows: Given a series of occasions, the probability \({ }^{2}\) of the occurrence of a certain event at each of which is known relative to certain initial data \(h\), on what proportion of these occasions may we reasomaly anficipate the wourrence of the ewent! (iiven, that is to say, the individual probability of each of a series of events \({ }^{a}\) priori, what statistical frequency of occurrence of these events is to be anticipated over the whole series? Beginning with Bernoulli's Theorem, we will consider the various solutions of this problem which have been propounded, and endeavour to

\footnotetext{



}
determine the proper limits within which each method has validity.
2. Bernoulli's Theorem in its simplest form is as follows: If the probability of an event's occurrence under certain conditions is \(p\), then, if these conditions are present on \(m\) occasions, the most probable number of the event's occurrences is \(m p\) (or the nearest integer to this), i.e. the most probable proportion of its occurrences to the total number of occasions is \(p\); further, the probability that the proportion of the event's occurrences will diverge from the most probable proportion \(p\) by less than a given amount \(b\), increases as \(m\) increases, the value of this probability being calculable by a process of approximation.

The probability of the event's occurring \(n\) times and failing \(m-n\) times out of the \(m\) occasions is (subject to certain conditions to be elucidated later) \(p^{\prime \prime} q^{\prime \prime-n}\) multiplied by the coefficient of this expression in the expansion of \((p+q)^{\prime \prime \prime}\), where \(p+q=1\). If we write \(n=m p-h\), this term is \((m p-h)!(m q+h)!p^{p^{\prime \prime} q^{\prime \prime-}-n}\). It is easily shown that this is a maximum when \(h=0\), i.e. when \(n=m p\) (or the nearest integer to this, where \(m p\) is not integral). This result constitutes the first part of Bernoulli's Theorem.

For the second part of the theorem some method of approximation is required. Provided that \(m\) is large, we can simplify the expression \(\frac{m!}{(m p-h)!(m q+h)!} p^{\prime \prime} q^{n \prime-x}\) by means of Stirling's Theorem, and obtain as its approximate value
\[
\frac{1}{\sqrt{2} \pi m p q} e^{2_{2}^{\prime 2}}
\]

As before, this is a maximum when \(h=0\), i.e. when \(n=m p\).
It is possible, of course, by more complicated formulae to obtain closer approximations than this. \({ }^{1}\) But there is an objection, which can be raised to this approximation, quite distinct from the fact that it does not furnish a result correct to as many places of decimals as it might. This is, that the approximation is independent of the sign of \(h\), whereas the original expression is not thus independent. That is to say, the approximation implies a symmetrical distribution for different values of \(h\) about

\footnotetext{
\({ }^{1}\) See, e.g., Bowley, Elements of Statistics, p. 298. The objection about to
} be raised does not apply to these closer approximations.
the value for \(h\) " ; white the expression umder approximation is unsymmetrical. It is analy sem that this want of symmetry is appreviable unless mpe" is large. We ought, therefore th have laid it down as a condition of our appoximation, not only that me must be large, hut alas that mpmost he large. I "ulikn mont of my criticisms, this is a mathematical, rather than a legrical point. I recur to it in § 15 .
"Par une fietion qui remdra les calculs phas faritus " (to quoti" Bertrand). we now replace the integer \(h\) be a comtimusus variatha and argue that the probability that the anomet of the divergence from the most probable value mp will lie het weon sand o rde. is
\[
\stackrel{1}{\sqrt{\prime}-\pi m v^{\prime} i^{\prime}} e^{z_{m}^{2}} z^{m} d z
\]

This 'fiction' will do no harm solong as it is remembered that we are now dealing with a particular kind of approximation. The probability that the divergence \(h\) from the most probable value \(m p\) will be less than some given quantity \(a\) is, therefore,
\[
\left.\begin{gathered}
1 \\
\sqrt{2}-\pi \mu!+1
\end{gathered} \right\rvert\, \quad 2 m p+i
\]

If we put \(\sqrt{\sqrt{2} m_{1} \mid \boldsymbol{l}}=t\), this is equal to


Thus, if we write \(a=\sqrt{ } / 2 m p q \%\) the probability \({ }^{1}\) thms tho number of occurrences will lie between
is measured by \(\mathrm{y}^{2}-\frac{-1}{-m} e^{\text {"d }} d\). This same expression measures

\footnotetext{

 li.. 1..1.... h yull \& 1 .


\[
\begin{array}{ll}
2 \\
1 & , j=m / y
\end{array}
\]

}
the probability that the proporion of occurrences will lie between

The different values of the integral \(\begin{gathered}2 \\ \sqrt{ } \pi \\ \int_{0}^{1} e\end{gathered}{ }_{0}^{\prime \prime \prime l}=-(-)(t)\) are given in tables. \({ }^{1}\)

The probability that the proportion of occurences will lie between given limits varies with the magnitude of \(\sqrt{2 p q} \begin{gathered}m\end{gathered}\), and this expression is sometimes used, therefore, to measure the 'precision' of the series. Given the à priori probabilities, the precision varies inversely with the square root of the number of instances. Thus, while the probability that the ubsolute divergence will be less than a given amount a decreases, the probability that the corresponding proportionate divergence (i.e. the absolute divergence divided by the number of instancei) will be less than a given amount \(b\), increases, as the number of instances increases. This completes the second part of Bernoulli's Theorem.
3. Bernoulli himself was not acquainted with Stirling's theorem, and his proof differs a good deal from the proof outlined in § 2. His final enunciation of the theorem is as follows: If in each of a given series of experiments there are \(r\) contingencies favourable to a given event out of a total number of contingencies 1 , so that \({ }_{t}^{r}\) is the probability of the event at each experiment, then, given any degree of probability c , it is possible to make such a number of experiments that the probability, that the proportionate number of the event's occurrences will lie between \(r+1\)
\(t\)\({ }^{r-1}{ }_{t}^{r-1}\), is greater than \(c .^{2}\)

Wahrscheinlichkeitsrechnung, vol. i. p. 121). As the whole formula is approximate, the simpler expression given in the text is probably not less satisfactory in practice. Seo also Czuber, Entwicklung, pp. 76, 77, and Eggenberger, Beiträge

\({ }^{1}\) A list of the principal tables is given by Czuber, loc. cit. vol. i. p. 122.
\({ }^{2}\) Ars Conjectandi, p. 236 (I have translated freely). There is a bricf account of Bernoulli's proof in Todhunter's History, pp. 71, 72. The problem is dealt with by Laplace, Throrie analytique livre ii. chap. iii. For an account of Laplace's proof see 'Todhunter's History, pp. 548-553.
4. We seem, therefore, to have proved that, if the \(\dot{a}\) priori probalility of an event under certain conditions is \(p\), the proportion of times most probable à priari for the event's occurrence on a series of occasions where the conditions are satisfied is also \(p\), and that if the series is a long one the proportion is very unlikely to differ widely from \(p\). This amounts to the principle which Ellis \({ }^{1}\) and Venn have employed as the defining axiom of probatility, save that if the series is 'long enough ' the proportion, according to them, will certuinly be \(p\). Laplace \({ }^{2}\) believed that the theorem afforded a demonstration of a general law of nature, and in his second edition published in 1814 he replaces \({ }^{3}\) the eloquent dedication, A J̌apoléon-le-firand, which prefaces the edition of 1812, by an explanation that Bernoulli's Theorem must always bring about the eventual downfall of a great power which, drunk with the love of conquest, aspires to a universal domination,-
 par de nombreuses et funestes expériences."
5. Such is the famous Theorem of Bernoulli which some have believed \({ }^{4}\) to have a universal validity and to he applicable to all 'properly calculated' probabilities. Yet the theorem exhibits algebraical rather than logical insight. And, for reasons about to be given, it will have to be conceded that it is only true of a special class of cases and requires conditions, hefore it can be legitimately applied, of which the fulfilment is rather the exception than the rule. For consider the case of a coin of which it is given that the two faces are either both heads or both tails : at every toss, provided that the results of the other tosses are unknown, the probability of heads is \(\frac{1}{2}\) and the probability of tails is \(\frac{1}{2}\) : yet the probability of \(m\) heads and \(m\) tails in \(2 m\) tosses
 given event be conrectly detormined, the evont will wh a long run of trials tend to recur with frequency proportional to this probability: This is generally
 unable to sever the judement that one event is mare likely to hapern than ancher from the belief that in the long run it will wewr more frequently."





Introdation. 11- lon. In
 view he adds: " It in false that the chancers must be realised in as series. It is, however, true that they mont probably will lex, and truo acain that this prob.

is zero, and it is certain \(\grave{a}\) priori that there will be either \(2 m\) heads or none. Clearly Bernoulli's Theorem is inapplicable to such a case. And this is but an extreme case of a normal condition.

For the first stage in the proof of the theorem assumes that, if \(p\) is the probability of one occurrence, \(p^{\prime \prime}\) is the probability of \(r\) occurrences rumning. Our discussion of the theorems of multiplication will have shown how considerable an assumption this involves. It assumes that a knowledge of the fact that the event has occurred on every one of the first \(r-1\) occasions does not in any degree affect the probability of its occurrence on the \(r\) th. Thus Bernoulli's Theorem is only valid if our initial data are of such a character that additional knowledge, as to the proportion of failures and successes in one part of a series of cases is altogether irrelevant to our expectation as to the proportion in another part. If, for example, the initial probability of the occurrence of an event under certain circumstances is one in a million, we may only apply Bernoulli's Theorem to evaluate our expectation over a million trials, if our original data are of such a character that, even after the occurrence of the event in every one of the first million trials, the probability in the light of this additional knowledge that the event will occur on the next occasion is still no more than one in a million.

Such a condition is very seldom fulfilled. If our initial probability is partly founded upon experience, it is clear that it is liable to modification in the light of further experience. It is, in fact, difficult to give a concrete instance of a case in which the conditions for the application of Bernoulli's Theorem are completely fulfilled. At the best we are dealing in practice with a good approximation, and can assert that no realised series of moderate length can much affect our initial probability. If we wish to employ the expression \(\frac{2}{\sqrt{\pi}} \int_{0}^{\gamma} e^{-t^{2}} d t\) we are in a worse position. For this is an approximate formula which requires for its validity that the series should be long; whilst it is precisely in this erent, as we have seen above, that the use of Bernoulli's Theorem is more than usually likely to be illegitimate.
6. The conditions, which have been described above, can be expressed precisely as follows:

Let \(x_{n}\) represent the statement that the event has occurred on \(m\) out of \(n\) occasions and has mot occurred on the others ；and let \({ }_{\mathbf{1}} x_{1} / h=p\) ，where \(h\) represents our id priori dulu，so that \(p\) is the à priori probability of the event in question．Bernoulli＇s Theorem then requires a series of conditions，of which the following is typical ：\(\ldots x_{1}, / x, h={ }_{1} x_{1} / h\) ，i．e．the probability of the exent on the \(n+1\) th oceasion must be unaffected by our knowledge of its proportionate frequency on the first \(n\) occasions，and must be exactly equal to its a priori probability before the first occasion．

Let us select one of these conditions for closer consideration． If \(y\) ，represents the statement that the esent has occurred on each

 must have \(y\) l！\({ }_{1} h=p\) for all values of \(s\) from 1 to \(r\) ．But in
 \(y_{\text {I }}\)／h \(\cdot p\) ．Bernoulli＇s Theorem，that is to say，temds，if it is carelessly applied，to exagereate the rate at which the probability of a given diverquce from the most probable decreases as the divergence increases．If we are given a panmy of which we have no reason to doubt the regularity，the probathility of heads at the first toss is ？：but if heads fall at evers one of the first 999 tosses，it becomes reatomable to estimate the probability of heads at the thousandth toss at much more than \(\frac{1}{2}\) ．For the it priori prohability of its being a conjurer＇s pemme or otherwise hiassed so as to fall heads almost invariably，is not usually so infinitesim－ ally small as \(\left(\frac{1}{2}\right)^{1 \times 4}\) ．We can only apply Bernoulli＇s Theorem with rigour for a predietion as to the pemye：behaviour over a series of a thousand tosses，if we have it priori such exhanstive knowledge of the penny＇s constitution and of the other com－ ditions of the problem that 999 heads rumning would not cause us to modify in any respect our prediction ì primri．

7．It seldom hapenens，therefore，that we can apply Bernoulli＇s Theorem with reference to a lons series of natural events．For in such cases we seldom possess the exhaustive knowleden which is necessary．Even where the serios is short，the perfectly rigorous application of the Theorem is not likely to be leqiti mate，and some degree of approximation will be involved in utilising its results．

Not so infrequently，howeyer．artificial serips ean he devised
in which the assumptions of Bernoulli's Theorem are relatively legitimate. \({ }^{1}\) Given, that is to say, a proposition \(a_{1}\), some series \(a_{1} a_{2} \ldots\) can be found, which satisfies the conditions:
\[
\begin{aligned}
& \text { (i.) } a_{1} / h=a_{2} / h \ldots=a_{/} / h . \\
& \text { (ii.) } a_{i} / a_{1} \ldots \bar{u}_{1} \ldots h=a_{r} / h .
\end{aligned}
\]

Adherents of the Frequency Theory of Probability, who use the principal conclusion of Bernoulli's Theorem as the defining property of all probabilities, sometimes seem to mean no more than that, relative to given evidence, every proposition belongs to some series, to the members of which Bernoulli's Theorem is rigorously applicable. But the natural series, the series, for example, in which we are most often interested, where the "'s are alike in being accompanied by certain specified conditions \(c\), is not, as a rule, rigorously subject to the Theorem. Thus 'the prolability of \(a\) in certain conditions \(c\) is \(\frac{1_{2}}{}\) ' is not in general equivalent, as has sometimes been supposed, to ' It is 500 to 1 that in 90,000 occurrences of \(c, a\) will not occur more than 20,200 times, and 500 to 1 that it will not occur less than 19,800 times.'
8. Bernoulli's Theorem supplies the simplest formula by which we can attempt to pass from the \(\grave{a}\) priori probabilities of each of a series of events to a prediction of the statistical frequene \(y\). of their occurrence over the whole scries. We have seen that Bernoulli's Theorem involves two assumptions, one (in the form in which it is nsually enunciated) tacit and the other explicit. It is assumed, first, that a knowledge of what has occurred at some of the trials would not affect the probability of what may occur at any of the other's ; and it is assumed, secondly, that these probahilities are all equal ì priori. It is assumed, that is to say, that the probability of the event's occurrence at the \(r\) th trial is equal à priori to its probability at the \(n\)th trial, and, further, that it is unaffected by a knowledge of what may actually have occurred at the \(n\)th trial.

A formula, which dispenses with the explicit assumption of equal ì priori probabilities at every trial, was proposed by Poisson, \({ }^{2}\) and is usually known hew his name. It does not dispense,

\footnotetext{
1 In the discussion in Chapter XVI., p. 170, of the probability of a divergence from an equality of heads and tails in coin-tossing, an example has been given of the construction of an artificial series in which the application of Bernoulli's Theorem is more leqitimate than in the natural series.

}
however, with the other inexplicit assumption. The difference between Poisson's Theorem and Burnoulli's is hest shown by reference to the ideal case of balls drawn from an urn. The typical example for the valid application of Bernoulli's Theorem is that of balls drawn from a single urn, containing black and white balls in a known proportion, and replaced after each drawing, or of talls drawn from a series of urns, each containing black and white balls in the same known proportion. The typical example for Poisson's Theorem is that of balls drawn from a series of urns, each containing black and white balls in different known proportions.

Poisson's Theorem may be enunciated as follows: \({ }^{1}\) Let 8 trials be made, and at the \(\lambda\) th trial \((\lambda=1,2 \ldots s)\) let the probalilities for the oecrurence and non-occurrence of the event be \(p_{\lambda}, q_{\lambda}\) respectively. Then, if \({ }^{-P_{\lambda}}=p\), the probability that the number of occurrences \(m\) of the event in the \(s\) trials will lie between the limits \(s p \pm l\) is given by

where \(k\)
\[
2-11^{\prime \prime}
\]
 in a form corresponding to that of Bernonlli's Theorem, \({ }^{2}\) namely :

The probability that the number of occurrences of the event will lie between sp土rk, /s is given bu

9. This is a highly ingenious theorem and extends the application of Bernoulli's results to some important types of cases. It embraces, for example, the case in which the successive terms of a series ate drawn from distinct populations known to be characterised by differing statistical frequencies; no further com-

\footnotetext{



}
plication being necessary beyond the calculation of two simple functions of these frequencies and of the number of terms in the series. But it is important not to exaggerate the degree to which Poisson's method has extended the application of Bernoulli's results. Poisson's Theorem leaves untouched all those cases in which the probabilities of some of the terms in the series of events can be influenced by a knowledge of how some of the other terms in the series have turned out.

Amongst these cases two types can be distinguished. In the first type such knowledge would lead us to discriminate between the conditions to which the different instances are subject. If, for example, balls are drawn from a bag, containing black and white balls in known proportions, and not replaced, the knowledge whether or not the first ball drawn was black affects the probability of the second ball's being black because it tells us how the conditions in which the second ball is drawn differ from those in which the first ball was drawn. In the second type such knowledge does not lead us to discriminate between the conditions to which the different instances are subject, but it leads us to modify our opinion as to the nature of the conditions which apply to all the terms alike. If, for instance, balls are drawn from a bag, which is one, but it is not certainly known which, out of a number of bags containing black and white balls in differing proportions, the knowledge of the colour of the first ball drawn affects the probabilities at the second drawing, because it throws some light upon the question as to which bag is being drawn from.

This last type is that to which most instances conform which are drawn from the real world. A knowledge of the characteristics of some members of a population may give us a clue to the general character of the population in question. Yet it is this type, where there is a change in knowledge but no change in the material conditions from one instance to the next, which is most frequently overlooked. \({ }^{1}\) It will be worth while to say something further about each of these two types. \({ }^{2}\)

\footnotetext{
\({ }^{1}\) Numerous instances could be quoted. To take a recent English example, reference may be made to Vule, Introduction to the Theory of Statistics, p. 251. Mr. Yule thinks that the condition of independence is satisfied if "the result of any one throw or toss does not affect, and is unaffected by, the results of the preceding and following tosses," and does not allow for the cases in which knowledge of the result is relevant apart from any change in the physical conditions.
\({ }^{2}\) The types which I distinguish under four heads (the Bernoullian, the
}
10. Fow prohtem- of the first type. where there is phy dial or material dependence between the successive trials, it is not possible, I think, to propose any general solution; since the
 of different ways. But for particular problems, if the conditions are precise enough, solutions can be devised. The problem, for instance, of an urn, containing black and white balls in known proportions, from which balls are drawn successively and not replaced, \({ }^{1}\) is ingeniously solved by Czuber \({ }^{2}\) with the aid of Stirling's Theorem. If \(\sigma\) is the number of balls and \(s\) the number of drawings, he reaches the interesting conclusion (assuming that \(\sigma, s\) and \(\sigma-s\) are all larme) that the probability of the number of black balls lying within given limits is the same as it would be if the balls were replaced after cach drawing and the number of drawings were \({ }_{\sigma} s\) instead of \(s\).

In addition to the assumptions already stated, Profession Czuber's solution applies only to those cases where the limits, for which we wish to determine the probability, are narrow compared with the total number of black balls \(p \sigma\). Professor Pearson \({ }^{3}\) has worked out the same problem in a much more general manner, so as to deal with the whole range, i.e the frequency or prol)ability of all possible ratios of hlack balls, even where \(s>p \sigma\). The various forms of curve, which result, according to the different relations existing between \(p, s\), and \(\sigma\), supply examples of each of the different types of frequency curve which arise out of a

\footnotetext{
 p. 15.7) (lassifies as follows:
 fon mill:


} at the earher stames, it has indipendanee:
(iii.) When they vary in a manner which depemeds upen what has happomel at the emerlierstates, it his comnerite.

 as a continmons variahla. This is the same hind of assumptom as that mowhe
 tions, -Ger rather the value of the resultes is limiteal in the same was
 rephaced, or are drawn simbltamesusly.


classification according to (i.) skewness or symmetry, (ii.) limitation of range in one, both or neither direction ; and he designates, therefore, the curves which are thus obtained as generalised probability curves. His discussion of the properties of these curves is interesting, however, to the student of descriptive statisties rather than to the student of probability. The most generalised and, mathematically, by far the most elegant treatment of this problem, with which I am acquainted, is due to I'rofessor Tschuprow. \({ }^{1}\)

Poisson, in attempting a somewhat similar problem, \({ }^{2}\) arrives at a result, which seems obviously contrary to good sense, by a curious, but characteristic, misapprehension of the meaning of 'independence' in probability. His problem is as follows: If \(l\) balls be taken out from an urn, containing \(c\) black and white balls in known proportions, and not replaced, and if a further number of balls \(\mu\) be then taken out, the probability that a given proportion \(\frac{m}{m+n}\) of these \(\mu\) balls will be black is independent of the number and the colour of the \(l\) balls originally draun out. For, he argues, if \(l+\mu\) balls are drawn out, the probability of a combination, which is made up of \(l\) black and white balls in given proportions followed by \(\mu\) balls, of which \(m\) are white and \(n\) hlack, must be the same as that of a similar combination in which the \(\mu\) balls precede the \(l\) balls. Hence the probability of \(m\) white balls in \(\mu\) drawings, given that the 7 balls have already been drawn out, must be equal to the probability of the same result, when no balls have been previously drawn out. The reader will perceive that Poisson, thinking only of physical dependence, has been led to his paradoxical conclusion hy a failure to distinguish between the cases where the proportion of black and white balls amongst the \(l\) balls originally drawn is known and where it is not. The fact of their having been drawn in certain proportions, provided that only the total number drawn is known and the proportions are unknown, does not influence the probability. Poisson states in his conclusion that the probability is inderendent of the number and colour of the \(l\) halls originally drawn. If he had added-as he ought-' provided the number of each colour is

\footnotetext{
1 "Zur Theorie der Stabilität statistischer Reihen," p. 216, published in the Skundinavisk Aktuarietidskrift for 1919.
\({ }^{2}\) Loc. cit. pp. 231, 23:.
}
untineme,' the air of paratox disappears. This is an exceedingly sood example of the failure tw pererine that a probability canmot be influenced by the emourrence of a material event but only by such linouledge, as we max hate, mopecting the occurrence of the event. \({ }^{1}\)
11. For problens of the second type, where knomledgen of the result of one trial is capable of intluencing the probability at the next apart from any chane in the material conditions, there is, likewise, no gomeral sulution. The following artificial example. however, will illustrate the onrt of considerations which are involved.

In the casm where Bornoulli's Theorem is applied to practical questions, the it primi probatility is generally obtained empirically the reformen to the statistical frequency of each alternative in past experience under apparently similar conditions. Thus the \(a_{\text {primer prob }}\) patility of a male hith is estimated by reference to the remordat propertion of male births in the past. \({ }^{2}\) The validity of sotimatine probahilities in this manner will be discussa! later. But for the purpers of this example let us as-sume that the if primer pmbatility has been calculatent on this basis. Thus the à priori probability \(p\left(=\frac{r}{s}\right)\) of an event is based on the observation of its occurrence \(r\) times out of \(s\) occasions on which the given conditions were present. Now, according to Bermolli: Theorem directly applied, the probability of the event's occurring \(n\) times running is \(p^{\prime \prime}\) or \(\binom{r}{s}\). But, if the event occurs at the first trial, the probability at the second

\footnotetext{

 solutions of this chapter are vitiated by his assuming in the course of them both that certain quantities are very large, and also, at a later atage, that the same quantities are infinitesimal. On this account, for exmmple, his solution of the following difficult froblem breaks down: (iven an urn A with \(m\) white and \(n\) black halls and an urn \(B\) with \(m^{\prime}\) white and \(n^{\prime}\) hack balls, if at cach move a ball is taken from A and put into B, and at the same time a ball is taken from Band put int.. 1 , what is the probability after \(x\) moves that the urne \(A\) and \(B\) shall have a given composition?
 a priori value to the chance \(p\) (i.e. of a male birth) as in the case of dice-throwing, but it is quite sulficiently accurate for practicnl purpeses to use the propertion of male birthe netually ohserved if that propertion be based on a mederately large number of olservations."
}
becomes \(\begin{aligned} & r+1 \\ & s+1\end{aligned}\), and so on. Hence the probability P, properly calculated, of \(n\) successive occurrences is
\[
i_{s}^{r+1} s+{ }_{s+2}^{r+2} \cdots \frac{r-n}{s+n-1} .
\]

Hence
\[
\begin{aligned}
& \mathrm{P}=\frac{(r: n-1)!(s-1)!}{(s+n-1)!(r-1):}
\end{aligned}
\]

Theorem, provided that \(r\) and \(s\) are large;
\[
\begin{aligned}
& \left.=\binom{\prime}{s}^{\prime \prime}\binom{n-1}{r}^{\prime \prime}{ }_{s}^{\prime \prime-1}\right)^{\prime \prime}
\end{aligned}
\]

Thus, in this case, the assumption of Bernoulli's 'Theorem is approximately correct, only if \(Q\) is nearly unity. This condition is not satisfied unless \(n\) is small both compared with \(r\) and compared with \(\varepsilon\). It is very important to notice that two conditions are involved. Not only must the experience, upon which the à priori probability is based, be extensive in comparison with the number of instances to which we apply our prediction; but also the number of previous instances multiplied by the probability based upon them, i.e. \(s p(=r)\), must be large in comparison with the number of new instances. Thus, even where the prion experience, upon which we found the initial probability \(P\), is very extensive, we must not, if \(P\) is very small, say that the probability of \(n\) successive occurrences is approximately \(p{ }^{\prime}\), unless \(n\) is alvo small. Similarly if we wish to determine, by the methods of Bernoulli, the probability of \(n\) occurrences and \(m\) failures on \(m\) i \(n\) occasions, it is necessary that we should have \(m\) and \(n\) small
compared with \(s\), \(n\) amatl comprated with \(r\), and \(m\) small conprared with \(s-r{ }^{1}\)

The case solved abowe is the rimpleat pereible. The eremeral problem is as follows: If an event haw necurred a times in tho first \(y\) trials. its probatility at the \(y: 1\) th is \(\begin{array}{r}r+x \\ s: y\end{array}\); determiter the a priori probability of the events securring \(p\) times in 4 trials.
 have \(\phi(p, q)=\begin{array}{rl}r+p & 1 \\ s+q & 1\end{array}(p-1 \cdot q-1) \quad s+q-1-r-p^{\prime}(p, y-1)\). I know of no solution of this. even approximate. But we may say that the comditions are these of supermormal dianerame as compared with Bermulli's comditions. That is tusay, the pult, ability of a propertion difiering widely from \({ }_{s}^{r}\) is greater than in Bermoullian comditions; for when the promenton begime 10 diverye it becomes more probable that it will comtinue to diw-ry. in the same direction. If, on the other hand, the conditions. of the problem had hern surh, that when the proportion besins tw diverge it becomes more probable that it will recoper itsell and tend hack fowards: \({ }^{r}\) (as when wr. draw halls without replacing thern from a bate of knewn (omperition), we should have subnormal dispersion. \({ }^{2}\)
12. The condition flucidated in the freceding parapmph i frequently werlowhed by statisticiats. The following "xample from Czuber \({ }^{3}\) will be sulliciont for the purpose of illu-tmant. Czuber's argument is as follows:

In the period 1866-1877 there were registered in Aust ria
\[
\begin{aligned}
& s=8,30: 3,269 ;
\end{aligned}
\]
\({ }^{1}\) This praractuph is concerned with a differont perint from that doalt with in Professor Pearson's article " On the Influcnee of P'ast Fixperienee on Future Expectation," to which it hears a superticial resemblance. P'rufeston floarmon' article which deals, net with Bernoulli's Theorem, but with Latplace's " Rule of Succession," will tre referred to in \$ 16 of this chapter and in \(\$ 12\) wf the wext.


 he is usually so carcfal an expenmet of theoretion statintios.
for the succeeding period, 1877-1899, we are given only
\[
m^{\prime}=6,533,961 \text { male births ; }
\]
what conclusion can we draw as to the number \(n\) ' of female births? We can conclude, according to Cyiber, that the most probable value
\[
n_{0}^{\prime}=\frac{n m^{\prime}}{m}=6,141, \tilde{2} 87,
\]
and that there is a probability \(\mathrm{P}=\cdot 9999779\) that \(n^{\prime}\) will lie between the limits \(6,118,361\) and \(6,164,813\).

It seems in plain opposition to good sense that on such evidence we should be able with practical certainty \(\left(\mathrm{P}=.9999779=1-\begin{array}{c}1 \\ 45250\end{array}\right)\) to estimate the number of female births within such narrow limits. And we see that the conditions laid down in § 11 have been llagrantly neglected. The number of cases, over which the prediction based on Bernoulli's Theorem is to extend, actually exceed's the number of cases upon which the \(a\) priori probability has been based. It may be added that for the period, 1877-1894, the actual value of \(n^{\prime}\) did lie between the estimated limits, but that for the period, 18951905, it lay outside limits to which the same method had attributed practical certainty.

That Professor Czuber should bave thought his own argument plausible, is to be explained, I think, by his tacitly taking account in his own mind of evidence not stated in the problem. He was relying upon the fact that there is a great mass of evidence for believing that the ratio of male to female births is peculiarly stable. But he has not brought this into the argument, and he has not used as his à priori probability and as his coefficient of dispersion the values which the whole masis of this evidence would have led him to adop,t. Would not the argument have secmed very preposterous if \(m\) had heen the number of males called George, and \(n\) the numier of fomalen called Mary? Would it not have seemed rather preposterous if \(m\) had been the number of legitimate births and \(n\) the number of illegitimate births? (learly we must take account of other considerations than the mere numerical values of \(m\) and \(n\) in estimating our \(\grave{a}\) priori probability. But this question belongs to the sulject-matter of later chapters,
and. quite apart from the manner of calculatime of the is fume probability, the argument is invalidated by the fant than ant a priori probability foumded on 8.3a3, 26: instances. withont corroborative evidence of a mom-statistical character. cammit be assumed stahbe through a calculation which astent- inor 12,700,000 instances.
13. Before we leave the theorems of Bernoulli and Poisson, it is necessary to call attontion to a sery monarhathe theormm ho Tchebycheff, from which both of the above theorems can be derived as special cases. This result is reached rigorously and without approximation, by turans of simple alemhta atind without the aid of the differential calculus. Apart from the beauty and simplicity of the proof, the theorem is so valuable and so little known that it will be worth while to quote it in full : 1

Let \(x, y, z \ldots\) represent certain magnitudes, of which \(x\) can take the values \(x_{1} x_{2} \ldots x_{1}\) with probabilities \(p_{1} p_{2} \ldots p\) respectively, \(y\) the values \(y_{1} y_{2} \ldots y\), with probabilities \(q_{1} q_{2} \ldots q\), \(z\) the values \(z_{1} z_{2} \ldots z_{\text {m }}\) with probatilities \(r_{2} r_{2} \ldots r_{\text {, and }}\) an \(\quad(m\). so that

Writ.
and
so that we can describe \(a\) as the mathematical expectation or average value of \(x\), and \(a_{1}\) as the mathematical expectation or average value of \(x^{2}\), etc.

The probability that the sum \(x+y+z+\ldots\) will have for its value \(x_{\kappa}+y_{\lambda}+z_{\mu}+\ldots\) is \(p_{\kappa} y_{\lambda} r_{\mu} \ldots\) (provided that the values of \(x, y, z \ldots\) are independent). Hence

\footnotetext{



 in Russian. It was not easily aceessible, therefore, until the pul, lieation at letrograd in 1907 of the colleceted edition of his works in Fronch. His theorems are, consequently, not nearly so well known as they doberve to tre, although his most impertant theorems were reproduced from time to time in

}
\[
\Sigma\left(z_{k}+y_{\lambda}+z_{\mu}+\ldots-a-h-c-\ldots\right)^{2} \prime_{k}^{\prime} I_{\lambda^{\prime}} \mu_{\mu} \ldots
\]
summed for all values of \(\kappa, \lambda, \mu\) is the average expectation for

Now
\[
\begin{aligned}
& \left(x_{\kappa}+y_{\lambda}+z_{\mu}+\ldots-a-b-c-\ldots\right)^{2} .
\end{aligned}
\]
\[
\begin{aligned}
& =u_{1}-2 u^{2}+u^{2}=u_{1}-u^{2} .
\end{aligned}
\]

Also \(\Sigma q_{\lambda}{ }_{\lambda}{ }_{\mu} \ldots=1\), summed for all values of \(\lambda, \mu \ldots\), and
\[
\begin{aligned}
& \Sigma 2\left(x_{k}-a\right)\left(y_{\lambda}-l\right)_{p_{k}}=\underset{1}{2} 2\left(x_{k} y_{\lambda}-7 p_{r_{k}}-\left(n y_{\lambda}+a b\right)_{p_{k}}\right.
\end{aligned}
\]
\[
\begin{aligned}
& =2\left(u y_{\lambda}-a b-u y_{\lambda}+c\left(l^{\prime}\right)=0 .\right.
\end{aligned}
\]

\[
=l_{1}+l_{1}+c_{1}+\ldots-u^{2}-l^{2}-c^{2}-\ldots,
\]
whence
\[
\begin{aligned}
& a^{2}\left(\mu_{1}+b_{1}+r_{1}+\ldots-u^{2}-l^{2} \quad r^{2} \ldots\right) \quad a^{2}
\end{aligned}
\]
where the summation extends over all values of \(\kappa, \lambda, \mu \ldots\) and \(a\) is some arbitrary number greater than unity.

If we omit those terms of the sum on the left-hand side of the above equation for which
\[
\begin{aligned}
& \left(w_{k}+y_{\lambda}+z_{\mu}+\ldots-a-h \quad r-\ldots\right)^{2} \\
& a^{2}\left(\prime_{1}+h_{1}+r_{1}+\ldots-i^{2}-i^{2}-r^{2}-\ldots\right)
\end{aligned}
\]
and write unity for this expression in the remaining terms, both these processes diminish the magnitude of the left-hand side. Hence \(\because p_{k} q_{2} r_{\mu} \ldots<_{a^{2}}^{1}\), where the summation covers those sets of values only for which
\[
\begin{gathered}
\left(r_{k} y_{A}+i_{1}+\ldots-d l-l^{\prime} \ldots\right)^{2} \\
n^{2}\left(\prime_{1}+l_{1}+c_{1}+\ldots-t^{2} \cdots l^{2}-r^{2} \ldots\right)
\end{gathered}
\]

If \(P\) is the probability that
\[
\begin{gathered}
\left.\left(r_{k}+!_{\lambda}+:_{k}+\ldots \cdots a-l\right)-c \ldots\right)^{2} \\
\left(1^{2}-1_{1}+l_{1}+l_{1}+\ldots-1^{2}-l_{1}-r^{2}-\ldots\right)
\end{gathered}
\]
is equal to or less than unity, it follows that
\[
\begin{array}{cc}
1 & 1 \\
1 & 1 \\
1 & 1 \\
& 1
\end{array}
\]

Hence the probability that the sum


is greater than \(1 \ldots\), where \(a\) is some number greater than unity.

This result constitutes 'Tchebychefl's Theorem. It may also be written in the following form:

Let \(n\) be the number of the magnitudes \(x, y, z \ldots\), and write \(u=\wedge^{n}\); then the probability that the arithmetic mean \({ }^{A_{*}} y_{\lambda} \quad z_{\mu}: \cdots\) lies between the limits

is greator than I

It is also easy to show \({ }^{1}\) as a deduction from Thehereheff: Theorem that, if an amount \(A\) is won when an event of prohathilits \(p \mu=1 \quad y \mid\) occurs and an amount B lost when it fails, then in \(s\) trials the probability that the total wimnings (or lossess) will lie between the limits
\[
\text { 1.1 } 131111 \text { 13, }
\]
is greater than 1 a
14. From this very romeral result for the prohable limits of as sum comporad of at mamber of independently varyine magnitudes. Bermoultis Theorem is easily derised. For let there be
\(s\) observations or trials, and \(s\) magnitudes \(x_{1} x_{2} \ldots x\) corresponding, such that \(x=1\) when the event under consideration occurs, and \(x=0\) when it fails. If the probability of the event's occurrence is \(p\), we have \(a=p, b=p\), etc., and \(a_{1}=p, b_{1}=p\), etc. Hence the probability 1 ' that the number of the event's occurrences will lie between the limits \(s p \pm a \sqrt{ } s p-s p^{2}\), i.e. between the limits \(s p \pm a \sqrt{ } s p q\) where \(q=1-p\), is \(>1-\frac{1}{a^{2}}\). If we compare this formula with the formula for Bernoulli's Theorem already given, we find that, where this formula gives \(\mathrm{P}>1-\stackrel{1}{a^{2}}\), Bernoulli's Theorem with greater precision gives \(P=(-)\left(\begin{array}{c}a \\ 1 \\ 1\end{array}\right)\). The degree of superiority in the matter of precision supplied by the latter can be illustrated by the following table:


Thus when the limits are narrow and \(a\) is small, Bernoulli's formula gives a value of P very much in excessis of \(1-\frac{1}{a^{2}}\). But Bernoulli's formula involves a process of approximation which is only valid when \(s\) is large. Tchebycheff's formula involves no such process and is equally valid for all values of \(s\). We have seen in \(\S 11\) that there are numerous cases in which for a different reason Bernoulli's formula exaggerates the result:, and, therefore, Tchebychefl's more cautious limits may sometimes prove useful.

The deduction of a corresponding form of Poisson's Theorem from Tehehycheff's general formula obviously follows on similar lines. For we put \({ }^{1} a=p_{1}, b=p_{2}\), etc., and \(a_{1}=p_{1}, b_{1}=p_{2}\), etc.,

\footnotetext{
\({ }^{1}\) I am using the same notation as that used for Poisson's 'Theorem in § 8.
}
and find that the probability that the number of the event's occurrences will lie between the limits
ie. between the limits \(\because:=\backslash \because\),

is greater than \(t-\frac{1}{n^{2}}\).
 directly by a method similar to his general method, and also obtains several supplementary results such as the following :
I. If the chances of an event E in \(\mu\) consecutive trials are \(p_{1} p_{2} \ldots p_{0}\) respectively, and their sum is \(s\), the probability that E will occur at least \(m\) times is less than

pmovidmithat \(\ln ^{\prime \prime}\) *
II. and the probability that \(E\) will not occur more than \(n\) times is less than
\[
\text { provided that ". }-1 \text {. }
\]
 and more than \(n\) is creator than

\[
\text { pro.ininl } \mu * 1 . /
\]
15. Tehebyeheff's methods have been set out and his results admirably extended by A. A. Markoff. \({ }^{2}\) And some develop-

\footnotetext{




}
ments along the same lines by Tschuprow ("Zax Theorie der Stabilität statistischer Reihen," Skandinavisk Aktuarielidishrift, 1919) have convinced me that Tchebycheff"s discovery is far more than a technical device for solving a special problem, and points the way to the fundamental methoul for attacking these questions on the mathematical side. The Laplacian mathematics, although it still holds the field in most text-hooks, is really obsolete, and ought to be replaced by the very beautiful work which we owe to these three Russians.
16. There is one other investigation relating to Bernoulli's Theorem which deserves remark. I have already pointed out, in \(\$ 2\), that the dispersion about the most probable value, even when the conditions for the applicability of Bernoulli's Theorem in its non-approximate form are strictly fulfilled, is unsymneetrical. The fact, that the usual approximation for the probability of a divergence \(h\) from the most probable number of occurrences (the notation is that of \(\$ 2\) above) takes the form
\(1 e^{-i n}{ }^{\prime \prime} p q\), which is the same for \(+h\) as for \(-h\), has led / \(2 \pi m p q^{\prime}\)
to this want of symmetry being very generally overlooked; and it is not uncommon to assume that the probability of a given divergence less than \(p_{m}\) is equal to that of the same divergence in excess of pim, and, in general, that the probahility of the frequency's excceding \(p m\) in a set of \(m\) trials is comel to that of its falling short of \(p m\).

That this is not strictly the case is ob, ious. If a die is cast (60 times, the most probable number of : appearances of the ace is 10 ; but the ace is more likely to appear 9 times than 11 times; and much more likely (about 5 times as likely) not to appear at all than to aprear exactly 20 times. That this mest be so will be clear to the reader (without his requiring to trouble himself with the algelora), when he reflects that the ace cannot appear less often than not at all, whereas it may well appear more than 20 times, so that the smallness of the possible divergence in defect from the most probable value 10 , as compared with the possible divergence in excess, must be made up for by the greater

Tehebychetl's leading idea. Further references to later memairs, which, being in the Russian language, are inaccessible to me, will be found in the Bibliographis:
frequency of any wiven defection at comparal with the corre． sponding excess．Thus the act and frequency in a serime of trials of an event，of which the prohahility at each trial is leos than \(\frac{1}{2}\) ． is likely to fall short of its mest prablable value meme often than it exceeds it．What is in fact true is that the mathematical expertation of deficmoney is empal to the mathematical expecta tion of excess，i．e．that the sum of the possible deficiencies each multiplied by its probability is equal to the sum of the possible excesses each multiplied by its probability：

The actual measurement of this want of symmetry and the determination of the conditions，in which it can be safely neglected，involves laborious mathematics，of which I am only acquainted with one direct investigation，that published in the



For the details of the proof I must refer the reader to Mr． Simmons＇s article．His principal theorem \({ }^{2}\) is as follows：If 1 is the probability of the event at each trial and \(n(a+1)\) the \(a \div 1\) number of trials，\(n\) and \(a\) being integers，\({ }^{3}\) the probability that the frequency of occurrence will fall short of \(n\) is always greater than the probability that it will exceed \(n\) ；the difference between the two probabilities being a maximum when \(n=1\) ，constantly diminishing as \(n\) increases，lyine always between \(\begin{aligned} & 1 a-1 \\ & 3,1 \\ & a+1\end{aligned}\) times the greatest term in \(\left(\begin{array}{c}a \\ a+1\end{array} \frac{1}{a+1}\right)^{\prime}\) and \(\frac{1}{3} a-1\) times the

\footnotetext{
 his investigution，and so far as I know this cham is justitied：but rocent investigations obtaining closer approximations to Burnoulli＇s Thoorem by means of the Method of Moments are esentially directed towards the samo problem．

A somewhat analogons print has，however，been raised by l＇rofessor l＇earson
 Expectation．＂He brings out an exactly similar want of symmetry in the probabilities of the various possible frequencies about the most probabla fres． quency，when the calculation is based，not on Bornomlli＇s Theorem as in Mr． Simmons＇s insestigation，but on Laplace＇s rule of succession（sere next hapter）． The want of symmetry has also leen peinted out by I＇rofessor la＇xis（．Ih／uand－

\({ }^{2} 1\) am not giving his wwn enunciation of it．
\({ }^{3}\) Mr．Simmens dones mut seem to have heren ahle tor remmen this restriction on the generality of his theorem，hat there does now semen mush reasum to donate that it can le rennowad．
}
greatest term in \(\left(\begin{array}{c}a \\ a+1\end{array} \frac{1}{a+1}\right)^{(a+1)(1+1)}\), and being approximately equal, when \(n\) is very large, to \(\frac{1}{3} \frac{a-1}{\sqrt{2 \pi n a(a+1)}}\).

The following table gives the value of the excess \(\Delta\) of the probability of a frequency less than pm over the probability of a frequency greater than \(p i n\) for various values of \(p\) the probability and \(m\) the number of trials \(\left[p=\begin{array}{c}1 \\ (1+1\end{array}, m=n(a+1)\right]\), as calculated by Mr. Simmons:
\begin{tabular}{|c|c|c|}
\hline \({ }^{\prime \prime}\) & \({ }^{\prime \prime}\). & د. \\
\hline 1 & & \\
\hline 3 & 3 & -037037 \\
\hline 1 & & \\
\hline ; & 15 & -02243662 \\
\hline 1 & & \\
\hline : & 24 & - 0182706 \\
\hline 1 & & \\
\hline 4 & 4 & -054687 \\
\hline 1 & & \\
\hline 4. & 20 & -03201413 \\
\hline 10 & 10 & \\
\hline \[
\begin{gathered}
10 \\
1
\end{gathered}
\] & 10 & \(\cdot 084777\) \\
\hline 10 & 20 & .068673713 \\
\hline 1 & & \\
\hline 100 & 100 & -101813 \\
\hline 1 & 200 & \(\cdot 081324387\) \\
\hline 100
1 & 200 & -081324387 \\
\hline 1000 & 1000 & \(\cdot 103454\) \\
\hline
\end{tabular}

Thus unless not only \(m\) but \(m p\) also is large the want of symmetry is likely to be appreciable. Thus it is easily found that in 100 sets of 4 trials each, where \(p \int_{1}^{1}\), the actual frequency is likely to exceed the most probable 26 times and to fall short of it 31 times; and in 100 ) sets of 10 trials carch, where \(p=\frac{1}{10}\), to exceed 26 times and to fall short 34 times.

Mr. Simmons was first directed to this investigation through
noticing in the examination of sets of random diems that " "and digit presented itwelf, with umexperted framumes. lose that \(\frac{1}{10}\) of the number of times. Fin instance, in lonsponf lian digits earh, I foumd that a digit presentad itself in a ret more fropumble umber 15 times than over 15 times; similarly in the ease of 80 sets each of 250 digits, and also in other aggregations." Its possible bearing on such experiments with dice and roulette, as are described at the end of this chapter, is clear. But apart from these artificial experiments. it is sometimes worth ther statistician's while to bear in mind this apprectablo want of symmerty in the distribution about the mode or most probable value in many even of those cases in which Bernoullian conditions are strictly fulfilled.
17. I will conclude this chapter by an account of some of the attempts which have been made to verify a posterior the conclusions of Bernoulli's Theorem. These attempts are nearly useless, first, because we can seldom be certain a priori that the conditions assumed in Bemoullis Themenn ane fultilled, ambl secondly: becaluse the theorem predicts mot what will happen but only what is, on certain evidence, likely to happen. Thus even where our results do not verify Bernoulli's Theorem, the theorem is not thereby discredited. The results have bearing on the conditions in which the experiments took place, rather than upon the truth of the theorem. In spite, therefore, of the not unimportant place which these attempts have in the history of probability, their scientific value is very small. I record them, because they have a good deal of historical and psychological interest, and because they satisfy a certain idle curiosity from which few students of probability are altogether free. \({ }^{1}\)
18. The duta for these investigations have been principally drawn from four sources - coin-tossing, the throw of dice, lotteries, and roulette; for in such cases as these the comditions for Bernoulli's Theorem seem to be fulfilled most nearly. The earliest recorded experiment was carried out hy Bution, who, assisted

\footnotetext{

 periments personally, in order to acquire confidence in the use of the theory:. Mr. Vule himself has indulged moderately.


}
by a child tossing a coin into the air, played 2048 partis of the Petersburg game, in which a coin is thrown successively until the parti is brought to an end by the appearance of heads. The same experiment was repeated by a young pupii of De Morgan's 'for his own satisfaction.' \({ }^{1}\) In Buffon's trials there were 1992 tails to 2048 heads; in Mr. II.'s (De Morgan's pupil) 2044 tails to 2048 heads. A further experiment, due to Bution's example, was carried out by Quetelet \({ }^{2}\) in 1837. He drew 4096 balls from an urn, replacing them each time, and recorded the result at different stages, in order to show that the precision of the result tended to increase with the number of the experiments. He drew altogether 2066 white balls and 2030 black balls. Following in this same tradition is the experiment of Jevons, \({ }^{3}\) who made 2048 throws of ten coins at a time, recording the proportion of heads at each throw and the proportion of heads altogether. In the whole number of 20,480 single throws, he obtained heads 10,353 times. Hore recently Weldon \({ }^{4}\) threw twelve dice 4096 times, recording the proportion of dice at each throw which showed a number greater than three.

All these experiments, however, are thrown completely into the shade by the enormously extensive investigations of the Swiss astronomer Wolf, the earliest of which were published in 1850 and the latest in \(1893 .{ }^{5}\) In his first set of experiments Wolf completed 1000 sets of tosses with two dice, each set continuing until every one of the 21 possible combinations had necurred at least once. This involved altogether 97,899 tosses, and he then completed a total of 100,000 . These data enabled him to work out a great number of calculations, of which Czuber quotes the following, namely a proportion of \(\cdot 83533\) of unlike pairs, as against the theoretical value \(\cdot 83333\), i.e. \({ }_{6}^{5}\). In his second set of experi-
\({ }^{1}\) Formal Logic, p. 185, published 1847. De Morgan gives Buffon's results, as well as his pupil's, in full. Buffon's results are also investigated ly Poisson, Recherchtes, p. 1: 1:2-135.
\({ }^{2}\) Letters on the Theory of Probabilities (Eng. trans.), p. 37.
\({ }^{3}\) Principles of Science (2nd ed.), p. 208.
" Quoted by Edecworth, "Law of Error" (Vmely. Jicit. Ioth cal.), and by Yule, Introduction to Statistics, p. 254.
\({ }^{5}\) See Bibliography. Of the earliest of these investigations I have no firsthand knowledge and have relied upon the account given by Czuher, loc. cit. vol. i. p. 149. Fur a general account of empirical verifications of Bernoulli's Theorem reference may be made to Czuber, Wahrscheinlichkeitsrechnung, vol. i. pp. 139-152, and ('zuber, Entwicklung der Wuhrscheinlichkeitstheorie, pp. 88-91.
ments Wolf used two dice, one white and one red (in the first set the diee were imblistinguishable), and completed 20.000 tosses, the details of pach result being recorded in the Vierteljahrsichrift der Nourufirserkemen diesellschuft in Zïrich. He studied particularly the cumber of aequensern with each die, and the relative frequency of each of the 3 p prosible combinations of the two dice. The seruences were somewhat fewer than they ought to have been, and the relative fregumey of the different combinations very different indered from what theory would predict. \({ }^{1}\) The explanation of this is casiic- found ; for the records of the relative frequency , if cads face show that the dice must have been very irrezular, the sis face of the winte die. for example, falling 38 per cent more often than the four face of the same die. This, them, is the soln comeluaion of these immensely laborious experi-ments,-that Wolf's dice were very ill made. Indeed the experimems omuld have had no bearing except upon the accuracy of his dice. But ten sears later Wolf embarked upon one more series of experiments, using four distinguishable dice,-white, yellow, real, and blue. -and tosising this set of four 10,000 times. Winlf remond altomether, therefore, in the course of his life
 Wold had any well-defined whect in view in making these reconds, which are published in curions conjunction with various astronomical results, and they afford a wonderful example of the pure love of experiment and observation. \({ }^{2}\)
19. Another smins of calculations have been based upon the rady-made dula pmovided by the published results of lotteries and roulette. \({ }^{3}\)

\footnotetext{
\({ }^{1}\) Cauber quotes the principal results (loc. cil. vol. i. pp. 149-15l). The irequencies of only \(f\), instead of 18 , out of the 36 combinations lay within the probable limits, and the standard deviat ion was \(76 \cdot 8\) instead of \(23 \cdot 2\).
\({ }^{2}\) The latest experiment of the kind, of which I am aware, is that of Otto
 i43-15(j), who recorded 24 series of 180 throws each with four distinguishable tice.
 demand withe part of gamblers. An Almanach romain sur la loterie royale de France was published at Paris in 1 \(\$ 30\), which contained all the drawings of the French lothery itwo or thrce a month from 1758 to 1830 . Playors at Monte Carlo aro proviled with cards and pins with which to record the results of successive coupls, and tho results at the tables are recrularly published in Le Monum. Gitmblers study these returns on account of the bolief, which they nsually hold, that as the mumber of cases is increased the absolute deviation from the mast probathe promation becomes less, whereas at the best Bernoulli's
}

Czuber \({ }^{1}\) has made calculations based on the lotteries of Prague (2954 drawings) and Brün (2703) drawing: between the years 17.54 and 1886 , in which the actual results agree very well with theoretical predictions. Fechner \({ }^{2}\) mployed the lists of the ten State lotteries of Saxony between the years 1843 and 185\%. Of a rather more interesting character are Professor Karl Pearson's investigations \({ }^{3}\) into the results of Monte Carlo Roulette as recorded in Le Monaco in the course of cight weeks. Applying Bernoulli's Theorem, on the hypothesis of the equiprobability of all the compartments throughout the investigation, he found that the actuall! recorded proportions of red and black were not unexpected, but that alternations and long runs were so much in excess, that, on the assumption of the exact accuracy of the tables, the i prioni odds were at least a thonisand millions to one against some of the recorded deviations. Professor Pearson concluded, therefore, that Monte Carlo Roulette is not objectively a game of chance in the sense that the tables on which it is played are absolutely devoid of bias. Here also, as in the case of ITolf's dice, the conclusion is solely relevant, not to the theory or philosophy of Chance, but to the material shapes of the tools of the experiment.

Professor Pearson's investigations into Roulette, which dealt with \(33,0(0)\) Monte Carlo coups, have been overshadowed, just

Theorem shows that the proportionate deviation decreases while the absolute deviation increases. Cf. Houdin's Les Tricheries des Grecs dévoilées: "In a game of chance, the oftener the same combination has occurred in succession, the nearer wo are to the certainty that it will not recur at the next cast or turn-up. This is the most elementary of the theories on probabilities; it is termed the maturity of the chances." Laplace (E.ssai philosophique, p. 142) quotes an amusing instance of the same belief not drawn from the annals of gambling: "J'ai vu des hommes désirant ardemment d'avoir un fils, n'apprendre qu'avec peine les naissances des garçons dans le mois où ils allaient devenir pères. S'imaginant que le rapport de ces naissances à celles des filles devait être le même à la fin de chaque mois, ils jugaient que les garçons déjà nés rendaient plus probables les naissances jrochaines des filles."

The literature of gambling is very extensive, but, so far as 1 am acquainted with it, excessively lacking in variety, the maturity of the chances and the martingale continually recurring in one form or another. The curious reader will find tolerable accounts of such topics in Proctor's Chance and Luck, and Sir Hiram Naxim's Monte Carlo Fucts and Fallacies.

I Zum Ciewt: der arosern Zahlen. The results are stammarisent in lis If ahrwhrimlichlierts, rflumeng, vol. i. p. 139.
\({ }^{2}\) Kollektivmasslehre, p. 229. These results also are summarised by Czuber, loi: , \(\because\)
\({ }^{3}\) The Chances of Death, vol. i.
as all other tosses of coins and dice have been outdome by Wolf, by Dr. Karl Marbe, \({ }^{1}\) who has examined 80,000 coups from Monte Carlo and elsewhere. Dr. Marbe arrived at exactly opposite conclusions; for he claims to have shown that long runs, so far from being in excess, were greatly in defect. Dr. Marbe introduces this experimental result in support of his thesis that the world is so constituted that long runs do not as a matter of fact occur in it. \({ }^{2}\) Not merely are long runs very improbable. They do mot, aceording to him, oecur at all. But we may doubt whether roulette can tell us very much either of the laws of logic or of the constitution of the universe.
1)r. Marte's main thesis is identical, as he himself recognises, with one of the heterodox contentions of D'Alembert. \({ }^{3}\) But this principhe of variety, precisely opposite to the usual principle of Indurtion. can have no claim to be accepted à priori and, as a gemerul primwiple, there is no adequate evidence to support it from experimee. Its oricin is to be found, perhaps, in the fact that

\({ }^{2}\) Dr. Marbe's monograph has given rise in Germany to a good deal of discussion, not directed towards showing what a preposterous method this is for demonstrating a natural law, but because the experimental result itself does not really fullow from the duta and is due to a somewhat subtle crror in Marle's reasoning, by which he has been led into an incorrect calculation of the probable proportions a priori of the various sequences. The problem is discussed by
 to these see the Biblingraphy), and by Lexis (Alhandlungen, pp. 222-226) and Czuber (Wahrscheinlichkeitsrechnung, vol. i. pp. 144-149). Largely as a result of this controversy, Von Bortkiewicz has lately devoted a complete treatise (I)ie Iterationen) to the mathematics of 'runs.' Dr. Marbe has been given far more attention by his colleagues in Germany than he conceivably deserves.
\({ }^{3}\) I'Alembert's principal contributions to Probability are most accessible in the volumes of his Opuscules mathémutiques (1761). Works on Probability usually contain some reference to D'Alombert, but his sceptical opinions, rejected rather than answered by the orthodox school of Laplace, have not always received full justice. I'Alembert has three main contentions to which in his varions paprers he constantly recurs:
(1) Thin a prombility cory small mathemationty is reatly zom:
(2) That the probabilities of two successive throws with a die are not 10.4.7. Mhen:
(3) 'That 'mathematical expectation' is not properly mensured by the product of the probability and the prize.

The first and third of these were partly advanced in explanation of the Peterahurg paradox (see p). 316). Tho second is connected with the first, and was also usecl to support his incorrect evaluation of the probability of heads twice rumnine ; but I'Alembert, in spito of many of his results being wrong, does not altogether deserve the ridicule which he has suffered at the hands of writers, who accepted without sceptical doubts the hardly less incorrect condusions of the orthendox theory of that time.
in a certain class of cases, especially where conscious human agency comes in, it may contain some element of truth. The fact of an act's having been done in a particular way once may be a special reason for thinking that it will not be performed on the next occasion in precisely the same manner. Thints in many so-called random events some slight degree of causal and material dependence between successive occurrences may, nevertheless, exist. In these cases 'runs' may be fewer and shorter than those which we should predict, if a complete absence of such dependence is assumed. If, for example, a pack of cards be dealt, collected, and shuffled, to the extent that card-players do as a rule shuffle, there may be a greater presumption against the second hand's being identical with the first than against any other particular distribution. In the case of croupiers long experience might possibly suggest some psychological generalisation,--that they are very mechanical, giving an excess of numbers belonging to a particular section of the wheel, or, on the other hand, that when a croupier sees a run beginning, he tends to vary his spin more than usual, thus bringing runs to an end sooner than he ought. \({ }^{1}\) At any rate, it is worth emphasising once more that from such experiments as these this is the only kind of knowledge which we can hope to obtain,--knowledge of the material construction of a die or of the psychology of a croupier.
\({ }^{1}\) A good roulette table is, however, so delicate an instrument that no prohable degree of regularity of habit on the part of the spinner could be sufficieni to produce regularity in the result.

\section*{CHAPTER XXX}

\author{
ThE MATHEMATICAL USE OF STATISTICAL FREQUENCIES FOH THE DETERMINATION OF PROBABILITY i POSTERIORI-THE METHODS OF LAPLACE
}

\begin{abstract}
Utilissima est aestimatio probabilitatum, quanquam in exemplis juridicis politicisque plerumque non tam subtili calculo opus est, quam accurata

\end{abstract}
1. Is the preceding chapter we have assumed that the probability of an event at each of a series of trials is given, and have considered how to infer from this the probabilities of the various possible frequmaies of the event over the whole series, without discussing in detail by what method the initial probability had been determined. Instatistical inquiries it is generally the case that this initial probability is based, not upon the Principle of Indifference, but upon the statistical frequencies of similar events whith have been observed previously: In this chapter, therefore, we must commence the complementary part of our inquiry,namely, into the method of deriving a measure of probability from an observed statistical frequency.

I do not myself believe that there is any direet and simple mothod by which we can make the transition from an observed numorical frequency to a numerical measure of probability. The problem, as I viow it, is part of the wereral prohlem of foundine judements of probability upon experience, and can only be dralt with by the general methods of induction expounded in Part III. The nature of the problem precludes any other method, and direct mathematical devices can all be shown to depend upon insupportable assumptions. In the next chapters we will consider the applicability of general inductive methods to this problem, and in this we will endeavour to discredit the mathematical charlatanry by which, for a humdred years pant, the basis of theoretical statistics has been greatly undermined.
2. Two direct methods have been commonly employed, theoretically inconsistent with one another, though not in every case noticeably discrepant in practice. The first and simplest of these may be termed the Inversion of Bernoulli's Theorem, and the other Laplace's Rule of Succession.

The earliest discussion of this problem is to be found in the Correspondence of Leibniz and Jac. Bernoulli, \({ }^{1}\) and its true nature cannot be better indicated than by some account of the manner in which it presented itself to these very illustrious philosophers. The problem is tentatively proposed by Bernoulli in a letter addressed to Leibniz in the year 1703. We can determine from à priori considerations, he points out, by how much it is more probable that we shall throw 7 rather than 8 with two dice, but we cannot determine by such means the probability that a young man of twenty will outlive an old man of sixty. Yet is it not possible that we might obtain this knowledge a posteriori from the observation of a great number of similar couples, each consisting of an old man and a young man ? Suppose that the young man was the survivor in 1000 cases and the old man in 500 cases, might we not conclude that the young man is twice as likely as the old man to be the survivor? For the most ignorant persons seem to reason in this way by a sort of natural instinct, and feel that the risk of error is diminished as the number of observations is increased. Might not the solution tend asymptotically to some determinate degree of probability with the increase of observations? Nescio, Vir Amplissime, an speculationibus istis sulditatis aliquid inesse Tibi videatur.

Leibniz's reply goes to the root of the difficulty. The calculation of probabilities is of the utmost value, he says, but in statistical inquiries there is need not so much of mathematical subtlety as of a precise statement of all the circumstances. The possible contingencies are too numerous to be covered by a finite number of experiments, and exact calculation is, therefore, out of the question. Although nature has her habits, due to the recurrence of causes, they are general, not invariable. Yet empirical calculation, although it is inexact, may be adequate in affairs of practice. \({ }^{2}\)

\footnotetext{
1 For the exact references see Bibliography.
\({ }^{2}\) Leibniz's actual expressions (in a letter to Bernoulli, 1)ecember 3, 1703) are as follows: Utilissima est aestimatio probabilitatum, quanquam in exemplis juridicis politicisque plerumque non tam subtili calculo opus est, quam accurata omnium circumstantiarum enumeratione. Cum empirice aestimamus proba-
}

Bernoulli in his amswer foll back upm the amatogy of halls drawn from an urn, and mantaned that without estimating each separate contingency we might determine within narrow limits the proportion favourine each alternative. If the true propertion were \(2: 1\), we might wtimate it with moral certaints. it postrimit as lying between 201: 1001 and 199: 1000 . " (ertus *um." he concluded the controwersy, " Tibi placituram demonst rationmm. "um publicavero." But whether he was impressed by the just caution of Leibniz. or whether death intercepted him, he ariances matters no further in the Ars Conjectandi. Aiter dealing with some of Leibniz's objections \({ }^{1}\) and seeming to promise some monde of estimatimp pohabilities it posteriori by an inversion of his theorem, he proves the diee theorem only and the book is suddenly at an end.
3. In dealing with the correspondence of Leibniz and Bermoulli, I have menthan matuly intlumend by the hiotorical intereat of it. The view of Leibniz, dwelling mainly on considerations of analory, and demanding " not so much mathematical subtlets. as a precion statment of all the circumstances." is, suhstamially, the viow which will ber supperted in the followinge chapters. The desire of Bermuili for an exact formula, which would derive from the numerical frequency of the experimental results a numerical measure of their probability, preludes the exact formulas of later and less cautious mathematicians, which will be examined immediately.
4. During the greater part of the eightemth century there is me trace. I think, of the explicit use of the Insersion of Bernoullis Themem. The insestigations carriod out ha 1) Wembert, 1) andel Bermulli, and others whed upen the typu of argument examined in Chapter XXV. They showed, that is to say, that certain whared serises of cremts would have heen very imperobable, if we had supposed independence between some two factors or if

\footnotetext{










}
some occurrence had been assumed to be as likely as not, and they inferred from this that there was in fact a measure of dependence or that the occurrence had probability in its favour. But they did not endeavour to pass from the observed frequency of occurrence to an exact measure of the probability. With the advent of Laplace more ambitious methods took the field.

Laplace began by assuming without proof a direct inversion of Bernoulli's Theorem. Bernoulli's Theorem, in the form in which Laplace proved it, states that, if \(p\) is the probability \(\grave{a}\) priori, there is a probability P that the proportion of times \(\begin{gathered}m \\ m+n\end{gathered}\) of the event's occurrence in \(\mu(=m+n)\) trials will lie between \(p \pm \gamma / \int_{\mu}^{2 p q}\), where \(\mathrm{P}=\stackrel{2}{\Omega} \int_{0}^{\gamma} \int_{0}^{\prime \prime d} d+\frac{1}{\sqrt{2} \pi \mu \nu, q^{\gamma}} e^{\gamma \prime}\). The inversion of the theorem, which he assumes without proof, states that, if the event is observed to happen \(m\) times in \(\mu\) trials, there is a probability P that the probability of the event \(p\) will lie between \({ }_{\mu}^{m} \pm \gamma / \frac{2 m n}{\mu^{3}}\), where \(\mathrm{P}=\left.\int_{\pi}^{2}\right|^{2} e^{-t^{2}} d t+\frac{1}{\sqrt{2 \pi \mu_{\mu}} \mu^{2}} e^{-\gamma^{2}}\). The same result is also given
by Poisson. \({ }^{1}\) Thus, given the frequency of occurrence in \(\mu\) trials, these writers infer the probability of occurrence at subsequent trials within certain limits, just as, given the i pmon fookability, Bernoulli's Theorem would enable them to predict the frequency of occurrence in \(\mu\) trials within corresponding limits.

\footnotetext{
\({ }^{1}\) For an account of the treatments of this topic both by Laplace and by Poisson, see 'Todhunter's IIistory, pp. 554-557. Both of them also obtain a formula slightly different from that given above by a method analogous to the first part of the proof of Laplace's Rule of Succession ; i.e. by an application of the inverse principle of probability to the assumption that the probability of the probability's lying within any interval is proportional to the length of the interval. This discrepancy has given rise to some discussion. See Todhunter,
 Inversion of Bernoulli's Theorem in I'robabilities: and Czuber, Entwicklung, Tp. 83, 84. But this is not the important distinction between the two mathematical methods by which this question has been approached, and this minor point, which is of historical interest mainly, I forbear to enter into.
}

If the number of trials is at all numerous, these limits are narmon and the parport of the inversion of Bernoulli's Theorem may therefore be put briwly as follows. By the direct theorem, if \(p\) monatires the probability, \(p\) also measures the most probable value of the frequency ; by the inversion of the theorem, if \(m\)
\[
m+n
\]
measures the frequency, \(m\) also measures the most probable value of the probability. The simplicity of the process has recommended it, since the time of Laplace, to a great number of writers. Czuber's argument, criticised on p. 351, with reference to the proportions of male and female births in Austria, is based upon an ung talifiod me of it. But examples abomed thomathont the litamare of the sinject, i , which the theorem is emplowed in circumstances of greater or less validity.

The theorem was originally given without proof, and is indeed incapable of it, unless some illegitimate assumption has been introduced. Rut, apart from this, there are some obvious objections. We have seen in the preceding chapter that Bernoulli's Theorem itself camnot be applied to all kinds of data indiscriminately, but only when certain rather stringent conditions are fulfilled. Corresponding conditions are required equally for the inversion of the theorem, and it cannot possibly be inferred from a statement of the number of trials and the frequency of occurrence merely, that these have been satisfied. We must know, for instance, that the examined instances are similar in the main relevant particulars, both to one another and to the unexamined instances to which we intend our conclusion to be applicable. An unanalysed statement of frequency cannot tell us this.

This method of passing from statistical frequencies to prob)abilities is not, however, like the method to be discussed in a moment, radically false. With due qualifications it has its place in the solution of this problem. The conditions in which an inversion of Bernoulli's Theorem is lecritimate will be clucidated in Chapter XXXI. In the meantime we will pass on to Laplace's second method, which is more powerful than the first and has obtained a wider currency. The more extreme applications of it are no loner ventured upon, but the theory which underlies it is still widely adopted, especially be French writers upon

5. The formula in question, which Vem \({ }^{1}\) has called the Pule of Succession, declares that, if we know no more than that an event has occurred \(m\) times and failed \(u\) times under given conditions, then the probability of its occurrence when those conditions are next fulfilled is \(\begin{gathered}m+1 \\ m+n+2\end{gathered}\). It is necessary, however, before we examine the proof of this formula, to disens in detail the reasoning which leads up to it.

This preliminary reasoning involves the Laplacian theory of 'unknown probabilities.' The postulate, upon which it depends, is introduced to supplement the Principle of Indifference, and is in fact the extension of this principle from the probabilities; of arguments, when we know nothing about the arguments, to the probabilities that the probabilities of arguments have certain values, when we know nothing about the probabiiities. Laplace's enunciation is as follows: " (Quand la probabilité d'un événement simple cist inconnue, on prut lui supposer également toutes les valeurs depuis zéro jusqu’à l'unité. La probabilité de chacune de ces hypothèses tirée de l'événement observé est . . . une fraction dont le numérateur est la probabilité de !'événement dans cette hypothèse, et dont le dénominateur est la somme des probabilités semblables relatives à toutes les hypothèses. . . ." 2

Thus when the probability of ain event is unknown, we may suppose all possible values of the probability between 0 and 1 to be equally likely à priori. The probability, after the event has occurred, that the probability a priori was \({ }_{r}^{1}\) (say), is meatured by a fraction of which \({ }_{r}^{1}\) is the numerator and the sumis of all the possible \(\dot{a}\) priuri values the denominator. The origis of this rule is evident. If we consider the problem it: which a ball is drawn from a bas containing an infinite number of black and white balls in unknown proportions, we have hypotheses, corresponding to each of the possible constitutions of the bag, the assumption of which yidds in turn every value between 0 and 1 as the \(i\) imiori probability of drawing a white ball. If we could asstame that these constitutions are equally probable is prori, we should obtain probabilities for each of them it powtrioni according to Laplace's rule.

\footnotetext{
i longir a) (liture, 1) l!oi).

}

On the analogy of this haphace assumes in momeral that, where averything is unknown, we mar suppose an infinito number of possibilities. vach of which is equally likely, and each of which heads to the event in quest ion with a diff.rent deemee of probahility, a) that fope very value hetween 1 and I there is one and only one hypothetical constitution of thines, the as-umption of which invests the event with a probability of that value.
6. It might be an almost sufficient criticism of the above to wint out that these assumptions are entirely haseless. But the theory has taken so important a place in the development of probability that it deserves a detailed treatment.

What, in the first place, does Laplace mean by an unknoun probability? He dons mot mean a probability, whose value is in fact unknown to \(u s\), because we are unable to draw conclusions which comid be drum from the dutu: and he seems to apply the twou to any pormbility whor value, actorting to the argument of Chapter III., in mumpratly indowerminte. Thus he assumes that merry prombility has a numerical value amel that, in those cases where there seems to be no numerical value, this value is not non-existent but unknown ; and he proceeds to argue that whow the mumerical value is unknown, or as \(\boldsymbol{I}\) should say wher there is no such value, every value between 0 and 1 is equally probable. With the possible interpretations of the term 'unknewn probatility. :ont with the theory that exery protratility can be measured by one of the real numbers between 0 and 1 , I hase dealt, as carefully is I cati, in 'hapter III. If the view batan there is correct. Laplace's theme heals down immediately. But even if we were to answer these questions, not as they have been answered in Chapter III., but in a manner favourable to Laphanes thenes it remains donhtral whether we conld lewitimately attribute a value to the probability of an unknown probability's having such and such a value. If a probability is unknown, surely the probability, relative to the same evidence, that this probability has a given value, is also unknown ; and we are involved in an infinite regress.
7. This point leads on to the second objection; Laplace's theory requires the employment of both of two inconsistent methods. Let us consider a number of alternatives \(a_{1}, a_{2}\), etc., having probabilities \(p_{1}, p_{2}\), etc.; if we do not know anything about \(n_{1}\), we do not know the value of its probahility \(p_{1}\), and we
must consider the various possible values of \(p_{1}\), namely \(b_{1}\), \(b_{2}\), ete. the probabilitios of these poisible values loning \(q_{1}, q_{2}\), etc. respectively. There is no reason why this process should ever stop. For as we do not know anything about \(b_{1}\), we do not know the value of its probability \(!_{1}\), and we must consider the various possible values of \(q_{1}\). mamely \(c_{1}, c_{2}\), etc., the probabilities of these possible values being \(r_{1}, r_{2}\), etc. respectively ; and so on. This method consists in supposing that, when vec do not know anything about an alternative, we must consider all the possible values of the probability of the alternative ; these possible valus con form in their tum a set of alternetives, and so on. But this method by itself can lead to no final conclusion. Laplace superimposes on it, therefore, bis other method of determining the prohabilities of alternatives abont whic!, we know nothing--nancly, the Principle of indifference. According to this method, when we know nothing about a set of alternatives, we suppose the probabilities of each of them to be equal. In some parts of his writings-and this is true also of most of his followers--he applies this method from the beginning. If, that is to say, we know nothing about \(a_{1}\), since ( 1 , and its contradictory form a pair of exhanstive alternatives two in mumer, the probability of these alternatives is equal and each is \(\frac{1}{2}\). But in the reasoning which leads up to the Taw of Succession he chooses to ayply this mothod at the second stage, having used the other method at the first stage. If, that is to say, we know nothing about \(a_{1}\), its probalility \(p_{1}\) may have any of the values \(b_{1}, b_{2}\), we. where \(b_{1}\) is ans fraction between 0 ant! ! and, as we know nothing about the probabilities \(q_{1}, q_{2}\) e ete wi these altematives \(\left\langle_{3}, b_{2}\right.\), etc.. we may by the Principle of inditherence suppose then to be equal. This account may seem rather confused; but it is not easy to give a lucid account of so confused a doctrine.
8. Turning asid from the considerations, let us examine the theory, for a moment, from another ide. When we reach the Rule of Succeesion, it will be seem that the hypothetical ie priori probabilities are tranted as if they were possible couses of the event. It is assumed, that in to say, that the mumber of possible sets of antecedent comditions is proportional to the number of real numbers hetween 0 and 1 ; and that these fall info eymal grouns, each group (ormespoading to one of the real himber
hetween 1 and 1, this number measuring the degree of probability with which we could predict the ewont, if we knew that an antecedent comdition belomging to that group was fultilled. It is then assumed that all of theor posible antecondent conditions are is pive equally likels. The argument has arisen be false analoge from the problem in which a hall is drawn from an urn contaning an iafinite number of black ame white halls. But for the assumption that we have in general the kind of knowledge which is necomary about the pro-ible anteredents. no reasomable foundation has been suggested.
le Morgan endeavoured to deal with the difficulty in much the same way in the following passage : \({ }^{1}\) "In determining the thatce which wist (moder kitow eircumstances) for the happening of an event a number of times which lies between certain limits, we are involved in a consideration of some difficulty, namely, the probabitity of a probabitity, or, as we have called it, the presumption of a probability. To make this idea more clear, remember that any state of probability may be immediately mand the expres inst of the rentt of a set of circumstances, which being introduced into the question, the difficulty disappears. The word presumption refers distinctly to an act of the mind, or a state of the mind, while in the word probability we feel disposed rather th think of the external arranmments on the knowledge of which the strengith of our presumption ought to depend, than of the presumption itself." The print of this explanation lies in the a- umption that " ans state of probability may be imme-
 It cannot be allowed that this is generally true ; \({ }^{2}\) and even in those cases in which it is true we are thrown loack on the a priori probabilities of the various sets of circumstances which need not

9. The proof of the Rule of Succession, which is based upon this theory of unknown probabilities, is, briefly, as follows:

If \(x\) stands for the à priori probability of an event in given conditions, then the probability that the event will occur \(m\) fimns and fail \(n\) times in these conditions is \(x(1-x) "\). If, however, \(r\) is unknown, all values of it between 0 and

\footnotetext{

 an urn containing hack and white in unknown proportions, unlese the number 1.f inil: imfonio
}

I are ì priori equally probable. It follows from these two sets of considerations that, if the event has been observed to occur \(m\) times out of \(m+n\), the probability a posteriori that \(x\) lies between \(x\) and \(x+d x\) is proportional to \(x^{\prime \prime \prime}(1-x)^{n} d x\), and is equal, therefore, to \(\Lambda x^{\prime \prime \prime}(1-x)^{\prime \prime} d x\) where A is a constant. Since the event has in fact occurred, and since \(x\) must have one of its possible values, A is determined by the equation
\[
\int_{-1}^{1} A r^{\prime \prime \prime}(1-r)^{\prime \prime} d_{1}=1 \quad \therefore A=\frac{\Gamma^{\prime}(m+n+2)}{\Gamma^{\prime}(m+1) \Gamma^{\prime}(n+1)} .
\]

Hence the probability that the event will occur at the \((m+n+1)\) th trial, when we know that it has occurred \(m\) times in \(m+n\) trials, is
\[
\left.\mathrm{A}\right|_{n} ^{1}
\]

If we substitute the value of A found above, this is equal to \(m+1 \quad 1\)
\(m+n+2\)
The class of problem to which the theorem is supposed to apply is the following: There are certain conditions such that we are ignorant \(i\) priori as to whether they do or do not lead to the occurrence of a particular event ; on \(m\) out of \(m+n\) occasions, however, on which these conditions have been observed, the event has occurred; what is the probability in the light of this experience that the event will occur on the next occasion? The answer to all such problems is \(\begin{gathered}m+1 \\ m+n+2\end{gathered}\). In the cases where \(n=0\), i.e. when the event has invariably occurred, the formula

\footnotetext{
\({ }^{1}\) The theorem is sometimes enunciated by contemporary writers in a much more guarded form, e.g. by Czuber, Wahrscheinlichkeitsrechnung, vel. i. p. 197, and by Bachelier, Calcul des probabilités, p. 487. Bachelier, instead of assuming that the a priori probabilities of all possible values of the probability of the event are cqual, writes \(\hat{\omega}(y) d y\) as the \(\grave{a}\) priori probability that the probability is \(y\), so that after \(m\) occurrences is \(m+n\) trials the probability that the probability lies between \(y\) and \(y+d y\) is \(\begin{aligned} & y^{\prime \prime \prime}(\mathbf{I}-y)^{n} \hat{\omega}(y) d y \text {. If one has no idea of } \hat{\omega} \grave{a} \text {. } y^{\prime \prime \prime}(1-y)^{\prime \prime} \hat{\omega}(y) d y .\end{aligned}\). priori, he suggests that the simplest hypothesis is to put \(\hat{\omega}=1\), which leads, as above, to Laplace's Law of Succession. He also proposes the hypothesis \(\hat{\omega}(y)=a+a_{1} y+a_{2} y^{2}+\ldots\), in which case the denominator is a series of Eulerian integrals. There is a discussion of the Law of Succession, and of the contradictions and paradoxes to which it leads, by E. T. Whittaker and others in Part VI. vol. viii. (1920 of the Transactions of the Faculty of Acluaries in sootloul.
}
yields the result \(\begin{aligned} & m+1 \\ & m: 2\end{aligned}\) In the case where the conditions have been observed once only and the event has occurred on that occasion, the result is \(\stackrel{\because}{g}\). If the conditions have never been met with at all, the probability of the event is \(\begin{aligned} & 1 \\ & 2\end{aligned}\). And even in the case where on the only occasion on which the conditions were observed, the event did not occur, the probability is \(\frac{1}{?}\).
some of the flaws in this proof hate been already explaned. One minor objection may be pointed out in addition. It is assumed that if a is the it priori probathility of the event's happeninge once. theng is the it priori probability of its happening \(n\) times in sucersiom, whereas be the theorem's own showing the knowleder that the erent has happened onee modifies the probability of its hafpeminer a second time: its suceessive occurrences are but. therefore, inderembent. If the a priori probability of the "vent is \(\begin{aligned} & 1 \\ & .2\end{aligned}\), and if, after it has been observed once, the probability that it will oceur a second time is \(\frac{2}{3}\), then it follows that the \(i\) prinri probability of its occurring twice is not \({\underset{2}{1}}_{1}^{1}{ }_{2}^{1}\), but \({\underset{2}{1}}_{1}^{2} \times \frac{2}{3}\), i.e. \(\frac{1}{3}\) : and in qeneral the is primi prohability of it: happenimer \(n\) times in succession is not \(\left(\frac{1}{2}\right)\) but \({ }^{\prime \prime}: 1^{1}\).
10. But refmements of disproof are hardly needed. The principle ens anclusion is inconsistent with its premisises. We hersin with the assumption that the it primit pobability of an event. abont which we have no information and no experience, is unIfewor, and that all values between ") and 1 are equally probable. Whe end with the conclusion that the it priori probathlity of such an event is \(\frac{1}{2}\). It has been pointed out in \(\S 7\) that this contratiction Was latent, as som as the P'rinciple of Indifterence


The theorem's conclusions, moreover, are a reductio ad "thsurflum of the reasming upon which it is hased. Who conld suppere that the prohabilite of a purely hypothetical exont. of
whatever complexity, in fievour of which no positim argument exists, the like of which has never been observed, and which has failed to occur on the one occasion on which the hypothetical conditions were fulfillecl, is no less than \({ }_{3}^{1}\) ? Or if we do suppose it, we are involved in contradictions, -for it is easy to imagine more than three incomputible events which satisfy these conditions.
11. The theorem wes first suggested by the problem of the urn which contains black and white balls in unknown preportions: \(m\) white and \(n\) black balls have been successively drawn and replaced; what is the probability that the next draw will vield a white ball ? It is supposed that all compowitions of the urn are equally probable, and the proof then proceeds precisely as in the case of the more general rule of succession. The rule of succession has bem, sometimes, directly deduced from the case of the urn, by assimilating the occurmen of the erent to the duaving of a white ball and its nom-ocearrence to the drawing of a black ball.

On the hypothesis that all compositions of the um are equally probable, an hypothesis to which in general there is nothing corresponding, and on the further hypothesis that the mumber of balls is infinite, this solution is correct. \({ }^{1}\) But the rule of succession does not apply, as it is easy to demonstrate, even to the case of balls drawn from an urn, if the number of balls is finite. \({ }^{2}\)
12. If the Jule of Succession is to be adopted by adherents of the Frequency Theory of Probability, \({ }^{3}\) it is necessary that they should make sone modification in the preliminary reasoning on which it is based. By Dr. Venn, however, the rule has been

\footnotetext{
1 This second condition is often omitted (e.g. Bertrand, Calcul des probabilités, p. 172).
\({ }^{2}\) The correct solution for the case of a finite number of balls, on the hypothesis that each possible ratio is equally likely, is as follows: The probability of a black ball at a further trial, after black balls have been successively with-
 the sum of the \(r\) th powers of the first \(n\) natural numbers. This reduces to
 \(p\) black balls and \(q\) white balls have been drawn and replaced, the chance
\[
=r^{-1}\left(\begin{array}{ll}
r & r
\end{array}\right)
\]
that the next ball will be black is,
\[
\therefore r\left(\begin{array}{ll}
n & r
\end{array}\right)
\]
}

See (:1apin: 11!
-xplicitle rejected one the around that it does mot accord with exprimene. \({ }^{1}\) But Professor Karl Pearson, who accepts it, has made the necessary restatement, \({ }^{2}\) and it will be worth while to examine the reasoning when it is put in this form. Professor Pearson's proof of the Rule of Succession is as follows:
" I start, as most mathematical writers have done, with ' the "., un! distribution of 'morance.' or I assume the truth of Base'
 I thin with Edeosomt \({ }^{3}\) that the hypothesis of the equal distrilution of immune is. vi thin the limits of practical life, justified by our experience of statistical ratios, which ad prior are unknown, ie. such ratios do not tend to cluster markedly round any particular value. 'Chances' lie between 0 and 1 , but our experience does not indicate any tendency of actual chances to cluster round any particular value in this range. The ultimate basis of the theory of statistics is thus not mathematical but observational. Those who do not accept the hypothesis of the
 are compelled to produce definite evidence of the clustering of chances, or to drop, all application of past experience to the judgemont of probable future statistical ratios. . . .
"Let the chance of a given event occurring be supposed to lie between \(x\) and \(x+d x\), then if on \(n=p+q\) trials an event has been observed to occur \(p\) times and fail \(q\) times, the probability that the true chance lies between \(x\) and \(x+d x\) is, on the equal distribution of our ignorance.
\[
\left.P=\begin{array}{ll}
1 & 1
\end{array}\right)
\]
:" This is Bayes' Theorem. . . . \({ }^{4}\)

\footnotetext{





 constant about which we know nothing in particular is as likely to have one value as another is grounded upon the rough but solid experience that such constants do, as a matter of fact, as often have one value as another." See also (chapter til. s, il as o

4 I'rufessor Pearson's use of this title for the above formula is not, I think.
 it self, and not this extension of it.
}
"Now suppose that a second trial of \(m=r+s\) instances be made, then the probability that the given event will occur \(r\) times and fail \(s\), is on the \(\grave{i}\) priori chance being between \(x\) and \(x+d x\)
\[
=P_{1} \Gamma_{\prime \prime \prime}^{\prime \prime} \Gamma_{s}^{\prime \prime}\left(\begin{array}{ll}
1 & \quad,
\end{array}\right)^{s},
\]
and accordingly the fotal chance \(\mathrm{C}_{\text {, }}\), whatever \(x\) may be of the event occurring \(r\) times in the second series, is

This is, with a slight correction, Laplace's extension of Bayes' Theorem." \({ }^{1}\)
13. This argument can be restated as follows. Of all the objects which satisfy \(\phi(x)\), let us suppose that a proportion \(p\) also satisfy \(f(x)\). In this case \(p\) measures the probability that any object, of which we know only that it is \(\phi\), is in fact also \(f\). Now if we do not know the value of \(p\) and have no relevant information which bears upon it, we can assume ì priori that all values of \(p\) between 0 and \(I\) are equally likely. This assumption, which is termed the 'equal distribution of ignorance,' is justified by our experience of statistical ration. Our experience, that is to say, leads us to suppose that of all the theories, which could be propounded, there are just as many which are ahways true as there are which are always false, just as many which an true once in fifty times as there are which are true once in three times, and so on. Professor Pearson challenges those who do not accept this assumption to produce definite evidence to the contrary.

The challenge is casily met. It would not be difficult to produce 10,1\()\) (f) ponitive theories which are always false corresponding to every one which is always true and 10,000 coredations of posi-

\footnotetext{
1 The rest of the article is concerned with the determination of the probable crror when Laplace's Rule of Succession is used not simply to yicld the probability of a single additional occurrence, but to predict the probable limits within which the frequency will lie in a considerable series of additional trials. Professor Pearson's method applies more rigorous methorls of approximation to the fundamental formulae given above than have been sonctimes used. As my main purpose in this chapter is to dispute the general validity of the fundamental formulae, it is not worth while to consider these further developments here. If the validity of the fundamental formula were to be eranted, Professor Pearson's methods of approximation would, I think, be satisfactorv.
}
tive qualitins which hold hes often than once in three times for wery one we can name which halds meme often than once in three times. And the converse is the case for negative theories and corredions between nowative quatites: for corresponding to every positive theory which is true there is a negative theory which is false. and sw on. Thus wherience, if it shows any thing, shows what there is a rery marhed chastering of statistical ratios in the inghourhome of zero and unity,-of those for positive theorm and for contelations between positive qualities in the miphtourhome of zero. and of thuse for werative theories and for convations between bugrative qualitios in the neighberorhood of unity. Moreover, we are seldom in so complete a state of ignorance regarding the nature of the theory or correlation under insertigation az mot to know whether or not it i.s a positise theory or a convedion butween poxitix" qualitios. In general, therefore, whenever our investigation is a practical one, experience, if it tali: wes any hinge inll- us mot omly that the statistical ratios cluster in the neighbourhood of zero and unity, but in which of these two neighmorhouds the ratio in this particular case is most likely a priori to be found. If we seek to discover what proportion of the population suffer from a certain disease, or have red hair, or are callint Jom? . it is preperstrous to suppese that the propertion is as lihelf it priori to excemi as to fall short of (say) fifty per cent. As Profman Peareon applies this met hod to investigations where it is phain that the qualifes involved are positive, he secms to maintain that experience shows that there are as many positive attributes which are shared hy more than half of any perpulation as there are which are shared by less than half.

It is al worth while to feint out that it is formally impossible that it -henat he true of all characters. simple and complex, that they an a linely fol lave any one irequency as any other. For let us take a character \(e\) which is compumui of two characters a and 1, hetwem in hinh there is no asonciation, and let us suppose that 4. haw a farpu-ney of in the ferpulation in question and that \(b\) has a frequency \(y\), so that, in the absence of association, the frequency \(z\) of \(c\) is equal to \(x y\). Then it is easy to show that, if all values of \(x\) and \(y\) between () and 1 are equally probable, all values of \(z\) between 0 and 1 are not equally probable. For the value is more probable than any other, and the pussible values of
\(z\) become increasingly improbable as they difier more widely from \({ }_{2}^{1}\).

It may be added that the conclusions, which Professor Pearson himself derives from this method, provide a reductio ad absurdum of the arguments upon which they rest. IIe considers, for example, the following problem : A sample of 100 of a population shows 10 per cent alfected with a certain disease. What percentage may be reasonably expected in a second :ample of 100 ? By approximation he reaches the conclusion that the percentage of the character in the second sample is as likely to fall inside as outside the limits, \(7 \cdot 85\) and \(13 \cdot 71\). Apart from the preceding criticisms of the reasoning upori which this depends, it does not seem reasonable upon general grounds thet we should be able on so little evidence to reach so certain a conclusioti. The argument does not require, for example, that we have any knowledge of the manner in which the samples are chosen, of the positive and negative analogics between the individuals, or indeed anything at all beyond what is given in the above statement. The method is, in fact, mucu too powerful. It investo any poritive conclusion, which it is employed to support, with far too high a degree of probability. Indeed this is so foolish a thensem that to entertain it is discreditable.
14. The Rule of Succession has played a very important part in the development of the theory of probability. It is true that it has been rejected by Boole \({ }^{1}\) on the ground that the hypotheses on which it is based are arbitays, by Vemn on the ground that it does not accord with experience, by Bertrand \({ }^{2}\) because it is ridiculous, and doultess by others aloo. But it has been wery widely accepted,-by De Morgan, \({ }^{4}\) lig jevons, \({ }^{5}\) by I Anze, \({ }^{5}\) by Czuber, \({ }^{7}\) and by Profesisor Pearison, \({ }^{8}\) - \(t 0\) namessome representative writers of successive schools and periseds. And, in any (ase, it

"Calcul cirs pmintitilitis, 1. 17.
\({ }^{1}\) Article in C'abinet Eincyclopaedia, p. 64. \({ }^{5}\) Principles of Scicnce, p. 297.
 cing " to him " as the more obscure analysis, by which it is usually obtained." The proof is among the worst ever conceived, and may be commended to those who seek instances of the profound credulity of even considerable thinkers.
 and with more qualifications than in the form discussed above.
a Lore cil.
is of interest as being one of the most characteristic results of a way of thimking in probatility introduced by Laplace, and never thoroughly discarded to this dey. Even amengst those writers who have rejected or aroided it, this rejection has been due more to a di- trust of the praticular applications of wheh the law is susceptible than to fundamental oljections acainst almost every step and every prexumtion upon which its proof depends.

Some of these particular applications have certainly been surprising. The law as is evident, provides a numerical measure of the probshility of ary :imple induction, provided only that our ignmance of its conditions is sufficimitly complete amd, althongh. when the number of cases dealt with is small, its results are incratible there is. whea the number dealt with is large, a certain plausibility in the results it gives. But even in these cases paradoxical conclusions are not far out of sight. When laphace prowes that, acoment indige taken of the experience of the human rame the probahility of the - min's rising fe-momew is \(1.82 t, 214\) to 1. this large number may seem in a kind of way to represent our state of mind about the matter. But an ingenious German, Prufo. or Poblal.. \({ }^{1}\) hat pur hout ho aremment a dearee further. and proves her mate of thes anm principles that the prelability of the sun's rising every day for the next 4000 years, is not more, approximately, than two-thirds,-a result less dear to our natural prejudices.


\section*{CHAPTER XXXI}

\section*{THE INVERSION OF BERNOULLI'S THEOREM}
1. I conclude, then, that the application of the mathematical methods, discussed in the preceding chapter, to the general problem of statistical inference is invalid. Our state of knowledge about our material must be positive, not negative, before we can proceed to such definite conclusions as they purport to justify. To apply these methods to material, unanalysed in respect of the circumstances of its origin, and withont reference to our general body of knowledge, merely on the basis of arithmetic and of those of the characteristics of our material with which the methods of descriptive statistics are competent to deal, can only lead to error and to delusion.

But I go further than this in my opposition to them. Not only are they the children of loose thinking, and the parents of charlatanry. Even when they are employed by wise and competent hands, I doubt whether they represent the most fruitful form in which to apply technical and mathematical methods to statistical problems, except in a limited class of special cases. The methods associated with the names of Lexis, Von Bortkiewicz, and Tschuprow (of whom the last named forms a link, to some extent, between the two schools), which will be hriefly described in the next chapter, seem to me to be much more clearly consonant with the principles of sound induction.
2. Nevertheless it is natural to suppose that the fundamental ideas, from which these methods have sprung, are not wholly égarés. It is reasonable to presume that, subject to suitable conditions and qualifications, an inversion of Bernoulli's Theorem must have validity. If we linew that our material could be likened to a game of chance, we might expect to infer chances from frequencies, with the same sort of confidence as that with
which we infer frequencies from chances. This part of our inquiry will not be complete, therefore, until we have endeavoured to elucidate the conditions for the validity of an Inversion of Bernoulli's Theorem.
3. The problem is usually discussed in terms of the happening of an event under certain conditions, that is to sar, of the coexistence of the conditions, as affecting a particular event, with that event. The same problem can be dealt with more generally and more conveniently as an investigation of the correlation hetween two characters \(\mathrm{A}(x)\) and \(\mathrm{B}(r)\), which, as in Part III., are propositional functions which may be said to concur or coexist when they are both true of the same argument \(x\). Given that, within the field of our knowledge, \(B(x)\) is true for a certain proportion of the values of \(x\) for which \(\Lambda(x)\) is true, what is the probahility for a further value a of \(x\) that, if \(\mathrm{A}(a)\) holds, \(\mathrm{B}(a)\) will hold also ?

Let us suppese that the occurrence of an instance of \(\mathrm{A}(x)\) is a sign of one of the events \(e_{1}(x), e_{2}(x) \ldots\) or \(e(x)\), and that these are exhanstive exclusive, and ultimate alternatives. Br exhaustion it is meant that, whenever there is an instance of \(\mathrm{A}(x)\), one of the e's is present: hy corclusive, that the presence of one of the is is mot a sign of the preseme of any other, but not that the concurrence of two ar more of the es is in fact impossible: hy ultimate, that no one of the \(e\) 's is a disjunction of two or more alternatives which might themselves be members of the \(e\) 's. Lut us assumn that these altornations are initially and throughout the argument "qually probable. which, subject to the above conditions, is justifiat by the P'rincipte of !ndifference. We have no reason, that is to say, and no part of our evidence ever gives us che. for thimking that \(A(\) (c) is more likely to he a sign of one of the is than of any wher, or ceren for thin'inge that some e's, although we do not know which, are more likely to occur than others. Let us also assume that, out of \(e_{1}(x), e_{2}(x) \ldots e_{\mu}(x)\), the set \(e_{1}(x), e_{2}(x) \ldots e(x)\), and these only, are signs or occasions of \(B(.1)\) : and further that we have no evidence bearing on the actual magnitude of the interers \(/\) and \(m\), so that the ratio \(/\) 'm is the only factur of which the probability varies as the evidence accumulates. Let us assume, lastly, that our knowlolere of the several instances of \(\mathrm{B}(x)\) is adequate to establish a perfect analogy between them ; the instances \(a\), etc... of \(B(x)\), that is to sas, must
not have anything in common except \(B\), unless we have reason to know that the additional resemblances are immaterial. Even by these considerable simplifications not every difficulty has been avoided. But a development along the usual lines with the assistance of Bernoulli's Theorem is now possible.

Let \(l / m=q\). It the value of \(q\) were known, the problem would be solved. For this numerical ratio would represent the probability that A is, in any random instance, a sign of B ; and no further evidence, which satisfies the conditions of the preceding hypothesis, can possibly modify it. But in the inverse problem \(q\) is not known; and our problem is to determine whether evidence can be forthcoming of such a kind, that, as this evidence is increased in quantity, the probability that A will be in any instance a sign of B , tends to a limit which lies betrveen two determinate ratios. just as the probability of an inductive generalisation may tend towards certainty, when the evidence is increased in a manner satisfying given conditions.

Let \(f(q)\) represent the proposition that \(q\) is the true value of \(l / m\). Let \(q^{\prime}\) represent the ratio of the number of instances actually before us in which A has been accompanied by B to that of the instances in which A has not been accompanied by B ; and let \(f^{\prime}\left(q^{\prime}\right)\) be the proposition which asserts this. Now if the ratio \(q\) is known, then, subject to the assumptions already stated, the number \(q\) must also represent the à priori probability in any instance, both before and after the results of other instances are known, that A, if it occurs, will be accompanied by B. We have, in fact, the conditions as set forth in Chapter XXIX., in which Bernoulli's Theorem can be validly applied, so that this theorem enables us to give a numerical value, for all numerical values of \(q\) and \(q^{\prime}\), to the probability \(f^{\prime}\left(q^{\prime}\right) / h . f(q)\), -which expression represents the likelihood à priori of the frequency \(q^{\prime}\), given \(q\).

An application of the inverse formula allows us to infer from the above the \(\grave{a}\) posteriori probability of \(q\), given \(q^{\prime}\), namely :
\[
\begin{aligned}
& f(q) / h \cdot f\left(q^{\prime}\right) / h \cdot f(q) \\
& \stackrel{\rightharpoonup}{f}(q) / h \cdot f\left(q^{\prime}\right), h \cdot f(q)
\end{aligned}
\]
where the sumination in the denominator covers all possible values of \(q\). In rough applications of this inverse of Bernoulli's Theorem it has been usual to suppose that \(f(q) / h\) is constant for all values of \(q\),-that, in other words, all possible values of the
ratio \(q\) are it priori equally likely. If this supposition were legitimate, the formula could be reduced to the algebraical expression
\[
\begin{aligned}
& f\left(q^{\prime}\right) h \cdot f(\eta), \\
& -f\left(q^{\prime}\right) h \cdot f(\eta),
\end{aligned}
\]
all the terms of which can be determined numerically be Bernoulli's Theorem. It is easy to show that it is a maximum when \(q=q^{\prime}\). i.e. that \(q^{\prime}\) is the most probable value of \(l \mathrm{~m}\), and that. when the instances are sery numerous, it is very improbable that I, \(m\) differs from \(q^{\prime}\) widely. If, therefore, the number of instances is increased in such a manner that the ratio continues in the neighbourhood of \(q^{\prime}\), the probability that the true value of l.m is nearly \(q^{\prime}\) tends to certainty ; and, consequently, the probability, that A is in any instance a sign of B , also tends to a magnitude which is measured by \(q^{\prime}\).

I sse, however. no justification for the assumption that all posishle values of the ratio \(q\) are it priori equally likely. It is not exan equivalent to the assumptions that all integral values of \(l\) an: d m respectively are "qually probable. 1 am not satisfied fither that different values of \(q\). or that different values of \(I I\). satisfy the conditions which have been latd down in Part I. for alternatives which ase equal before the Principle of Indifference. There sem.m. for instaner, to be relerant differences between the statement that I can arise in exactly two ways and the statement that it can arise in exactly a thousand ways. We must. therefore, be contont with some lesser assumption and with a less precise form for our final conclusion.
4. Since, in accordance with our hypothesis, m. cannot exeeed some finite number, and since \(l\) must necessarily be less than \(m\), the possible values of \(m\), and therefore of \(q\), are finite in number. Perhaps we can assume therefore, as ome of our fundamental assumpuimes, that there is a priori a finite probability in favour of rach of these possible values. Lut \(\mu\) be the finite number which in camot exceed. Then there is a fimite probability for each of the intervals \({ }^{1}\)
\[
\text { I to } \stackrel{\ddot{\partial}}{\mu},{\underset{\mu}{\mu}}_{\mu}^{2} \text { to }_{\mu}^{3}, \ldots \mu_{\mu}^{\mu} \text { to } 1
\]

\footnotetext{
 limit.
}
that \(q\) lies in this interval ; but we cannot assume that there is an equal probability for each interval.

We must now return to the formula
\[
\begin{aligned}
& f(q) / h \cdot f\left(q^{\prime}\right) / h f(q) \\
& f(q) / h \cdot f\left(q^{\prime}\right) / h f(q)^{\prime}
\end{aligned}
\]
which represents the \(\grave{a}\) posteriori probability of \(q\), given \(q^{\prime}\). Since by sufficiently increasing the number of instances, the sum of terms \(f\left(q^{\prime}\right) / h f(q)\) for possible values of \(q\) within a certain finite interval in the neighbourhood of \(q^{\prime}\) can be made to exceed the other terms by any required amount, and since the sum of the values of \(f(q) / h\) for possible values of \(q\) within this interval is finite, it clearly follows that a finite number of instances can make the probability, that \(q\) lies in an interval of magnitude \(1 / \mu\) in the neighbourhood of \(q^{\prime}\), to differ from certainty by less than any finite amount however small.
5. We have, therefore, reached the main part of the conclusion after which we set out-namely, that as the number of instances is increased the probability, that \(q\) is in the neighbourhood of \(q^{\prime}\), tends towards certainty ; and hence that, subject to certain specified conditions, if the irequency with which \(B\) accompanies A is found to be \(y^{\prime}\) in a great number of instances, then the probability that A will be accompanied by B in any further instance is also approximately \(q^{\prime}\). But we are left with the same vagueness, as in the case of generalisation, respecting the value of \(\mu\) and the number of instances that we require. We know that we can get as near certainty as we choose by a finite number of instances, but what this number is we do not know. This is not very satisfactory, but it accords very well, I think, with what cormmon sense tells us. It would be very surprising, in fact, if logic could tell us exactly how many instances we want, to yield us a given degree of certainty in empirical arguments.

Nobody supposes that we can measure exactly the probability of an induction. Yet many persons seem to believe that in the weaker and much more difficult type of argument, where the association under examination has been in our experience, not invariable, but merely in a certain proportion, we can attribute a definite measure to cur future expectations and can claim practical certainty for the results of predictions which lie within relatively narrow limits. Coolly considered, this is a preposter-
ous claim, which would have been universally rejected long ago, if those who made it had not so successfully concealed themselves from the eyes of common sense in a maze of mathematics.
6. Meantime we are in danger of forgetting that, in order to reach wen our modified conclusion, material assumptions have been introduced. In the first place, we are faced with exactly the same difficulties as in the case of universal induction dealt with in Part III., and our original starting-point must be the same. We have the same dificulty as to how our initial probability is to beobtained : and I have no better sugerestion to offer in this than in the former case-namely, the supposed principle of a limitation of independent varietr in experience. We have to suppose that if \(A\) and \(B\) oecur torether (i.e. are true of the same object), this is some just appreciahle reason for supposing that in this instance they have a common cause : and that, if A oecurs again, this is a just apprectable reason for supposing that it is due to the same cause as on the former occasion. But in addition th, the usual inductive hypothesis, the argument has rested on two particularly important assumptions, firat. that we have no reason for supposing that some of the. events of which A may be a sign are more likely to be exemplition in some of the particular instances than in others, and secondre that the analogy amongst the examined B's is perfect. Tho first assumption amoments, in the languare of statisticians, io an assumption of rutudom sumpling from anongst the I's. The seeond assumption corresponds precisely to the similar condition which we discussed fullv in connection with inductive seneralisation. The instances of \(\mathrm{A}(x)\) may be the result of random sampling, and yet it may still be the case that there are material cireumstances. common (1) all the examined instances of \(B(x)\). yet not coverod by the statement \(\mathrm{A}(r) \mathrm{B}(r)\). In so far as there two assumptions are not justified, an element of doubt and vameness, which is not easily. measured, assails the argument. It is an element of doubt precisely similar to that which exists in the case of erneralisation. But we are most likely to fored it. For havine owercomthe difticulties pereuliar to correlation.' it is, posisibly: not un-

\footnotetext{
\({ }^{1}\) I am here using this torm in distinction to generalisation; that is to say, I call the statement that \(A(x)\) is always accompanied by \(B(x)\) a generalisation, and the statement that \(\mathrm{A}(x)\) is accompanied by \(\mathrm{B}(x)\) in a certain proportion of cases a correlation. This is not quite identical with its use by modern statiolwlul.
}
natural for a statistician to feel as if he had overcome all the difficulties.

In practice, however, our knowledge, in cases of correlation just as in cases of generalisation, will seldom justify the assumption of perfect analogy between the B's; and we shall be faced by precisely the same problems of analysing and improving our knowledge of the instances, as in the general case of induction already examined. If B has invariably accompanied A in 100 cases, we have all kinds of difficulties about the exact character of our evidence before we can found on this experience a valid generalisation. If B has accompanied A , not invariably, but only 50 times in the 100 cases, clearly we have just the same kind of difficulties to face, and more too, before we can announce a valid correlation. Out of the mere analysed statement that \(B\) has accompanied A as often as not in 100 cases, without precise particulars of the cases, or even if there were \(1,000,000\) cases instead of 100 , we can conclude very little indeed.

\section*{(IIAPTER XXXII}

\section*{THE INDUCTIVE USE OF STATISTICAL FREQUENCIES FOR THE DETERMINATION OF PROBABILITY A POSTERIORI-THE METHODS OF LEXIS}
1. No ome suppuses that a good induction can be arrived at mernly her combing cases. The business of strengthening the argument whefly amsist in determining whether the alleged association is stable, when the accompanying conditions are varime. This process of improving the Analogy, as I hate called it in Part III.. is, both logically and practically of the essence of the argument.

Now in statistical reasoning (or inductive correlation) that part of the argument. Whirh corresponds to rombting the (ases in motactios qeneralisation. way present considerable technical dilliculty: This is esperially so in the particularly comples cases of what in the nest chapher (\$9) I shall term (!umentulior C'orrelolion". Whish have ereatly occupied the attention of English statisticians in recent years. But clearly it would be an error to suppose that, when we have successfully overcome the mathematical or other terhmical ditheulties, we have made ans greater prospecs towards estahlishing our conclusion than when, in the case of inductive generalisation, we have counted the cases but have not yet analysed or compared the descriptive and nonnumerical differences and resemblances. In order to get a good scemtitie arernment we still have to pursue precisely the same scientifir mothods of experiment, amalysis, comparizon, and differentiation as are recomphised to be mecessary to estahlish ant sedontifer emoralisation. These methods are not reducible to a precise hathematical form for the reasons examined in I'art \(1 / I\). of this treatiore. But that is no reason for ignoring them, or for pretemtine that the calculation of a probabilits. which takes into
account nothing whatever except the numbers of the instances, is a rational proceeding. The passage already quoted from Leibniz (In exemijis juridicis politicisque pilerumque non tamen subtili calculo opres est, quam accuratu omnium circomstantiarum enumeratione) is as applicable to scientific as to political incuiries.

Generally speaking, therefore, I think that the business of statistical technique ought to be regarded as strictly limited to preparing the numerical aspects of our material in an intelligible form, so as to be ready for the application of the usual inductive methods. Statistical technique tells us how to 'count the cases when we are presented with complex material. It must not proceed also, except in the exceptional case where our evidence furnishes us from the outset with data of a particular kind, to turn its results into probabilities; not, at any rate, if we mean by probability a measure of rational belief.
2. There is, however, one type of technical, statistical investigation not yet discussed, which seems to me to be a valuable aid to inductive correlation. This method consists in breaking up a statistical series, according to appropriate principles, into a number of sub-series, with a view to analysing and measuring, not merely the frequency of a given character over the aggregate series, but the stability of this frequency amongst the subseries ; that is to say, the series as a whole is divided up by some principle of classification into a set of sub-series, and the fluctuation of the statistical frequency under examination between the various sub-series is then examined. It is, in fact, a technical method of increasing the Analogy between the instances, in the sense given to this process in Part III.
3. The method of analysing statistical series, as opposed to the Laplacian or mathematical method, one might designate the inductive method. Independently of the investigations of Bernoulli or Laplace, practical statisticians began at least as early as the end of the seventeenth century \({ }^{-1}\) to pay attention to the stability of statistical serices when analysed in this manner. Throughout the eightcenth century, students of mortality statistics, and of the ratio of male to female lirths (including Laplace himself), paid attention to the degree of constancy of the

\footnotetext{
\({ }^{1}\) Graunt in his Natural and Political Observations upon the Bills of Mortality has been quoted as one of the carliest statisticians to pay attention to these considerations.
}
ratios ove different parts of their series of instances as well as to their average value over the whole series. And in the early part of the nineteenth century, Quetelet, as we have already noticed, widely popularised the notion of the stability of various social statistics from year to year. Quetelet, however, sometimes asserted the existence of stability on insufficient evidence. and involved himself in theoretical errors through imitating the methods of Laplace too closely; and it was not until the last quarter of the nineteenth century that a school of statistical theory was fommed, which gave to this way of approaching the problem the system and technique which it had hitherto lacked, and at the same time made explicit the contrast between this analytical or inductive method and the prevailing mathematical theory: The sole founder of this school was the German economist, Wilhelm Lexis. whose theories were expounded in a swies of articles and mommraphs published between the years 1875 and 1859. For some years Lexis's fundamental ideas did not attract much notice, and he himself seems to have turned his attention in other directions. But more recently a comsiderable literature has grown up round them in Germany, and their fall purport has been expressed with more clearness than by Lexis himself -although mo one. with the exception of Ladislaus vom Bortkiewicz, has been able to make additions to them of any great signifiance. \({ }^{1}\) Lexis devised his theory with an immediate view to its practical application to the problems of sex ration and mortality. The fact that his general theory is so closely intermingled with these particular applications of it is, probably, a part explanation of the long interval which elapsed before the general theoretical inportance of his ideas was widely realised. I cannot help doubting how fully Lexis himself realised it in the first instance. It would certainly bee easy to read his earlier contributions to the question without appreciating their generaliowd significance. Atter leig lexis added mothing substantial to his earlier work, and lator developments are mainly due to Von

\footnotetext{
\({ }^{1}\) A list of Lexis's principal writings on these topics will be found in the Bibliography. There is little of first-rate importance which is not contained

 Statistik: In this latter volume the two important articles on "Die Theorie der stahilitat stati-tiaher Reeihen" and on "Das Gerchlechtaverhaltnis dor Geborenen und die Wahrscheinlichkeitsrechnung," originally published in Con-

}

Bortkiewicz. Those of the latter's writings, which have an important bearing on the relation between probability and statistics, are given in the Bibliography. \({ }^{1}\)

On the logic and philosophy of Probability writers of the school of Lexis are in general agreement with Von Kries ; but this seems to be due rather to the reaction which is common both to him and to them against the Laplacian tradition, than to any very intimate theoretical connection between Von Kries's main contributions to Probability and those of Lexis, though it is true that both show a tendency to find the ultimate basis of Probability in physical rather than in logical considerations. I am not acquainted with much work, which has been appreciably influenced by Lexis, written in other languages than German (including with Germans, that is to say, those Russians, Austrians, and Dutch who usually write in German, and are in habitual connection with the German scientific world). In France Dormoy \({ }^{2}\) published independently and at about the same time as Lexis some not dissimilar theories, but subsequent French writers have paid little attention to the work of either. Such typical French treatises as that of Bertrand, or, more recently, that of Borel. contain no reference to them. \({ }^{3}\) In Italy there has been some discussion recently on the work of Von Bortkiewicz. Among Englishmen Professor Edgeworth has shown a close acquaintance with the work of the German school, \({ }^{4}\) he providing for nearly forty years past, on this as on other matters where the realms of

\footnotetext{
\({ }^{1}\) The reader may be specially referred to the Kritische Betrachtungen zur theoretischen Statistik ( first instalment - the later instalments being of less interest to the student of Probability), the Anwendungen der W'ahrscheinlichkeitsrerhnung auf Statistik, and Homogeneität und Stubilitüt in der Statistik. Of other German and Russian writers it will be sufficient to mention here Tschuprow, who in " Die Aufgaben der Theorie der Statistik" (Schmoller's.Jahrbuch, 1905) and "Zur Theorie der Stabilität statistischer Reihen " (Skandinavisk Aktuarietidskrift) gives by far the best and most lucid general accounts that are available of the doctrines of the school, he alone amongst these authors writing in a style from which the foreign reader can derive pleasure, and Czuber, who in his Wahrscheinlichkeitsrechnung (vol. ii. part iv. section 1) supplies a useful mathematical commentary.
\({ }^{2}\) Journal des actuaires fruncais, 1874, and Théorie mathématique des ass,surances sur la vie, 1878 ; on the question of priority see Lexis, Abhandlungen, p. 130.
\({ }^{3}\) Though both these writers touch on closely cognate matters, where Lexis's investigations would be highly relevant-Bertrand, Calcul, pp. 312-314; Borel, Eléments, p. 160.
\({ }^{4}\) See especially his "Methods of Statistics" in the Jubilee Volume of the Stat. Journ., 1885, and "Application of the Calculus of Probabilities to Statistics," International Statistical Institute Bulletin, 1910.
}

Statistics and Probability overlap, almost the only comecting link between English and continental thought.

Nevertheless. an account in English of the main doctrines of this school is still lacking. It would be outside the phan of the present treatise to attempt such an account here. But it may. be useful to give a short summary of Lexis's fundanontal ideas. After giving this accoment I shall find it comvenient, in proceeding to my uwn incomplete observations on the matter, to approach it from a rather different standpoint from that of Lexis or of Vom Bortkiewicz, though not for that reason the less influenced or illuminated by their eminent contributions to this problem.
4. It will be chearer to begin with some analysis due to Vom Bortkiewicz.! and then to proceed to the method of Lexis himself, although the latter came first in point of time.

A group of observations may be made up of a number of sul)groups. to which different freguencies for the character under invertigation are properly applicable. That is to say, a proper1 ion \(z_{z}^{z_{1}}\) of the observations may belong to: aroup. for which, given the frequency, the it prion prohability of the character under ohservation in a particular instance would be \(p_{1}\), a propertion \({ }_{z}^{z_{2}}\) may helong to a second group for which \(p_{2}\) is the probability, and so on. In this case, given the frequencies for the sub-groups. the probability fo for the group as a whole wowh he made up as follows:
\[
r={ }_{z}^{z_{1}} h_{1} \quad{ }_{z}^{z_{2}} f_{2}: \ldots
\]

We may call \(p\) a general probabitity, and \(p\), etc., special probabilitios. Bint the sperial probabilities may in their furn ber general probabilities. so that there mat he mom than one way of resolving a general probability into special probabilities.

If \(p_{1}=p_{2}=\ldots=p\), then \(p\), for that particular way of resolving the total gromp, into partial groups, is, in Bortkiewic\%s termin chogy, imdifferent. If \(p\) is indifferent for all conceivable resolution. into partial groupr. \({ }^{2}\) then, borrowing a phrase from Von Kries. Bortkiewirz sats of it that it has a defimitire interperention. In
\({ }^{1}\) What follows is a free rendering of sume passages in his Kritische Bitrurhturifer.
\({ }^{2}\) This is clearly a very lonse statement of what Bortkiewicz reully means.
dealing with à priori probabilities, we can resolve a total probability until we reach the special probabilities of each individual case ; and if we find that all these special probabilities are equal, then, clearly, the general probability satisfies the condition for definitive interpretation.

So far we have been dealing with à priori probabilities. But the object of the analysis has been to throw light on the inverse problem. We want to discover in what conditions we can regard an observed frequency as being an adequate approximation to a definitive general probability.

If \(p^{\prime}\) is the empirical value of \(p\) (or, as I should prefer to call it, the frequency) given by a series of \(n\) observations, we may have
\[
p^{\prime}={ }_{n}^{n_{1}} p_{1}^{\prime}+{ }_{n}^{n_{2}} p_{2}^{\prime}+\ldots
\]

Even if this particular way of resolving the series of observations is indifferent, the actually observed frequencies \(p_{1}{ }^{\prime}, p_{2}{ }^{\prime}\), etc., may nevertheless be unequal, since they may fluctuate round the norm \(p\) ' through the operation of 'chance' influences. If, however. \(n_{1}, n_{2}\), etc., are large, we can apply the usual Bernoullian formula to discover whether, if there was a norm \(p^{\prime}\), the divergences of \(p_{1}{ }^{\prime}, p_{2}{ }^{\prime}\), etc., from it are within the limits reasonably attributable on Bernoullian hypotheses to 'chance' influences. We can, however, only base a sound argument in favour of the existence of a 'definitive' probability \(p^{\prime}\) by resolving our aggregate of instances into sub-series in a great variety of ways, and applying the above calculations each time. Even so, some measure of doubt must remain, just as in the case of other inductive arguments.

Bortkiewicz goes on to say that probabilities having definitive interpretation (definitive Bedeutung) may be designated elementary probabilities (Elementarwahrscheinlichkeiten). But the probabilities which usually arise in statistical inquiries are not of this type, and may be termed average probabilities (Durchschnittswahrscheinlichkeiten). That is to say, a series of observed frequencies (or, as he calls them, empirical probabilities) does not, as a rule, group itself as it would if the series was in fact subject to an elementary probability.
5. This exposition is based on a philosophy of Probability different from mine ; but the underlying ideas are capable of
translation. Suppose that one is endeavouring to estahlish an inductive correlation, e.g. that the chance of a male birth is \(m\). The conclusion, which we are seeking to establish, takes no account of the place or date of birth or the race of the parents. and assumes that these influences are irrelevant. Now, if we had statistics of birth ratios for all parts of the world throughout the nineteenth epntury: and added them all up and found that the average frequency of male births was \(m\), we should not be justified in arguing from this that the frequency of male births in England next year is very unlikely to diverge widely from \(m\). For this would involve the unwarranted assumption, in Bortkiewiczas terminology, that the empirical probability \(m\) is elementary for any resolution dependent on time or place, and is not an average prohahility compounded out of a series of groups, relating to different times or places, to each of which a distinct speerial prohaibity is applicable. And. in my terminology, it would assume that variations of time and place were irrelevant to the correlation. without any attempt having been made to employ the met hods of positive and negative Analogy to establish this.

We must, therefore, break up our statistical material into groups by date, place, and any other characteristic which our generalisation proposes to treat as irrelevant. By this meame we shall , htam a number of frequencies \({ }^{\prime}{ }_{1}{ }^{\prime}\). \(m_{2}^{\prime}, m_{3}^{\prime}, \ldots . m_{1}{ }^{\prime}\) ". \(m_{2}{ }^{\prime \prime}, m_{3}{ }^{\prime \prime}, \ldots\) etc., which are distributed round the average frequency in. For simplicity let us consider the series of frequencies \(m_{1}{ }^{\prime}, m_{2}{ }^{\prime}, m_{3}{ }^{\prime}\), . . . obtained by breaking up our material according to the date of the birth. If the observed diveremoes of these frequencies from their mean are not signiticant, we have the beginnings of an inductive argument for regarding dute as being in this connection irrelevant.
6. At this point Lexis's fundamental contribution to the problem must be introduced. He concentrated his attention in the mature of the dispersion of the frequencies \(m_{1}{ }^{\prime}, m_{2}{ }^{\prime}, m_{3}{ }^{\prime}\) romm their mean value \(m\); and he sought to devise a terhmical buethen for measuring the degree of stability displayed liy the suries of suld-frequencies, which are riedded by the various possible reriteria for resolving the aggregate statistical material into a number of constituent groups.

For this purpose he classified the various types of dispersion which could oceur. It may be the case that some of the sub)
frequencies show such wide and discordant variations from the mean as to suggest that some significant Analogy has been overlooked. In this event the lack of symmetry, which characterises the oscillations, may be taken to indicate that some of the subgroups are subject to a relevant influence, of which we must take account in our generalisation, to which some of the other subgroups are not subject.

But amongst the various types of dispersion Lexis found one class clearly distinguishable from all the others, the peculiarity of which is that the individual values fluctuate in a 'purely chance' manner about a constant fundamental value. This type he called typical (typische) dispersion. He meant by this that the dispersion conformed approximately to the distribution which would be given by some normal law of error.

The next stage of Lexis's argument \({ }^{1}\) was to point out that series of frequencies which are typical in character may have as their foundation either a constant probability, \({ }^{2}\) or one which is itself subject to chance variations about a mean. The first case is typified by the example of a series of sets of drawings of balls, each set being drawn from a similar urn ; the second case by the example of a series of sets of drawings, the urns from which each set is drawn being not similar, but with constitutions which vary in a chance manner about a mean.

As his measure of dispersion Lexis introduces a formula, which is evidently in part conventional (as is the case with so many other statistical formulae, the particular shape of which is often determined by mathematical convenience rather than by any more fundamental criterion). He expresses himself as follows. Where the underlying probability is constant, the probable error in a particular frequency à priori is \(r=\rho \int \begin{gathered}\frac{2 u(1-c)}{g}\end{gathered}\), where \(\rho=4769, v\) is the underlying probability, and \(y\) is the number of instances to which the frequency refers. This follows from the usual Bernoullian assismptions. Now let \(R\) be the corresponding expression derived \(a\) posteriori by reference to the actual deviations of a scries of ohserved frequencies from their mean, so that
\({ }^{1}\) I am here following fairly closely his paper, "Über die Theorie der Stabilität statisticher Recihen," "reprinted in his Abhundlunyen zur Theorie der Bevölkerungsund Moral-Statistik, pp. 170-212.
\({ }_{2}\) This mode of expression, which is not in accurate conformity with my philosophy of Probability, is Lexis's, not mine. His meaning is intelligible.
\(\mathrm{R}=\rho \leq \sum_{\left.2 \mid \delta^{2}\right\rfloor}^{\prime-1}\), where \(\left[\delta^{2}\right]\) is the sum of the squares of the deviations of the individual frequencies from their mean and \(u\) is their number. Now, if the ohserved facts are due to merely chance variations about a constant \(r\), we must have approximately \(\mathrm{R}=r\), though, if ! is small, comparatively wide deviations between R and, will mot he signiticant. If, on the other hand, \(r\) itself is not comstant hut is subject to chance variations, the case stands difterently. For the fluctuations of the observed frequencies are now due to two components. The one which would be present, even if the underlying probability were constant, laxis terms the ortinary or unessential compenent: the other har terms the physical component. If \(f\) is the probable deviation of the various values of \(v\) from their mean, then, on the same assumphions and as a deduction from the same theory as before. \(R\) will tend to equal not \(r\) but \(\sqrt{ } / r^{2} \cdots r^{2}\). In this event \(R\) cannot be less than \(r\). If, therefore, \(\mathrm{R}<r\), one must suppose that the individual instances of each several series on which each frequency is based are not independent of one another. Such a series Lexis terms an organic or dependent (gebundene) series, and explains that it camot he handled by purely statistical methods.

Since, therefore, we have three types of series, differing fundamentally from one another according as \(\mathrm{R}=r\), \(>r\), or \(<r\), R
Lexis puts. \({ }_{r}=Q\), and takes \(Q\) as his measure of dispersion. \({ }^{1}\) If \(Q=1\), we have normal dispersion; if \(Q>1\), we have supernormal dispersion ; and if \(Q<1\), we have subnormal dispersion, which is an indication that the series is 'organic.

If the mumber of instances on which the frequencies are based is very ereat. 6 becones negligible in comparisen with \(l\) (the physical component), and, therefore, \(\mathrm{R}=\sqrt{r^{2}+p^{2}}\) becomes apposimately \(R \quad \mu\). On the other hamd, if \(\mu\) is mot very large and the hase number of instances is small, \(p\) beeomes negligible


 set, \(n\) the number of sets, \(p_{k}\) the frequency for set \(k\), and \(p\) the mean of the \(n\) frequencies.
in comparison with \(r\), and we have a delusive appearance of normal dispersion. \({ }^{1}\) Lexis well illustrates the former point by the example that the statistics of the ratio of male to female births for the forty-five registration districts of England over the years 1859-1871 approximately satisfy the relation \(R=r\). But if we take the figures for all England over those thirteen years, although the extreme limits of the fluctuation of the ratio about its mean 1.042 are 1.035 and 1.047 , nevertheless \(R=2.6\) and \(r=1.6\), so that \(Q=1 \cdot 625\); the explanation being that the base number of instances, namely 730,000 , is so large that \(r\) is very small, with the result that it is swamped by the physical component \(p\). And he illustrates the latter point by the assertion that, if in 20 or 30 series each of 100 draws from an urn containing black and white balls equally, the number of black balls drawn each time were only to vary between 49 and 51, he would have confidence that the game was in some way falsified and that the draws were not independent. That is to say, undue regularity is as fatal to the assumption of Bernoullian conditions as is undue dispersion.
7. In a characteristic passage \({ }^{2}\) Professor Edgeworth has applied these theories to the frequency of dactyls in successive extracts from the Aeneid. The mean for the line is \(1 \cdot 6\), exclusive of the fifth foot, thus sharply distinguishing the Virgilian line from the Ovidian, for which the corresponding figure is \(2 \cdot 2\). But there is also a marked stability. "That the Mean of any five lines should differ from the general Mean by a whole dactyl is proved to be an exceptional phenomenon, about as rare as an Englishman measuring 5 feet, or 6 feet 3 inches. An excess of two dactyls in the llean of five lines would be as exceptional as an Englishman measuring 6 feet 10 inches." But not only so--the stability is excessive, and the fluctuation is less " than that which is obtained upon the hypothesis of pure sortition. If we could imagine dactyls and spondees to be mixed up in the pret's brain in the proportion of 16 to 24 and shaken out at random, the modulus in the number of dactyls would be \(1 \cdot 38\), whereas we have constantly obtained a smaller number, on an average (the square root of the average fluctuation) \(1 \cdot 2\)." (On Lexian principles these statistical results would support the hypothesis that the

\footnotetext{
\({ }_{1}\) This is part of the explanation of Bortkiewicz's Law of Small Numbers. See also p. 401.

2 "On Methorls of Stitistics," Jubilee Volume of the Royal Statistical Socicty, p. 211.
}
series under investigation is 'orqanic' and not subject to Bernoullian conditions. an hypothesis in accordance with our ideas of poetry. That Edeworth should have put forward this example in criticism of Lexiss conclusions, and that Lexis \({ }^{1}\) should have retorted that the explanation was to be found in Edgeworth's srries' not consisting of an adequate number of separate ohservations, indicates, if I do not misapprehend them. that these authorities are at fault in the principles, if not of Probability, of Poetry.

The dactuls of the Virgilian hexamoter are, in fact. a very grood example of what has been termed comnexité, leading to subnormal dispersion. The quantities of the successive feet are not independent, and the appearance of a dactyl in one foot dimimishes the probability of amother dactyl in that line. It is like the case of drawing hatek and white balls out of an urn. where the balls are not replaced. But Lexis is wrong if he suppeses that a superneromel diapersion cammot also arise out of connesile, on organic connection between the successive terms. It might have been the case that the appearance of a dactyl in one foot increased the probability of another dactyl in that line. He should, I think. have contemplated the result \(R \rightarrow r\) an possibly indicating a non-typical. oremice suris. and shomld mot hase assumed that. where R is greater than \(r\), it is of the form \(\sqrt{ } r^{2}+p^{2}\).

In short. Iexis hats nom pushed his analysis far enough, and how has mot fully comprethemded the character of the underlying conditions. But this dowe mot affere the fact that it was he who made the vital advance of taking as the unit, not the single ohservation, but the freguency in given comditions, and of concerving the mature of statistical induction an consisting in ther examination. and if possible the measmement, of the stability of the frequency when the conditions are varied.
8. There is one sperial piece of work illustrative of the abowe methents. due to lim bortkiewiez, which mast not be overlooked, and which it is convenient to introduce in this place the socalled Low of simull . .imbines. \({ }^{2}\)

Quetente as wr hate sem in Chapter XXIIII.. called attemtion

\footnotetext{
1 ". 'her de. Wahrachemlichkeitarechnung," P. 44 (see Bihlingraphy).
\({ }^{2}\) There are numerous references to this phenomenon in periodical literature ;
 Z.ulico
}
to the remarkable regularity of comparatively rare events. Von Bortkiewicz has enlarged Quetelet's catalogue with modern instances out of the statistical records of bureaucratic Germany. The classic instance, perhaps, is the number of Prussian cavalrymen killed each year by the kick of a horse. The table is worth giving as a statistical curiosity. (The period is from 1875 to 1894; G stands for the Corps of Guards, and I.-XV. for the 15 Army Corps.)
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline & 75 & 76 & 77 & 78 & 79 & 80 & 81 & 82 & 83 & & 4 & 85 & 86 & 87 & 88 & 89 & 90 & 91:92 & 93 & 94 \\
\hline G. & . & 2 & 2 & 1 & & . & 1 & 1 & . & & 3 & . & 2 & 1 & . & - & 1 & 1 & & 1 \\
\hline I. & & . & & 2 & . & 3 & . & 2 & . & & . & . & 1 & 1 & 1 & - & 2 & . 3 & 1 & \\
\hline II. & & & & 2 & - & 2 & . & . & 1 & & 1 & . & . & 2 & 1 & 1 & . & . 2 & & \\
\hline III. & & & & 1 & 1 & 1 & 2 & . & 2 & & . & . & . & 1 & . & 1 & 2 & 1 . & & \\
\hline IV. & & 1 & . & 1 & 1 & 1 & 1 & . & . & & . & . & 1 & . & \(\because\) & \(\cdots\) & \(\cdots\) & 11 & & \\
\hline V. & & . & & . . & 2 & 1 & . & \(\because\) & 1 & & . & \(\cdots\) & 1 & \(\cdots\) & 1 & 1 & 1 & 11 & 1 & \\
\hline VI. & & . & 1 & . & 2 & . & & 1 & 2 & & . & 1 & 1 & 3 & 1 & 1 & 1 & 3 & \(\cdots\) & \\
\hline VII. & 1 & . & 1 & . & & . & 1 & . & 1 & & 1 & . & & 2 & . & . & 2 & 1 & 2 & \\
\hline VIII. & 1 & . & & & 1 & & & 1 & & & . & & & 1 & . & \(\cdots\) & \(\cdots\) & 11 & & 1 \\
\hline IX. & & & & & & 2 & 1 & 1 & 1 & & \(\cdots\) & 2 & 1 & 1 & . & 1 & 2 & . 1 & " & \\
\hline X. & & & 1 & 1 & & 1 & & 2 & & & 2 & . & . & & . & 2 & 1 & 3 . & 1 & 1 \\
\hline XI. & & & & & 2 & 4 & & 1 & 3 & & . & 1 & 1 & 1 & 1 & 2 & 1 & 3 1 & 3 & 1 \\
\hline XIV. & 1 & 1 & 2 & 1 & 1 & 3 & & 4 & & & 1 & \(\cdots\) & 3 & 2 & 1 & \(\cdots\) & 2 & 1 1 & & \\
\hline XV. & . & 1 & . & \(\ldots\) & . & & & 1 & & & 1 & 1 & . . & . & . & 2 & 2 & . . . & & \\
\hline
\end{tabular}

The agreement of this table with the theoretical results of a random distribution of the total number of casualties is remarkably close : \({ }^{1}\)
\begin{tabular}{c|cc|}
\hline \begin{tabular}{c} 
Casualties in a \\
Year.
\end{tabular} & \begin{tabular}{c} 
Number of Occasions on which the Annual \\
Casualties in a Corps reach the Figure \\
in Column 1.
\end{tabular} \\
& \multicolumn{3}{c|}{\begin{tabular}{c} 
Actual.
\end{tabular}} \\
0 & 144 & Theoretical. \\
0 & 91 & \(143 \cdot 1\) \\
1 & 32 & \(92 \cdot 1\) \\
2 & 11 & \(33 \cdot 3\) \\
3 & 2 & \(8 \cdot 9\) \\
4 & \(\ldots\) & \(2 \cdot 0\) \\
5 and more & & \(0 \cdot 6\)
\end{tabular}

Other instances are furnished by the numbers of child suicides in Prussia, and the like.

It is Von Bortkiewicz's thesis that these observed regularities

\footnotetext{
1 Bortkiewicz, op. cit. p. 24.
}
have a good theoretical explanation hehind them, which he dignifies with the name of the Law of simull Numbers.

The reader will recall that, according to the theory of Lexis, his measure of stability Q is, in the more general case. made up of two components, \(r\) and \(p\), combined in the expresion \(\sqrt[1]{1, \cdots} \quad i^{2}\). of which one is due tw tluctuations from the a verage of the conditions governing all the nembers of a series, which iumishess us with one of our observed frequencies, and of which the other is due to fluctuations in the individual members of the series ahout the true norm of the series. Bortkiewicz carries the same analysis a little further, and shows that Lexis's \(Q\) is of the form \(\sqrt{ } 1-(n-1)^{2}\). where \(n\) is the number of times that the event occurs in each series. \({ }^{1}\) That is to say, \(Q\) increases with \(n\), and, when \(n\) is small, \(Q\) is likely to exceed unity to a less extent than when \(u\) is large. To postulate that \(n\) is small, is, when we are dealing with observations drawn from a wide field, the same thing as to say that the event we are looking for is a comparatively rare one. This, in hrief, is the mathematical hasis of the Jaw of Small Numbers.

In his latest publinhed work on these toppics, \({ }^{2}\) Von Bontkiewiez builds his mathematical -tructure considerably higher. without. however. any further underpinning of the logical foundations of it. He has there worked out further statistical constants. arising out of the conceptions on which 1 axis s \(Q\) is hased (the precise beariny of which is not made any (learem ly his callingr thent corefficionls of symmeromy). Which are explicitly dependent on the value of 1 ; and he elatorately compares the theoretical value of the coefficients with the observed value in certain actual statistical material. He comeludes with the thesis, that Homogeneity and stability (defined as he defines them) are opposed conceptions, and that it is not correct to premise, that the larger statistical mass is as a rule more stable than the smaller, unless
\({ }^{1}\) I refer the reader to the original, op. rit. pl, \(2, \ldots 1\), for the inturpretation
 invertigatmon and for the mathematieal argument by wheh the abone result is justilial.
 nurisk Altumulalskift, 191s. Those readers. who lowk up my referencer. will, I think, agree with me that Von Purtkiewic\% does not fit any lews whseure as he trus on. The mathomatieal aremment is ripht chombh, atul wfen brilliat. Biat what it 18 all trally abons, what if all really amounts to.

we also assume that the larger mass is less homogeneous. At this point, it would have helped, if Von Bortkiewicz, excluding from his vocabulary homogeneity, paralromy, \(\gamma_{\mathrm{M}}^{\prime}\), and the like, had stopped to tell in plain language where his mathematics had led him, and also whence they had started. But like many other students of Probability he is eccentric, preferring algebra to earth.
9. Where, then, though an admirer, do I criticise all this? I think that the argument has proceeded so far from the premisses, that it has lost sight of them. If the limitations prescribed by the premisses are kept in mind, I do not contest the mathematical accuracy of the results. But many technical terms have been introduced, the precise signification and true limitations of which will be misunderstood if the conclusion of the argument is allowed to detach itself from the premisses and to stand by itself. I will illustrate what I mean by two examples from the work of Von Bortkiewicz described above.

Von Bortkiewicz enunciates the seeming paradox that the larger statistical mass is only, as a rule, more stable if it is less homogeneous. But an illustration which he himself gives shows how misleading his aphorism is. The opposition between stability and homogeneity is borne out, he says, by the judgment of practical men. For actuaries have always maintained that their results average out better, if their cases are drawn from a wide field subject to variable conditions of risk, whilst they are chary of accepting too much insurance drawn from a single homogencous area which means a concentration of risk. But this is really an instance of Von Bortkiewicz's own distinction between a general probability \(\rho\) and special probabilities \(p_{1}\) etc., where
\[
p={ }_{z}^{z_{1}} p_{1} \quad{ }_{z}^{z_{2}} p_{2}+\ldots
\]

If we are basing our calculations on \(p\) and do not know \(\mu_{1}, p_{2}\), etc., then these calculations are more likely to be borne out by \({ }^{\circ}\) the result if the instances are selected by a method which spreads them over all the groups \(1, \ddot{2}\), etc., than if they are selected by a method which concentrates them on group 1. In other words, the actuary does not like an undue proportion of his cases to be drawn from a group which may be subject to a common relevant influence for which he has not allowed. If the a priori calculations are based on the average over a field which is not homogeneous
in all its parts. greater stability of result will be obtamed if the instances are drawn from all parts of the non-homonemeous total field. than if they are drawn now from one homogemeous sub-field and now from another. This is not at all paradoxical. Yet I believer, though with hesitation, that this is all that Von Bortkiewic\%s elaborately supported mathematical conclusion really amounts to.

My second example is that of the Law of Small Numbers. Here also we are presented with an apparent paradox in the statement that the reqularity of ofecurrence of rare events is more stable than that of commoner events. Here, I suspect, the paradoxical result is really latent in the particular measure of stahility which has heen selected. If we look back at the figures. which I have quoted above, of Prussian eavalromen killed he the kirk of a horses, it is evident that a measure of stability could he chosen according to which exceptional instability would be displayed be this particular material ; for the frequence varies from (1) to 4 round a mean somewhat less than unity, which is a wery great percenterge fluctuation. In fact, the particular measure of stability which Von Bortkiewiez has adopted from Lexis has about it howewer mafol and comment it may he esperially for mathematical manipulation, a great deal that is arhitrary and conventional. It is only one out of a great many possible formulae which might be emploved for the mumerical measurement of the conception of stability. which, quantitatively at least, is not a perfectly precise one. Theso-called law of Small Numbers is. therefore little moere than a demonstration that, where rare erents are concerned, the Lexian measure of stability does not lead to satisfactory results. Like some other formulae which involve a nse of Bernoullian methods in an approximative form, it does not lead to reliable results in all circumstances. I should add that there is one other mement which may contribute to the total pischenterical reaction of the reader's mind to the Law of small Numbers. namely, the surprising and piquoml examples which are eited in suppert of it. It is startling and even ammsine to be told that horses kiek cavalremen with the same sort of regularity as characterises the rainfall. But our surprise at this particular example's fulfilling the l.aw of (ireat Numbers has little of monhing to do with the excentenal stability about which the Law of small Numbers purpert sonnern itself.

\section*{CHAPTER XXXIII}

\section*{OUTLINE OF A CONSTRUCTIVE THEORY}
1. Tumer is a great difference between the proposition "It is probable that erery instance of this generalisation is true " and the proposition "It is probable of any instance of this generalisation taken at random that it is true." The latter proposition may remain valid, even if it is certain that some instances of the generalisation are false. It is more likely than not, for example, that any number will be divisible either by two or by three, but it is not more likely than not that all numbers are divisible either by two or by three.

The first type of proposition has been discussed in Part III. under the name of Unicersal Induction. The latter belongs to Inductive Correlation or Statistical Induction, an attempt at the logical analysis of which must be my final task.
2. What adrocates of the Frequency Theory of Probability wrongly helieve to he characteristic of all probabilities, namely, that they are essentially concerned not with single instances but with series of instances, is, I think, a true characteristic of statistical imluction. A statistical induction either asserts the probability of an instance selectell ret rundom from a series of propositions. or else it assigns the probability of the assertion, that the truth frequency of a series of propositions (i.e. the proportion of true propesitions in the series) is in the neighbourhood of a given value. In either case it is asserting a characteristic of a serios of propsitions, rather than of a particular proposition.

Whilst, therefore, our unit in the case of Cniversal Induction is a single instance which satisfies both the condition and the conclusion of our generalisation, our unit in the case of Statistical

Induction is not a single instance, but a set or series of instances, all of which satisfy the condition of our generalisation but which satisfy the conclusion only in a certain proportion of cases. Ind whilst in Cniveraal Induction we build up our argument hy examining the known positive and negative Analogy shown in a series of single instances, the corresponding task in Statistical Induction consists in examining the Analogy shown in a series of series of instances.
3. We are presented, in problems of Statistical Induction, with a set of instances all of which satisly the comditions of our generalisation, and a propertion \(f\) of which satisfy its comelusion: and we seek to generalise as to the probable proportion in which further instances will satisfy the conclusion.

Now it is useless merely to pay attention to the proportion (or frequency) \(\int\) dismsered in the aggregate of the instances. For any collection whatever, comprising a definite number of ohjects, must, if the objectis be classified with reference to the presence or absence of any specified characteristic whatever. show some definite propertion or statistical frequency of oecurrence: so that a mere knowledge of what this frequency is can have no appreciable bearing on what the corresponding frequence will he for some other collection of oljeces. or on the probability of finding the characteristic in an object which does not belong to the original collection. We should be arguing in the same sort of way as if we were to base a universal induction as to the concurrence of two characteristics on a single observation of this concurrence and without any analysis of the accompanying circumstances.

Let the reader be clear about this. To argue from the mere fact that a given erent has occurred invariably in a thousand instances under observation. without any analysis of the circumstances accompanying the individual instances, that it is likely. to occur invariably in future instances, is a feeble inductive argument hecanse it takes no accoment of the Analogy. Neserthe. less an argment of this kind is mon entirely worthless. as we have seen in Part III. But to argue, without analysis of the instances. from the mere fact that a given event has a frequency of 10 per cent in the thousand instances under ohservation, or exen in at million instanes, that its probability is \(1 / 10\) for the next instamee. or that it is likely to have a frequency near to 1 ! 10 in a further
set of observations, is a far feebler argument ; indeed it is hardly an argument at all. Yet a good deal of statistical argument is not free from this reproach ;-though persons of common sense often conclude better than they argue, that is to sav, they select for credence, from amongst arguments similar in form, those in favour of which there is in fact other evidence tacitly known to them though not explicit in the premisses as stated.
4. The analysis of statistical induction is not fundamentally different from that of universal induction already attempted in Part III. But it is much more intricate ; and 1 have experienced exceptional difficulty, as the reader may discover for himself in the following pages, both in clearing up my own mind about it and in expounding my conclusions precisely and intelligibly. I propose to begin with a few examples of what conmonly impresses us as good arguments in this field, and also of the attendant circumstances which, if they were known to exist, might be held to justify such a mode of reasoning ; and, having thus attempted to bring before the reader's mind the character of the subjectmatter, to proceed to an abstract analysis.

Example One.-Let us investigate the generalisation that the proportion of male to female lirths is \(m\). The fact that the aggregate statistics for England during the nineteenth century yield the proportion \(m\) would go no way at all towards justifying the statement that the proportion of male lirths in Cambridge next year is likely to approximate to \(m\). Our argument would be no better if our statistics, instead of relating to England during the nineteenth century, covered all the descendants of Adam. But if we were able to break up our aggregate series of instances into a series of sulb-series, classified according to a great variety of principles, as for example by date. he season. by locality, hy the class of the parents, by the sex of previous children, and so forth, and if the proportion of male births throughout these subseries showed a significant stability in the neighbourhood of \(m\), then indeed we have an argument worth something. Otherwise we must either abandon our generalisation, amplify its conditions, or modify its conclusion.

Example Two.-Let us take a series of objects \(s\) all alike in some specified respect, this resemblane constituting membership, of the class F ; let us determine of how many members of the series a certain property \(\phi\) is true, the frequency of which is to be
the subject of our generaliation: and if a proportion \(f\) of the series \(s\) have the property \(\phi\), we may say that the series \(s\) has a frequency \(f\) for the property \(\phi\).

Now if the whole fied \(F\) has a finite number of constituents. it must have some determinate frequency \(p\), and if, therefore. we increase the comprehensiveness of \(s\) until eventually it includes the whole fiedd. \(f\) must come in the end to be equal to \(\mu\). This is ohsions and without interest and not what we mean be the law of great numbers and the stability of atatiotical frequency.

Let us now divide up the field F , according to some determinate principhe of division D , into subliolds \(\mathrm{F}_{1}\), \(\mathrm{F}_{2}\). etce; and let the series \(s_{1}\) be taken from \(\mathrm{F}_{1}, s_{2}\) from \(\mathrm{F}_{2}\), and so on. Where \(\mathrm{F}_{1}, \mathrm{~F}_{2}\), etc., have a finite number of constituents, \(s_{1}, s_{2}\), etc., may possibly coincide with them ; if \(s_{1}, s_{2}\), etc., do not coincide with \(F_{1}, F_{2},+1 c .\). hut are chosen from them, let us suppose that they are chosen according to some principle of random or unbiassed selection \(-s_{1}\), that is to say, will be a random sample from \(\mathrm{F}_{1}\). Now it may happen that the frequencies \(f_{1}, f_{2}\), etc., of the series \(s_{1}, s_{2}\), etc., thus selected cluster round some mean frequency \(f\). If the frequencies show this characteristic (the measurement and pre(ise determination of which I an not now (onsidering), then the series of series \(s_{1}, s_{2}\). etc., has a stable frequency for the classification D. 'Great numbers' only come in because it is difficult to ascertain the existence of stable frequency unless the series \(s_{1}\), \(s_{2}\). ete.. are themselver numerous and unless each of these comprises numerous individual instances.

Let us then apply a different principle of division I)', leading
 a thitat principte of division \(D\) )" leading 10 frequencies \(f_{1}{ }^{\prime \prime}\). \(f_{2}\) ", atce: and shon. to the full extent that our knowledee of the differences betwern the indivitual instaners permits us. If the frequencies
 we have an inductive ground of some weight for asserting a statistical generalisation.
1.4 the field F. for example. comprise all Englishmen in their sixtieth sare and let the property \(\phi\), about the frequeney of which we are erneralisinge the their death in that year of the ir age.


in their sixtieth year in \(1901, \mathrm{~F}_{2}\) in 1902, and so on ; or we might classify them according to the districts in which they live; or according to the amount of income tax they pay ; or according as they are in workhouses, in hospitals, in asylums, in prisons, or at large. Let us take the second of these classifications and let the subfields \(\mathrm{F}_{1}, \mathrm{~F}_{2}\), etc., be constituted by the districts in which they live. If we take large random selections \(s_{1}, s_{2}\). etc., from \(\mathrm{F}_{1}, \mathrm{~F}_{2}\), etc., respectively, and find that the frequencies \(\int_{1}, f_{2}\), etc., fluctuate closely round a mean value \(f\), this can be expressed by the statement that there is a stable frequency \(f\) for death in the sixtieth year in different English districts. We might also find a similar stability for all the other classifications. On the other hand, for the third and fourth classifications we might find no stability at all, and for the first a greater or less degree of stability than for the second. In the latter case the form of our statistical generalisation must be modified or the argument in its favour weakened.

Example Three.-Let us return to the example given in Chapter XXVII. of the dog which is fed sometimes by scraps at table and so judges it reasonable to be there. From one year to another, let us assume, the dog gets scraps on a proportion of days more or less stable. What sorts of explanation might there be of this? First, it might be the case that he was fed on the movable feasts of the Church; there would be the same number of these in each year, but it would not be easy for any one who had not the clue to discover any regularity in the occasions of their individual occurrence. Second, it might be the case that he was given scraps whenever he looked thin, and that the scraps were withheld whenever he looked fat, so that if he was given scraps on one day, this diminished the likelihood of his getting scraps on the next day, whilst if they were withheld this would increase the likelihood; the dog's constitution remaining constant, the number of days for scraps would tend to fluctuate from year to year about a stable value. Third, it might be the case that the company at table varied greatly from day to day, and that some days people were there of the kind who give dogs scraps and other days not; if the set of people from whom the company was drawn remained more or less the same from year to year, and it was a matter of chance (in the objective sense defined in § 8 of Chapter XXIV. above) which of them were
there from day to day, the proportion of days for scraps might again show some degree of stability from year to year. Lastly. a combination between the first and third type of circumstance gives rise to a variant deserving separate mention. It might be the case that the dog was only given scraps by his master, that his master generally went away for Saturday and Sunday, and was at home the rest of the week unless something happened to the contrary, and that "chance" causes would sometimes intervene to keep him at home for the week-end and away in the week; in this case the frequency of days for scraps would probably fluctuate in the neighbourhood of five-sevenths. In circumstances of this third type, however. the degree of stability. would probably be less than in circumstances of the first two types: and in order to get a really stable frequency it might he necessary to take a longer period than a year as the basis for each series of observations, or even to take the average for a number of dogs placed in like circumstances instead of one dog only.

It has been assumed so far that we have an opportunity of ohserving what happens on every day of the year. If this is not the case and we have knowledge only of a random sample from the days of each year, then the stability, though it will he less in degree may be nevertheless observable, and will increaw as the number of days included in each sample is increased. This applies equally to fach of the three types.
5. What is the correct logical analysis of this sort of reasoning? If an inductive generalisation is a true one. the conclusion which it asserts about the instance under inquiry is, so far as it ques. definit, and final. and cannot be modified by the acquisition of more detailed knowledge about the particular instance. But a statistical induction, when applied to a particular instance, is not like this: for the acquisition of further knowledge might remder the - tatistical induction, though not in itself less probable than before, inapplicable to that particular instance.

This is due to the fact that a statistical induction is met really about the particular instance at all, but has its subject. about which it generalises, a series; and it is omly applicable to the particular instance, in so far as the instance is relative to our knowledtee a random member of the series. If the acquisition of new knowledge affords us additional relevant information about
the particular instance, so that it ceases to be a random member of the series, then the statistical induction ceases to be applicable ; but the statistical induction does not for that reason become any less probable than it was--it is simply no longer indicated by our data as being the statistical generalisation appropriate to the instance under inquiry. The point is illustrated by the familiar example that the probability of an unknown individual posting a letter unaddressed can be based on the statistics of the Post Office, but my expectation that \(I\) shall act thus, cannot be so determined.

Thus a statistical generalisation is always of the form: 'The probability, that an instance taken at random from the series S will have the characteristic \(\phi\), is \(p\);' or, more precisely, if \(a\) is a random member of \(\mathrm{S}(x)\), the probability of \(\phi(a)\) is \(p\).

It will be convenient to recapitulate from Chapter XXIV. § 11 the definition of 'an instance taken at random' : Let \(\phi(x)\) stand for ' \(x\) has the characteristic \(\phi\),' and \(\mathrm{S}(x)\) for ' \(x\) is a member of the class \(\mathrm{S}^{\prime}\); then, on evidence \(h, a\) is a random member of the class S for characteristic \(\phi\), if ' \(x\) is \(a\) ' is irrelevant to \(\phi(x) / \mathrm{S}(x) . h,{ }^{1}\) i.e. if we have no information about \(a\) relevant to \(\phi(a)\) except \(\mathrm{S}(a)\).

Or alternatively we might express our definition as follows : Consider a particular instance \(a\), where the object of our inquiry is the probability of \(\phi(a)\) relative to evidence \(h\). Let us discard that part of our knowledge \(h(a)\) which is irrelevant to \(\phi(a)\), leaving us with relevant knowledge \(h^{\prime}(u)\). Let the class of instances \(a_{1}, \mu_{2}\), etc., which satisfy \(h^{\prime}(x)\) be designated by S. Then, relative to evidence \(h, a\) is a random member of the class or series S for the characteristic \(\phi\).

Let us denote the proposition ' \(x\) is, on evidence \(h\), a random member of S for characteristic \(\phi^{\prime}\) by \(\mathrm{R}(x, \mathrm{~S}, \phi, h)\); then our statistical generalisation is of the form \(\phi(x) / \mathrm{R}(x, \mathrm{~S}, \phi, h) . h=p\).

If \(\mathrm{R}(a, \mathrm{~S}, \phi, h)\) holds, then, on evidence \(h\), S is the appropriate statistical series to which to refer \(a\) for the purposes of the characteristic \(\phi\).

It is not always the case that the evidence indicates any series at all as 'appropriate' in the above sense. In particular,

\footnotetext{
\({ }^{1}\) The use of variables in probability, as has been pointed out on p. 58, is very dangerous. It might therefore be better to enunciate the above : \(a\) is a random member of \(S\) for characteristic \(\phi\), if \(\phi(a) / \mathcal{S}(a) \cdot h=\phi(b) / \mathcal{S}(b) . h\) where \(\dot{S}(b) . h\) contains no information about \(b\), except that \(b\) is a member of S
}
if evidence \(h\) indicates \(S\) as the appropriate series. and evidence \(h^{\prime}\) indicates \(\mathrm{S}^{\prime}\) as the appropriate series. then relative to evidence h \(h^{\prime}\) (assuming these to be not incompratible). it may be the case that no determinate series is indicated as appropriate. In this (ase the method of statistical induction fails us as a means of determining the probability under inquiry.
6. We can now remove our attention from the individual instance \(a\) to the properties of the series S . What sort of evidence is capable of justifying the conclusion that \(\mu\) is the probability that a random member of the series \(S\) will have the characteristic \(\phi\) ?

In the simplest case. is is a finite series of which we know the truth frequency for the characteristic \(\phi\), namely \(f\). \({ }^{1}\) Then by a straightforward application of the Principle of Indifference we have \(p=f\), so that \(\phi(x) / \mathrm{R}(x, \mathrm{~S}, \phi, h) . h=f\).

In another important type \(S\) is a series, with an indefinite number of members which, however, group themselves in such a way that for every member of which \(\phi(x)\) is true. there corresponds a determinate number of members of which \(\phi(. r)\) is false. The series, that is to say, contains an indefinite number of atoms, but each atom is made up of a set of molecules of which \(\phi(, r)\) is true and false respectively in fixed and determinate proportions. If this determinate proportion is known to be \(f\), we have, as before, \(\mu-f\). The typical instance of this type is afforded by games of chance. Every possible state of affairs which might lead to a divergence in one direction is balanced by another probability leading in the opposite direction ; and these alternative possibibilites are of a kind to which the Principle of Indifference is applicable. Thus for every poise of the dice box which leads to the fall of the six-face, there is a corresponding poise which leads to the fall of each of the other faces: so that if \(S\) is the series of possible poises, we may equate \(p\) to \(\frac{1}{6}\) where \(\phi\) is the fall of the six-face. It is not necessary. in order to ohtain this remult, th assert that is is a finite series with an actual determinate frequency \(f\) for the fall of each face.
sofar no inductive element enters in. But in general we do not know the constitution of S for certain, and can only infer it inductively from its rememblance to other series of which we know the constitution. This presents a nornal inductive problem-

\footnotetext{

}
the determination by an analysis of the positive and negative analogies as to whether the respects in which S differs or may differ from the other series is or is not relevant in the particular context \(\phi\); and it involves the same sort of considerations as those discussed in Part III.

There is, however, a further difficulty to be introduced before we have reached the typical statistical problem. In the case now to be considered our actual data do not consist of positive knowledge of the constitutions either of S itself or of other series more or less resembling S, but only of the frequency of the characteristic in actually observed sets of selections, great or small, either from S itself or from other series more or less resembling S.

Thus in the most general case our inquiry falls into two parts. We are given the observed frequency in statistical sets selected from \(S_{1}, S_{2}\), etc., respectively. The first part of our inquiry is the problem of arguing from these observed frequencies to the probable constitutions of \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., i.e. of determining the values of \(\phi(x) / \mathrm{R}\left(x, \mathrm{~S}_{1}, \phi, h\right) . h\), etc.; we may call this part the statistical problem. The second part of our inquiry is the problem of arguing from the probable constitutions of \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., to the probable constitution of S , where \(\mathrm{S}, \mathrm{S}_{1}, \mathrm{~S}_{2}\) resemble one another more or less, and we have to determine whether the differences are or are not relevant to our inquiry; we may call this part the inductive problem.

Now if the observed statistical sets are made up of random instances of \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., we can argue in certain conditions from the observed frequencies to the probable constitutions of the series, out of which the random selections have been made, by an inverse application of Bernoulli's Theorem on the lines explained in Chapter XXXI. Moreover, if the series \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., are finite series and the observed selections cover a great part of their members, we can reach an at least approximate conclusion without raising all the theoretical difficulties or satisfying all the conditions of Chapter XXXI. The commonly received opinions as to the bearing of the observed frequencies in a random sample on the constitution of the universe out of which the sample is drawn, though generally stated too precisely and without sufficient insistence on the assumptions they involve. our actual evidence not warranting in general more than an
approximate result, are not. I think, fundamentally erroneons. The most usual error in modern method consists in treating too lightly what 1 liave termed above the inductice problem. i.e. the problem of passing from the series \(S_{1}, \mathcal{S}_{2}\). etce, of which we have observed samples. to the series is of which we have not observed samples.

Let us, then, assume that we have ascertained \(p_{1}, p_{2}\), etc., with more or less exactness, hy examining either all the instances of the series \(S_{1}, S_{2}\), ete., or random selections from them, i.e. \(\phi(. r)_{i} \mathrm{R}\) \(\left(x, S_{1}, \phi, h\right) \cdot h=\mu_{1}\), etc. This can be expressed for short by saying that the series \(s_{1}\), \(s_{2}\), etc., are subject to probable-frequencies \(p_{1} \cdot P_{2}\). \(+1 c_{c}\)., for the characteristic \(\phi\). Our problem is to infer from this the probable-frequency \(p\) of the unexamined seriess. The class characteristics of the series \(S_{1}, \stackrel{S}{2}_{2}\), etc., will be partly the same and partly different. Using the terminology of Part III. we may term the class characteristics which are commen to all of them the Positive Analogy, and the class characteristics which are not cominon to all of them the Negative Analogy.

Now, if the , heserved or inferred probable-frequencies of the series \(S_{1}, S_{2}\) are to form the basis of a statistical induction, they must -how a stuble value; that is to say, either we must have \(\rho_{1}-\mu_{2}=\) etc... or at least \(\mu_{1}, \mu_{2}\), etce, must be stably grouped about their mean value. Our next task, therefore, must be to discover whether the probable-frequencies \(\mu_{1}\), \(p_{2}\). etce, display a significant stability. It is the great merit of Lexis that he was Whe lirst to investigate the problem of stability and to attempt its measurement. For, until a primâ fucie case has been established for the existence of a stable probable-frequency, we have hot a flimsy basi for any statistical induction at all ; indeed we ate limited to the class of case where the instance under inquiry is a member of identically the sume series as that from which our -amples were drawn, i.e. where \(S=S_{1}^{\prime}\), which in social and scientific inquiries is seldom the case.

What is the meaning of the assertion that \(p_{1}, p_{2}\), etc., are stably grouped about their mean value? The answer is mot simple and not perfectly precise. We could propound varions formake for the measurement of statility and disperion, respectisely, and the problem of translating the conception of mability: which is mot quantitatively precise, into a numerical formula involves an arhtrars or approximative element. For practical
purposes, however, I doubt if it is possible to improve on lexis's measure of stability \(Q\), the mathematical definition of which has been given above on p. 399. Lexis describes the stability as subnormal, normal, or supernormal according as \(Q\) is less than, equal to, or greater than 1. This is too precise, and it is better perhaps to say that the stability about the mean is normal if the dispersion is such as would not be improbable \(a\) priori, if we had assumed that the members of \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., were obtained by random selection out of a single universe \(U\), that it is subnormal if the dispersion is less than one would have expected on the same hypothesis, and that it is supernormal if the dispersion is greater than one would have expected.

Let us suppose that we find that on this definition \(p_{1}, p_{2}\), etc., are stable about \(p\), and let us postpone consideration of the cases of subnormal or supernormal dispersion. This is equivalent to saying that the frequencies of \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., are within limits which we should expect \(a\) priori, if we had knowledge relative to which their members were chosen at random from a universe U of which the frequency was \(p\) for the characteristic under inquiry. We next seek to extend this result to the unexamined series \(S\) and to justify anticipations about it on the basis of the members of S also being chosen at random from the universe U. This leads us to the strictly inductive part of our inquiry.

The class characteristics of the several series \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., will be partly the same and partly different, those that are the same constituting the positive analogy and those that are different constituting the negative analogy, as stated above. The series S will share part of the positive analogy. The argument for assimilating the properties of S , in relation to the characteristic under inquiry, to the properties of \(\mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., in relation to this characteristic depends on the differences between \(\mathrm{S}, \mathrm{S}_{1}, \mathrm{~S}_{2}\), etc., being irrelevant in this particular connection. The method of strengthening this argument seems to me to be the same as the general inductive method discussed in Part III. and to present the same, but not greater, difficulties.

In general this inductive part of our inquiry will be best advanced by classifying the aggregate series of instances with which we are presented in such a way as to analyse most clearly the significant positive and negative analogies, to group them, that is to say, into sub-series \(S_{1}, S_{2}\). ete., which show the most
marked and definite class characteristics. Our knowledere of the differences between the particular observed instances which constitute our original data will suggest to us one or more principles of classification. such that the members of each sub)series all have in common some set of positive or nequative characteristics, not all of which are shared in common by all the members of any of the other sub-series. That is to say, we classify our whole set of instances into a series of series \(S_{1}\), \(S_{2}\), etce.. which have frequemeies \(f_{1}\). \(f_{2}\). etce.. for the characteristic under inquiry : and then again we classify them by another principle or criterion of classification into a seeond series of series \(\mathrm{S}_{1}{ }^{\prime}\). \(\mathrm{S}_{2}{ }^{\prime}\), ette.. with frequencies \(f_{1}^{\prime} \cdot f_{2}^{\prime}\).ettin and so on, so far as our knowledge of the presihle relevant differences between the instances extends: the whole result heing then summed up in a statement of the positive and nosative analogies of the series of series. If we then find that all the frequencies \(f_{1} \cdot f_{2}\), ette.. \(f_{1}^{\prime} \cdot f_{2}^{\prime}\). etce... are stable about a value \(\mu\), and if. on the basis of the above positive and negative analogies. we have a uomal inductive argument for assimilating the unexamined series St to the examined series \(S_{1}\). S. ette., \(\mathrm{S}_{1}{ }^{\prime}\), \(\mathrm{S}_{2}^{\prime}\), etce, in respect of the characteristic under inguiry, in this case we have not conclusive grounds, but grounds of some weight for asserting the probabilits \(p\), that an instance taken at random from \(S\) will have the characteristic in question.

Let me recapitulate the two essential stages of the argument. We first find that the observed frequencies in a set of series are such as would have been not improbable à priori if, relative to our knowledge, these series had all been made up of random members of the same universe U : and we next argue that the positive and negative analogies of this set of series furnish an inductive argument of some weight for supposing that a further unexamined series is resembles the former series in having a frequency for the characteristic under inguiry such as would have been not improbable \(\dot{a}\) primi if, relative to our knowlediee. St was also made up of random members of the hypothetical universe U .
7. It is sers perplexing to decide how far an argument of this character involves any new and theoretically distinct difficultics or assumptions, beyond those already admitted as inherent in Iniversal Induction. I believe that the foremoing analysis is along the right lines and that it carries the
inquiry a grood deal further than it has been carried hitherto. But it is not conclusive, and I must leave to others its more exact elucidation.

There is, however, a little more to be said about the half-felt reasons which, in my judgment, recommend to common sense some at least of the scientific (or semi-scientific) arguments which run along the above lines. In expressing these reasons I shall be content to use language which is not always as precise as it ought to be.

I gave in Chapter XXIV. §§ 7-9 an interpretation of what is meant by an 'objectively chance' occurrence, in the sense in which the results of a game, such as roulette, may be said to be governed by 'objective chance.' This interpretation was as follows: "An event is due to objective chance if in order to predict it, or to prefer it to alternatives, at present equi-probable, with any high degree of probability, it would be necessary to know a great many more facts of existence about it than we actually do know, and if the addition of a wide knowledge of general principles would be little use." The ideal instance of this is the game of chance ; but there are other examples afforded by science in which these conditions are fulfilled with more or less perfection. Now the field of statistical induction is the class of phenomena which are due to the combination of two sets of influences, one of them constant and the other liable to vary in accordance with the expectations of ohjective chance,--Quetelet's ' permanent causes' modified by 'accidental causes.' In social and physical statistics the ultimate alternatives are not as a rule so perfectly fixed, nor the selection from them so purely random, as in the ideal game of chance. But where, for example, we find stability in the statistics of crime, we could explain this by supposing that the population itself is stably constituted, that persons of different temperaments are alive in proportions more or less the same from year to year, that the motives for crime are similar, and that those who come to be influenced by these motives are selected from the population at large in the same kind of way. Thus we have stable causes at work leading to the several alternatives in fixed proportions, and these are modified by random influences. Generally speaking, for large classes of social statistics we have a more or less stable population including different kinds of persons in certain proportions and on the other
hand sets of enviromments; the proportions of the different kinds of persons, the propertions of the different kinds of environments, and the manner of allotting the enviromments to the persons vary in a ramlom mamer from year to year (or, it may be, from district to district). In all such cases as these. however. prediction heroud what has been observed is clearly open to sources of error which can be neglected in comsidering, for example. games of thance;-our so-called 'permanent ' caluses are always chameins is little and are liable at any moment to radical alteration.

Thas the more als.s. that we find the conditions in scientific examples assimilated to those in games of chance, the more antidently dues common sense recommend this method. The rather surprisine frequencr with which we find apparent stability in human statiotio may possibly be explainet, therefore, if the binhwical heory of Mendelism can he established. According to this theors the qualities apparent in any generation of a given race appar in proprortions which are determined by methods very closely analogous to those of a game of chance. To take a sprecific example (I am giving not the correct theory of sex but an artificially simplifion form of it), suppose there are two kinds of spermatozoa and two kinds of ova and of the four possible kinds of union two produce males and two femates, then if the kinds of spermatozo and ora exist in equal numbers and their union is determined by random considerations in precisely the same sensw in which a game of chance such as roulette depends upon random comsiderations, we should expect the observed proportions to vary from equality, as indeed they do, in the same manner as variations Cona equality of red and hark onecur at roulette.' If the sphere of influmen of Mendelian considerations is wide. We haw both an explanation in part of what we observe and also : laver opmertunity in future of using with profit the methods of statistical analysis.

This is all familiar. This is the way in which in fact we do think and argue. The inquiry as to how far it is covered by the abo-dact analyoin of the preceding pragraphs. and he what

\footnotetext{
I The fluctuations in the: proportion of the sexes which, as is well known, is mot in fact one of equality, correspond, as Lexis has shown, to what one Would expect in a game of chance with an astonishing exactitude. But it is difficult to find any other example, amongst natural or social phenomena, in whith his criteria of stability are by any means as equally well satisfied.
}
logical principle the use of this analysis can be justified as rational, I have pushed as far as I can. It deserves a profounder study than logicians have given it in the past.
8. Two subsidiary questions remain to be mentioned. The first of these relates to the character of series which, in the terminology of Lexis, show a subnormal or supernormal stability ; for I have pressed on to the conclusion of the argument on the assumption that the stabilities are normal. Subnormal stability conceals two types: the one in which there is really no stability at all and the results are in fact chaotic ; and the other in which there is mutual dependence between the successive instances of such a kind that they tend to resemble one another so that any divergence from the normal tends to accentuate itself. Supernormal stability corresponds in the other direction to the second of these two types ; that is to say, there is mutual dependence of a regulative kind between the successive instances which tends to prevent the frequency from swinging away from its mean value. The case, where the dog was fed with scraps when he looked thin and not fed when he looked fat, illustrated this. The typical example of this type is where balls are drawn from urns, containing black and white balls in certain jroportions and not replaced ; so that every time a black ball is drawn the next ball is more likely than before to be white, and there is a tendency to redress any excess of either colour beyond the proper proportions. Possibly the aggregate annual rainfall may afford a further illustration.

Where there is no stability at all and the frequencies are chantic, the resulting series can be described as 'non-statistical.' Amongst 'statistical series,' we may term 'independent series' those of which the instances are independent and the stability normal, and 'organic series," those of which the instances are mutually dependent and the stability abnormal, whether in excess or in defect. 'Organic series' have been incidentally discussed elsewhere in this volume. I shall not pursue them further now, because I do not think that they introduce any new theoretical difficulty into the general problem of statistical inference; although the problem of fitting them into the general theoretical scheme is not easy. \({ }^{1}\)

\footnotetext{
1 The following more precise definitions bring these ideas into line with what has gone before : consider the terms \(a_{1}, a_{2} \ldots a_{n}\) of a series \(s(x)\); let ' \(a_{r}\) is \(g\) '
}
9. The seeond question is concerned with the relation between the Inductive Correlation. which has been the subject-matter of this chapter, and the Correlation Coeflicient, or, as I should prefer to call it, the Qumbitatire ('orrelation, with which recent English statistical theory has chiefly occupied itself. I do not propose to discuss this theory in detail. because I suspect that it is much more concerned. at any rate in its present form, with statistical description than with statistical induction. The transition from defining the 'correlation coeflicient' as an algehraical expression to its employment for purposes of inference is very far from clear even in the work of the best and most systematic writeron the subject. surth as Mr. Yule and Professor Bowley.

In the motation cmployed in the carlier part of this chapter I have classified carh examined instance " according as it did or did mot possess the characteristic \(\phi\). i.e. satisfy the propersitional function \(\phi(x)\), wr. in wher words, aceording as \(\phi(a)\) was true or false. Thu only two possible alternatives were contemplated, and \(\phi\) was mot comsidered as a quantitative characteristic which the instance could satisfy in greater or less degree. Equally the common element in all the instances, required to constitute them as instances for the purpose of our statistical generalisation (or. as I havesometimes put it. required to satisfy the comdition of the generalisation). was regarded as definite and unique and not capable of quantitative variation. That is to say, all the instances satisfied a function \(\psi(\).\() , and the question was, what propertion\)

\footnotetext{
\(\equiv g_{r}\) and let \(g_{r}{ }^{\prime} h \quad p_{r}\), where \(h\) is our data. Then, if \(g_{r} / g_{g} \ldots g_{t} \ldots h-p_{r}\) for all values of \(r, s, \ldots, t \ldots\), the terms of the series are independent relative to \(h\). If \(p_{1}-p_{z}=\ldots=p\) the terms are uniform. If the terms are both independent and uniform, the series may be called an independent Bernoullian series, subject to a Bernoullian probability p. If the terms are independent but not uniform, the series may be called an independent compound series, subject to a compounded probubility \(1 / n こ p_{p}\). If the terms are not independent, the series is an organic s.ries

The same terminology can then be applied to the series \(\mathrm{S}_{1}, \mathrm{~S}_{2}, \ldots \mathrm{~S}_{n}\), regarided as members of the series of series \(\mathrm{S}(x)\). Let the frequencies of the series for the
 probability of a frequency \(x_{1}\) in the first series. Then if \(x_{r} x_{p} \ldots h-\theta_{r}\left(x_{r}\right)\) for all
 the frequencies are slable. If the frequencies are stable and independent, the series of series may be called Giaussian. If the frequencies are stable and independent, and if in addition each individual series is subject to a Bernoullian probability, the probable dispersion of the frequeney is normal and symmetrical. If the individual series are organic, the dispersion of the frequencies may be normal, submormal, or supernormal. If the series of series is Gaussian, nad the individual series Bernoullian, we have the type of the perfect statistical series
}
of them also satisfied the function \(\phi(x)\). A typical example was that of sex-ratio \(-\psi(x)\) being the birth of a child and \(\phi(x)\) its sex, where there is no question of degree in either \(\psi(x)\) or \(\phi(x)\).

It might be the case, however, that the characteristics under examination were capable of degree or quantitative variation ; for example \(\psi(x)\) might be the age of the mother and \(\phi(x)\) the weight of the child at birth. In this case we should have a series \(\psi_{1}(x), \psi_{2}(x)\), etc., corresponding to the various age-periods of the mothers, and a series \(\phi_{1}(x), \phi_{2}(x)\), ete., corresponding to the various weights of the children. Now if we concentrated our attention on \(\psi_{1}(x)\) and \(\phi_{1}(x)\) alone, i.e. on mothers of a particular age and the proportions of their children which had a particular weight at birth, we have a one-dimensional problem of the same lind as before ; out of all the instances which satisfy \(\psi_{1}^{\prime}(x)\) a certain proportion satisfy \(\phi_{1}(x)\) also. But clearly we can push our observations further and we can take note what proportion of the instances which satisfy \(\psi_{1}(x)\) satisfy \(\phi_{2}(x), \phi_{3}(x)\), and so on, respectively ; and then we can do the same as regards the instances which satisfy \(\psi_{2}(x), \psi_{3}(x)\), etc. The total results of this twodimensional set of observations can then be tabulated in what is called a twofold correlation table. Thus if \(f_{r s}\) is the proportion of instances satisfying \(\psi_{s}(x)\) which also satisfy \(\phi_{r}(x)\) we have a table as follows :


We could, further, increase the complexity and completeness of our observations to any required degree. For example we might take account also of \(\theta(x)\), the age of the father, and construct a threefold table where \(f_{\text {rst }}\) is the proportion of instances satisfying \(\phi_{r}(x), \psi_{s}(\cdot), \theta_{t}(x)\); and so on up to an \(n\)-fold table.

Clearly it is not necessary for the construction of tables of
this kind that \(\phi(r)\) and \(\psi(r)\) should stand for degrees of the same quantitative characteristie: they might lee any set of exclusive alternatives ; for example. \(\psi(t)\) might be the colour of the baby's eyes, and \(\phi(x)\) its Christian name.

But in order that the correlation table may be of any practical interest for the purposes of inference, it is necessaryand this, I think, is one of the critical assumptions of correlation - that \(\phi_{1}(.),, \phi_{2}(t)\). . . and also \(\phi_{1}(r), \phi_{2}(r)\). . . should be arranged in an order that is significent, i.e. such that we have some à priori reason for expecting some connection to exist between the order of the \(\phi\) ss and the order of the \(\phi\) s. The peint of this will be illustrated by concentrating our attention on the
slight one. for supposing that there might be some comectom between ther are of the mother and the weight of the babe, then. if in a particular set of instances the frequemeies were grouped about the diamonal as sugrested above this might he taken as affording some inductive support for the hypothesis.

Now the theory of correlation, as it is expounded in the text-homks, is almost entirely comeerned with meaturing how nearly the ohserved frequencies are grouned about the diagemal of the table (though the complete theory is not, of course, so restrictedas this). The 'conefficient of correlation' is an alesedraical formula which may he rewarded as measuring thi phemomenom in a way that is sulficiontly satisfactore for all ondinary purpeses. If it is defined thus, it is simply a statistical description of a particular ont of ohservations arransed in a particular order. How can we make use of this coefficient for the purposes of inference?

Dr. Bowley faces this problem a little more definitely than do
of them also satisfied the function \(\phi(x)\). A typical example was that of sex-ratio,- \(\psi(x)\) being the birth of a child and \(\phi(x)\) its sex, where there is no question of degree in either \(\psi(x)\) or \(\phi(x)\).

It might be the case, however, that the characteristics under examination were capable of degree or quantitative variation ; for example \(\psi(x)\) might be the age of the mother and \(\phi(x)\) the weight of the child at birth. In this case we should have a series \(\psi_{1}(x), \psi_{2}(x)\), etc., corresponding to the various age-periods of the mothers, and a series \(\phi_{1}(x), \phi_{2}(x)\), ete., corresponding to the various weights of the children. Now if we concentrated our attention on \(\psi_{1}(x)\) and \(\phi_{1}(x)\) alone, i.e. on mothers of a particular age and the proportions of their children which had a particular weight

\author{
ERRATA
}

Page 423, 1. 8. For the first-mentioned \(\phi_{1}, \phi_{2}\) read \(\psi_{1}, \psi_{2}\).
1. 11. For the first-mentioned \(\phi\) read \(\psi\).
1. 13. For the first-mentioned \(\phi\) read \(\psi\).
1. 16. For \(\phi_{1}, \phi_{2}\) read \(\psi_{1}, \psi_{2}\).
lalle as sullumi.


We could, further, increase the complexity and completeness of our observations to any required degree. For example we might take account also of \(\theta(x)\), the age of the father, and construct a threefold table where \(f_{\text {rst }}\) is the proportion of instances satisfying \(\phi_{r}(x), \psi_{s}(\cdot), \theta_{t}(x)\); and so on up to an \(n\)-fold table.

Clearly it is not necessary for the construction of talles of
this kind that \(\phi(r)\) and \(\psi(x)\) should stand for degrees of the same quantitative characteristic: they might lee any set of exclusive alternatives ; for example. \(\psi(1)\) might be the colour of the baby's eyes, and \(\phi(x)\) its Christian name.

But in order that the correlation table may be of any practical interest for the purposes of inference, it is necessaryand this. I think, is one of the critical assumptions of correla-tion-that \(\phi_{1}(t), \phi_{2}(t)\). . . and also \(\phi_{1}(t), \phi_{2}(t)\). . . shoult be arranged in an order that is significont. i.e. such that we have some à priori reason for expecting some connection to exist bee ween the order of the \(\phi\) s and the order of the \(\phi\) s. The peint of this will her illustrated ley concentrating our attention on the simplest type of case where \(\phi(r)\) and \(\phi(x)\) are quantitative characteri-tirs aranged in order of magnitude. Now suppose it wen the rase that the romener mothers tended to bear heavier hahios. then if \(\phi_{1}(r) \phi_{2}(\cdot)\) are the ages increasing upwards and \(\phi_{1}(1) \phi_{2}(.1)\) the werights diminishing dewnwards. \(f_{11}\) would probably be the ereatest of the \(f_{r 1}\) 's and. qenerally speraking. \(f_{r 1}\) would he greater than \(f_{r+1,1}\) : alow. \(f_{22}\) might he the greatest of the \(f_{r 2} s\) and so on ; so that the frequencies lying on the diagonal of the table would be the greatest and the frequencies would tend to be less the farther they lay from the diagonal. If we had some reason à priori (i.e. based on our pre-existing knowledge), if only a slight one. for supposing that there might be some comection between the age of the mother and the weight of the baby, then, if in a particular set of instances the frequencies were grouped about the diagonal as sugrested above this might he taken as affording some inductive support for the hypothesis.

Now the theory of correlation, as it is expounded in the text-homks, is almost emtiely concerned with measuring how nearly the oboerved frequencies are grouned about the diagenal of the table (thenugh the complete theory is not, of comeres. - w restricted as hisis). The 'conefficient of correlation' is an alqebraical formula which may be regarded as measuring this phemomenon in a way that is sufficiontly satisfactory for all ordinary purposes. If it is defined thus. it is simple a statiotical descoption of a particular onf of ohservations arraned in a particular order. How can we make use of this coefficient for the purposes of inferanm:

Dr. Bowley faces this problem a little more definitely than do
most statistical writers. Mr. Yule warns the student that the problem exists, \({ }^{1}\) but he does not himself attack it systematically or do more than apply common sense to particular problems. So much greater emphasis, however, has been laid hitherto on the mathematical complications, that many statistical students hazily float from defining the correlation coefficient as a statistical description to emploving it as a measure of the probability of a statistical generalisation as to the association between quantitative variations of \(\phi(x)\) and \(\psi(x)\) respectively. If, for example, it is found in a particular set of observations of mothers' ages and babies' weights that the frequencies are closely ranged about the diagonal, this is considered a sufficiently good reason for attributing probability to a generalisation as to the 'correlation' (i.e. tendency to quantitative correspondence) between the age of the mother and the weight of the baby.

Dr. Bowley's line of thought is as follows. He begins by defining the correlation coefficient \(r\) merely as a statistical description (Elements of Statistics, p. 354). He then shows (p. 355), as an illustration of the nature of \(r\), that if \(x\) and \(y\) are two variable quantities which depend (more strictly, are linown to depend) on other variables \(\mathrm{U}, \mathrm{V}, \mathrm{W}\) in such a way that
\[
\begin{aligned}
& \mathrm{X}_{t}={ }_{1} \mathrm{U}_{t}+{ }_{2} \mathrm{U}_{t}+\ldots+{ }_{p} \mathrm{U}_{t}+{ }_{1} \mathrm{~V}_{t}+{ }_{2} \mathrm{~V}_{t}+\ldots+{ }_{q} \mathrm{~V}_{t}+ \\
& \mathrm{Y}_{t}={ }_{1} \mathrm{U}_{t}+{ }_{2} \mathrm{U}_{t}+\ldots+{ }_{p} \mathrm{U}_{t}+{ }_{1} \mathrm{~W}_{t}+{ }_{2} \mathrm{~W}_{t}+\ldots{ }_{q}+\mathrm{W}_{t}
\end{aligned}
\]
where \({ }_{1} \mathrm{U}_{t},{ }_{2} \mathrm{U}_{t} \ldots{ }_{1} \mathrm{~V}_{t},{ }_{2} \mathrm{~V}_{t} \ldots{ }_{1} \mathrm{~W}_{t},{ }_{2} \mathrm{~W}_{t} \ldots\) are selected at random each from an independent group of quantities (more strictly, are relative to our data, random members of independent groups) ; then, if we know à priori certain statistical coefficients descriptive of the constitution of these groups, the value of \(r\) will probably tend towards a certain value. So far we are on fairly safe, but not very fruitful, ground. We have no basis for arguing backwards from the observed value of \(r\); but, provided we have rather extensive and peculiar knowledge à priori as to how \(\mathrm{X}_{t}\) and \(\mathrm{Y}_{t}\) are constituted, then we have calculable expectations as to the limits within which the value

\footnotetext{
1 Introduction to the Theory of Statistics, p. 191: "The coefficient of correlation, like an average or a measure of dispersion, only exhibits in a summary and comprehensible form one particular aspect of the facts on which it is based, and the real difficulties arise in the interpretation of the coefficient when obtained."
}
of \(r\), namely the correlation coeflicient between X and Y . will probably turn out to lie, when we have observed it.

Dr. Bowley's next move is more dubious. If the constitutions of the independent groups are similar in a certain statistical respect (ife if the hate the same standard deviations), then, Dr. Bowley concludes, \(r=\frac{\eta}{\mu+q}\), which "expressed in words shows that the correlation coneflicient tends to be the ratio of the number of causes common in the genesis of two variables to the whole number of independent causes on which each depends." By this time the student's mind. unless anchored hy a more than ordinary scepticism, will have been well launched into a vague, fallacious sea.

Nerlecting. however, the diefmon just quoted, we find that the secomd stage of the argument consists in showing that. if we hate a cortain sor of kowlodge a priori as to how our variables are constituted, then the various possible values for the coefficients of correbation, which would be yielded by actual sets of observations made in preseribed comditions. will have. à priori, and hefore the observations have been made, calculable probabilities. certain ranes of values being probable and others improbable.

As a rule. however, we are not arguing from knowleder about the variables to anticipations about their correlation coefticient: but the other way round, that is from ohservations of their correlation coefficients to theories about the nature of the variables. Dr. Bowley percepives that this involves a third stage of the argument, and appeals accordingly (p. 409) to "the diflicult and elusive theory of inveree probability." He apprehends the difficulty but he does not pursue it; and, like Mr. Yule. her really falls hack for practical purpeses on the criteria of common sense, an expedient well enough in his case, but not a universal safeguard.

The general argument from inverse probability to which 1)r. Bowley makes his vague appeal is doubtless on the following lines: If there is no causal connection between the two sets of guantitios then a flowe erompinge of the frequencies about the diamonal would he is primi improbathe (and the serater the number of the indicidual whervations. the ereater the improha bility since, if the quantities are independent, there is, then, all the more opportunity for 'averaging out') ; therefore, inversely,
if the frequencies do group themselves about the diagonal, we have a presumption in favour of a causal connection between the two sets of quantities.

But if the reader recalls our discussion of the principle of inverse probability, he will remember that this conclusion cannot be reached unless a priori, and quite apart from the observations in question, we have some reason for thinking that there may be such a causal connection between the quantities. The argument can only strengthen a pre-existing presumption; it cannot create one. And in the absence of reasons peculiar to the particular inquiry, we have no choice but to fall back on the general methods and the general presumptions of induction.

It is apparent that, where the correlation argument seems plausible, some tacit assumption must have slipped in, if we return to the case where our correlation table relates to the weights of the babies and their Christian names. Either by accident or because we had arranged the order of the Christian names to suit, it might happen with a particular set of observations, even a fairly numerous set, that the correlation coefficient was large. Yet on that evidence alone we should hardly assert a gencralisation connecting the weights of babies with their Christian names.

The truth is that sensible investigators only employ the correlation coefficient to test or confirm conclusions at which they have arrived on other grounds. But that does not validate the crude way in which the argument is sometimes presented, or prevent it from misleading the unwary,- since not all investigators are sensible.

If we abandon the method of inverse probability in favour of the less precise but better founded processes of induction, 'quantitative correlation,' as I should like to term this particular branch of statistical induction, is more complicated than, but not theoretically distinct from, the kind of arguments which have occupied the earlier paragraphs of this chapter. The character of the additional complication can be dess ribed by saying that we are presented with a two-dimensional problem instead of a one-dimensional problem. The mere existence of a particular correlation coefficient as descriptive of a group of observations, even of a large group, is not in itself a more conclusive or significant argument than the mere existence of a particular frequency coetlicient would be. Of course if we have a considerable body
of pre-existing knowledge relevant to the particular inquiry, the calculation of a small number of correlation coefficients may bee crucial. But otherwise we must proceed as in the case of frequency conflicients; that is to say we must have before us. in order to found a satisfachory argument, many sets of ohservations. of which the correlation coefficients display a significant stability in the midst of variation in the non-essential class characteriatios (i.e. those class characterintics which our wenera!isation propuses to neglect) of the different sets of ohsiservations.
10. I am now at the conclusion of an inquiry in which, berimnine with fundamentalquestions of logie. I have endeavonred to pusio fonward to the analysis of some of the actual arguments which imperess us as rational in the progress of knowledge and the practice of empirical science. In writing a book of this kind the author mut. if he is to put his point of view clearly, pretend sometimes to a little more conviction than he feels. He must give his own argument a chance, so to speak, nor be too ready to depres its ritality with a wet choud of douht. It is a heavy task to write on these problems; and the reader will perhaps excuse me if I have sometimes pressed on a little faster than the difficalties were osercome and with decidedly more contidence than I have always felt.

In laying the fommations of the subject of Probahility, I have departed a wool deal from the conception of it which governed the minds of Laplace and Quetelet and has dominated through their influene the thenolit of the past century, though I beliewe that Leibniz and Hume might have read what I have writem with sympathe. But in taking leave of l'robability. I should like to say that, in my judement. the practical usefulness of these modes of inference, here termed Unisersal and Statistical Induction. on the validity of whid the beasted knewledge of modern sciene depends, can only exist and I do not now pause to inquire again whether surf an argument must he circular if the uniwerse of phenomoma dew- in fact present these pecentiar characteristios of atomism and limited variety which appear more and more clearly as the ultimate result to which material arimen is tending

\section*{fateare necessest}
materiem quoqu: fimitis differe figuri-.
The physicists of the minementh contury have reduced matter to
the collisions and arrangements of particles, between which the ultimate qualitative differences are very few; and the Mendelian biologists are deriving the various qualities of men from the collisions and arrangements of chromosomes. In both cases the analogy with the perfect game of chance is really present ; and the validity of some current modes of inference may depend on the assumption that it is to material of this kind that we are applying them. Here, though I have complained sometimes at their want of logic, I am in fundamental sympathy with the deep underlying conceptions of the statistical theory of the day. If the contemporary doctrines of Biology and Physics remain tenable, we may have a remarkable, if undeserved, justification of some of the methods of the traditional Calculus of Probabilities. Professors of probability have been often and justly derided for arguing as if nature were an urn containing black and white balls in fixed proportions. Quetelet once declared in so many words- " l'urne que nous interrogeons, c'est la nature." But again in the history of science the methods of astrology may prove useful to the astronomer ; and it may turn out to be true-reversing Quetelet's expression-that " La nature que nous interrogens, c'est une urne."

BIBLIOGRAPHY

\title{
BIBLIOGRAPHY
}

\author{
INTRODUCTION
}

There is no opinion, however absurd or incredible, which has not been maintained by some one of our philosophers.-Descartes.

The following Bibliography does not pretend to be conplete. but it contains at much longer list of what has been written about Probability than can be found elsewhere. I have hesitated a little hefore burdening this volume with the titles of many works, so few of which are still valuable. But I was myself much hampered, when first I embarked on the study of this subject, the the absence of guide-posts to the scattered but extensive literature of the subject ; and a list which I drew up for my own convenience, without much attention to bibliographical nicety or to exact uniformity in the style of entry, may be useful to others.

It is rather an arbitrary matter to decide what to include and what to exclude. Probability onerlaps many other topics. and some of the most important references to it are to be found in bools, the main topic of which is something else. On the other hand it would be absurd to include every casual reference : and no useful purpose would hase been served by ratalnguing the very numerous volumes dealing with Insurance. (iames of (hance, Statistics. Errors of (h)wervation, and Loast Squares, which treat in detail thees various applications of the Theors of I'robability. It has been a matter of some difliculty, therefore, to know precisely where to draw the line. Where the main subject of a hook or papere is Probabilit! preper. I have included it, nearly regardless of my own view as to its importance, and have not attempted to act as censor; but where Probalility is mot the main subject or where an application of I'robatilite is concerned, the chief interest of which is
solely in the application itself, I have only included the entry where I think it important, intrinsically or historically or from the celebrity of the author. In particular, the existence of Professor Mansfield Merriman's very extensive bibliography, published in the Transuctions of the Connecticut Academy for 1877, has made it possible to deal very lightly (and to the extent of but few entries) with the inordinately large literature of Least Squares. This list comprises 408 titles of writings relating to the Method of Least Squares and the theory of accidental errors of observation, and is sufficiently exhaustive so far as relates to memoirs on this topic published before 1877.

Of bibliographical sources for Probability proper, Todhunter's History of the Mathematical Theory of Probability and Laurent's Calcul des probabilités are alone important. Of mathematical works published hefore the time of Laplace, Todhunter's list, and also his commentary and analysis, are complete and exact,-a work of true learning, beyond criticism. The bibliographical catalogue at the conclusion of Laurent's Calcul (published in 1873) is the longest list published hitherto of general works on Probability. But it is unduly swollen by the inclusion of numerous items on Insurance and Errors of Observation, the bearing of which on Probability is very slight; \({ }^{1}\) it is chiefly mathematical in bias; and it is now nearly fifty years old.

I have not read all these books inyself, but I have read more of them than it would be good for any one to read again. There are here enumerated many dead treatises and ghostly memoirs. The list is too long, and I have not always successfully resisted the impulse to add to it in the spirit of a collector. There are not above a hundred of these which it would be worth while to preserve,-if only it were securely. ascertained which these hundred are. At present a bibliographer takes pride in numerous entries; but he would be a more useful fellow, and the labours of research would be lightened, if he could practise deletion and bring into existence an accredited Inder Expmigutorius. But this can only be accomplished by the slow mills of the collective judgment of

\footnotetext{
\({ }^{1}\) Laurent's list contains 310 titles, of which I have excluded 174 from my list as being insufficiently relevant.
}
the learned: and I have already indicated my own favourite authors in copious footnotes to the main body of the text.

The list is long: yet there is, perhaps, no, subject of equal importance and of equal fascination to men's minds on which so little has been written. It is mow rifty-five years since Dr. Venn, still an accustomed figure in the streets and courts of C'ambridge, first published his Loogic of C'lumene : yet amongst systematic works in the English language on the logical foundations of Probalility my Treatise is next to his in chronological order.

The student will find many fanous names here recorded. The sulject hats preserved its inystery, and has thus attracted the motice profound or, more often. casual, of most speculative minds. Leibniz, Pascal, Arnath, Hureres. Spinoza, Jacques and Daniel Bernoulli, Hume, 1)'. Nembert, Condorcet, Euler. Laplace. Poisson. Cournot, Quetelet, Ciauss, Mill, Boole. Tchehschef, Levis, and Poincaré, to name those only who are dead, are catalogued below.

Abbotт, T. K. "On the Probability of Testimony and Arguments." Phil. Mag. (4), vol. 27, 1864.
Adrain, R. "Research concerning the Probabilities of the Errors which happen in making Observations." The Analyst or Math. Museum, vol. 1, pp. 93-109, 1808.
[This paper, which contains the first deduction of the normal law of error, was partly reprinted by Abbe with historical notes in Amer. Journ. Sci. vol. i. pp. 411-415, 1871.]
Ammon, O. "Some Social Applications of the Doctrine of Probability." Journ. Pol. Econ. vol. 7, 1899.


 Berlin, pp. 3-32, 179+5.
Abbuthnot, J. Of the Laws of (hance, or a Method of calculation of the Hazards of Game plainly Demonstrated. 16 mo . London, 1692.
[Contains a translation of Huygens, De ratiociniis in ludo aleae.]
4th edition revised by John Hans. By whom is added a demonstration of the gain of the banker in any circumstance of the game call'd Pharaon, ete. Sim. 8vo. London, 1738.
[For a full account of this book and discussion of the authorship, see Todhunter's History, pp. 4s.73.]
"An Argument for Divine Providence, taken from the constant Rogular-
 \(190(1710-12)\).
[Argucs that the excess of male births is so invariable, that we may conclude that it is nut an even chance whether a male or female be born.]

Aristotle. Anal. Prior. ii. 27, \(70^{a} 3\).
Rhet. i. 2, 1357 a 34. [See Zeller's Aristotle for further references.]
Arnauld. (The Port Royal Logic.) La Logique ou l'Art de penser. 12 mo . Paris, 1662. Another ed. C. Jourdain, Hachette, 1846. Transl. into Eng. with introduction by T. S. Baynes. London, 1851. xlvii +430. See especially pp. 351-370.

Babbage, C. An Examination of some Questions connected with Cames of Chance. 4to. 25 pp. Trans. R. Soc. Edin., 1820.
Bachelier, Louis. Calcul des probabilités. Tome i. 4to. Pp. vii +517. Paris, 1912.

Le Jeu, la chance, et le hasard. Pp. 320. Paris, 1914.
[Bailey, Samuel.] Essays on the pursuit of truth, on the progress of knowledge and on the fundamental principle of all evidence and expectation. Pp. xii +302 . London, 1829.
Baldwin. Dictionary of Philosophy. Bibliographical volumes; s.v. "Probability."
Baniol, A. "Le Hasard." Revue Internationale de Sociologie. Pp. 16. 1912.

Barbeyrac. Traité du jeu. 1st ed. 1709. 2nd ed. 1744.
[Todhunter states (p. 196) that Barbeyrac is said to have published a discourse "Sur la nature du sort."]
Bayes, Thomas. An Essay towards solving a Problem in the Doctrine of Chances. Phil. Trans. vol. liii. pp. 370-418, 1763. A demonstration, etc. Phil. Trans. vol. liv. pp. 296-325, 1764.
[Both the above were communicated by the Rev. Richard Price, and the second is partly due to him.]

German transl. Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung. Herausgegeben von H. E. Timerding. Sm. 8vo. Leipzig, 1908. Pp. 57.
Bequelin. "Sur les suites ou séquences dans le loterie de Gênes." Hist. de l'Acad. Pp. 231-280. Berlin, 1765.
"Sur l'usage du principe de la raison suffisante dans le calcul des probabilités." Hist. de l'Acad. Pp. 382-412. Berlin, 1767. (Publ. 1769.)
Bellavitis. "Osservazioni sulla theoria delle probabilità." Atti del Instituto Veneto di Scienze, Lettere, ed Arti, Venice, 1857.
Benard. "Note sur une question de probabilités." Journal de l'École royale politechnique. Vol. 15, Paris, 1855.
Bentham, J. Rationale of Judicial Evidence.
See Introductory View, chap. xii., and Bk. i. chaps. v., vi., vii.
Bernoulli, Daniel. "Specimen theoriae novae de mensura sortis." Comm. Acad. Sci. Imp. Pet. vol. v. pp. 175-192, 1738.

Germ. transl. 1896, by A. Pringsheim: Die Grundlage der modernen Wertlehre. Versuch einer neuen Theoric der Wertbestimmung von Glücksfällen (Einleitung von Ludvig Fick). Pp. 60. Leipzig, 1896.
"Recueil des pièces qui ont remporté le prix de l'Académie Royale des Sciences." 1734. iii. pp. 95-144.
[On "La cause physique de l'inclinaison des plans des orbites des planètes par rapport au plan de l'équateur de la révolution du soleil autour de son axe."]
"Essai d'une nouvelle analyse de la mortalité causée par la petite vérole." Hist. de l'Acad. pp. 1-45. Paris, 1760.
De usu algorithmi infinitesimalis in arte conjectandi specimen. Novi Comm. Petrop., 1766. xii. pp. 87-98. A 2nd memoir. Petrop., 1766. xii. pp. 99-126. See a criticism by Trembley, Mém. de l'Acad., Berlin, 1799.

\section*{Bernoulli, Daniel.---contimued.}

Disquisitiones analytiquae de novo problemate conjecturali. Novi Comm. Petrop. xiv. pp. 1-25, 1769. A 2nd memoir, Petrop. xiv. pp. \(26.45,1769\).
- Dijudication maxime prohabilis phurium observationum discrepantium atque verisimillima inductio inde formanda." Acta Acad., pp. 3-23. Petrop., 1777. Crit. by Euler, pp. 24-33.
Bernoulli, Jac. Ars conjectandi, opus posthumum. Pp. ii \(+306+3 \overline{5}\). Sm. 4to, Basileae, 1713.
[Published by N. Bernoulli eight years after Jac. Bernoulli's death.]
Part 1. Reprint with notes and additions of Huygens, De ratiociniis in lude alear.

Part II. Doctrina de permutationibus et combinationibus.
Part III. Explicans usum praccedentis doctrinae in variis sortitionibus et ludis aleae. [Twenty-four problems.]

Part IV. Tradens usum et applicationem praecedentis doctrinae in civilibus, moralibus et oeconomicis.

Tractatus de seriebus infinitis. [Not connected with the subject of Probability.]

Iettre à un amy, sur les partis du jeu de paume.
[The most important sections, including Bernoulli's Theorem, are in
Part IV. For a very full account of the whole volume see Todhunter's History, chap. vii.]
Engl. Transl. of Part II. only, vide Museres.
Fr. transl. of Part I. only, vide Vastel.
Germ. transl. : Wahrscheinlichkeitsrechnumes. I Teile mit dhem Anhamen :
Brief an einem fremme üher das hallopiel, ühers. u. hrow. b. R. Hawsmer. 2 vols. Sm. 8vo. 1899. [See also Leibniz.]
Berxoulli, John. De alea, sive arte conjectandi, problemata quacdam. Collected ed. vol. iv. pp. 28-33. 1742.
Bernoulli, Johs (grandson). "Sur les suites ou séquences dans la loterie de Gênes." Hist. de l'Acad, pp. 234-253. Berlin, 1769.
"Mémoire sur un problème de la doctrine du hasard." Hist. de l'Acad., pp. 384-408. Berlin, 1768.
Bernoulli, Nicholas. Specimina artis conjectandi, ad quaestiones juris applicatae. Basel, 1709. Repr. Act. Erud. Suppl., pp. 159-170, 1711.
Bertrand, J. Calcul des probabilités. 1'p. 1vii +332 . Paris, 1889 .
"Sur l'application du calcul des probabilités à la théorie des jugements." Comptes rendus, 1887.
"Les Lois du hasard." Rev. des Deux Mondes, p. 758. Avril 1884.
 Astr. Nachrichten, vol. xv: pp. 369-404, 18:38.

Also Abhandl. von Bessel, vol. ii. pp. 372-391. Leipzig, 1875.
Bicquilley, C. F. de. Du calcul des probabilitís. 164 pp., 1783. 2nd ed. 180.5
(ierm. transl. by C. F. Rüdiger. Leipzig, 1788.
Bienaymé, J. "Sur un principe que Poisson avait cru découvrir et qu'il avait appelé loi des grands nombres." Comptes rendus de l'Acad. des Sciences morales, 1850.
[Reprinted in Journal de la Soc. do Statistiques de Paris, pl, 199-204, 1876. 1


"Sur la probabilité des résultate moyens dos observations, etc." Sav. Etranyero. v., 1 n3s.

Bienayme, J.-continued.
"Théorème sur la probabilité des résultats moyens des observations." Procès-verbaux de la Soc. Philomathique, 1839.
"Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés." Comptes rendus des séances de l'Académie des Sciences, vol. xxxvii., 1853.
[Reprinted in Journal de Liouville, 2nd series, vol. xii., 1867, pp. 158-176.]
" Remarques sur les différences qui distinguent l'interpolation de Cauchy de la méthode des moindres carrés." Comptes rendus, 1853.
"Probabilité des erreurs dans la méthode des moindres carrés." Journ. Liouville, vol. xvii., 1852.
Binet. "Recherches sur une question de probabilité" (Poisson's Theorem). Comptes rendus, 1844.
Blaschke, E. Vorlesungen über mathematische Statistik. Pp. viii +268 . Leipzig, 1906.
Вовек, K. J. Lehrbuch der Wahrscheinlichkeitsrechnung. Nach System Kleyer. Pp. 296. Stuttgart, 1891.
Bommans, G. "Die Grundbegriffe der Wahscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung." Atti del IV Congr. intern. dei matematici, Rome, 1909.
Boole, G. Investigations of Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities. Pp. ix +424 . London, 1854.
"Proposed Questions in the Theory of Probabilities." Cambridge and Dublin Math. Journal, 1852.
" On the Theory of Probabilities, and in particular on Michell's Problem of the Distribution of the Fixed Stars." Phil. Mag., 1851.
"On a General Method in the Theory of Probabilities." Phil. Mag., 1852.
"On the Solution of a Question in the Theory of Probabilities." Phil. Mag., 1854.
" Reply to some Observations published by Mr. Wilbraham in the Phil. Mag. vii. p. 465, on Boole's 'Laws of Thought.' "' Phil. Mag., 1854.
"Further Observations in reply to Mr. Wilbraham." Phil. Mag., 1854.
"On the Conditions by which the Solutions of Questions in the Theory of Probabilities are limited." Phil. Mag., 1854.
"On certain Propositions in Algebra connected with the Theory of Probabilities." Phil. Mag., 1855.
"On the Application of the Theory of Probabilities to the Question of the Combination of Testimonies or Judgments." Edin. Phil. Trans. vol. xxi. pp. 597-652, 1857.
"On the Theory of Probabilities." Roy. Soc. Proc. vol. xii. pp. 179184, 1862-1863.
Borchasit, B. Einführung in die Wahtschemlichkeitslehre. vi +86 . Berlin, 1889.
Bordoni, A. Sulle probabilità. 4to. Giorn. dell' I. R. Instit. Lombardo di Scienze. T. iv. Nuova Serie. Milano, 1852.
Borel, E. Éléments de la théorie des probabilités. 8vo, pp. vii +191. Paris, 1909. 2nd ed. 1910.

Le Hasard. Pp. iv +312. Paris, 1914.
"Le Calcul des probabilités et la méthode des majorités." L'Année psychologique, vol. 14, pp. 125-151. Paris, 1908.
"Les Probabilités dénombrables et leurs applications arithmétiques." Rendiconti del Circolo matematico di Palermo, 1909.
"Le Calcul des probabilités et la mentalité individualiste." Revue du Mois, vol. 6, pp. 641-650, 1908.

Borel. E.-cm.efinual.
 1, pp. 424-437, 19046.
 1911.
 Leipzig, 189s.
 cyklopädie der mathematischen Wissenschaften, Band 1, Heft 6.
"Wahrahminti hhemathoorie und Erfahrung." \%it...trift iür Philn. sophie und philosophische Kritik, vol. 121, pp. 71-81. Leipzig, 1903.

"Kritische Betrachtungen zur theoretischen Statistik." Jahrb. f. Nationalök. u. Stat. (3), vol. 8, pp. 641-680, 1894; vol. 10, pp. 321-360, 189j; vol. 11, pp. 671-705, 1896.
"Die erkenntnistheoretischen Grundlagen der Wahrscheinlichkeitsrechnung." Jahrb. f. Nationalök. u. Stat. (3), vol. 17. pp. 230-244. 1899.
[Criticised by Stumpf., q.v., who is answered by Bortkiewic\%, loc. cit. vol. 18, pp. 239-242, 1899.]
" Zur Verteidigung des Gesetzes dor kleinen Zahlen." Jahrb, f. Nationalök. u. Sttat. (3), vol. 39, pp. 218-236, 1910.
[The literature of this topic is not fully dealt with in this Bibliography, but very full references to it will be found in the above article.]
"Über den Präzisionsgrad des Divergenzkoeffizientes." Mitteil. des Ver bandes der österr. und ungar. Versicherungstechniker, vol. 5.
"Realismus und Formalismus in der mathematischen Statistik." Allg. Stat. Archiv, vol. ix. pp. 225-256. Munich, 1915.

Die Iterationen: ein Beitrag zur Wahrscheinlichkeitstheorie. Pp. \(x i i+205\). Berlin, 1917.

Die radioaktive Strahlung als Gegenstand wahrscheinlichkeits. theoretischer Lntersuchungen. Pp. 84. Berlin, 1913.
 guote bei Zwillings Gebieten." Sitzungsber. der Berliner Math. (ies., vol. \(x\) vii. pp. 8-14, 1918.

Homogeneität und Stabilität in der Statistik. I'p. 81 (Extracted from the Skandinavisk Aktuarietidskrift.) Uppsala, 1918.
Bostwick, A. E. "The Theory of Probabilities." Science, iii., 1896 p. 66.
 Mois, vol. \(5, \mathrm{p} \mu .641-154,1908\).
Bowrey, A. L. Elaments of Statistics. Pr. xi +459 . 4th ad. Londun. 1920.


 tion d'un point." Mém. Sav. vol. 9. pp. 255.332, P’aris, 1846.


Broals, (: D). "The Pelation between Induction and Probability." Mind, vol. xxvii. (1918). \(\mathrm{P}_{\mathrm{p}}\), 389-404, and vol. xxix. (1920) pp, 11-45.


 Leipzig, 1901. Pp, 145-153.


Bruny, Dr. Hermann. "Über ein Paradoxon der Wahrscheinlichkeitsrechnung." Sitzungsberichte der philos.-philol. Klasse der K. bayrische Akademie, pp. 692-712, 1892.
Bruns, H. Wahrscheinlichkeitsrechnung und Kollektivmasslehre. 8vo. Pp. viii \(+310+\) 18. Leipzig, 1906 .
"Das Gruppenschema für zufällige Ereignisse." Abhandl. d. Leipz. Ges. d. Wissensch. vol. xxix. pp. 579-628, 1906.
Bryant, Sophie. "On the Failure of the Attempt to deduce inductive Principles from the Mathematical Theory of Probabilities." Phil. Mag. S. 5, No. 109, Suppl. vol. 17.
Buffon. "Essai d’arithmétique morale." Supplément à l'Histoire Naturelle, vol. 4, 103 pp. 4to. 1777. Hist. Ac. Par. pp. 43-45, 1733.
Bunyakovski. Osnovaniya, etc. (Principles of the Mathematical Theory of Probabilities.) Petersburg, 1846.
Burbury, S. H. "On the Law of Probability for a System of correlated variables." Phil. Mag. (6), vol. 17, pp. 1-28, 1909.

Campbell, R. "On a Test for ascertaining whether an observed Degree of Uniformity, or the reverse, in tables of Statistics is to be looked upon as remarkable." Phil. Mag., 1859.
"On the Stability of Results based upon average Calculations." Journ. Inst. Act. vol. 9, p. 216.

A popular Introduction to the Theory of Probabilities. Pp. 16, Edinburgh, 1865.
Cantelli, F. P. "Sulla applicazione delle probabilità parziali alla statistica." Giornale di Matematica finanziaria, vol. i. (1919), pp. 30-44.
Cantor, G. Historische Notizen über die Wahrscheinlichkeitsrechnung. 4to. 8 pp. Halle, 1874.
Cantor, M. Politische Arithmetik oder die Arithmetik des täglichen Lelens. Pp. \(x+155\). Leipzig, 1898, 2nd ed. 1903.
Canz, E. C. Tractatio synoptica de probabilitate juridica sive de praesumtione. 4to. Tübingen, 1751.
Caramuel, John. Kybeia, quae combinatoriae genus est, de alea, et ludis fortunae serio disputans. 1670. [Includes a reprint of Huygens, which is attributed to Longomontanus.]
Cardan. De ludo aleac. fo., 15 pp . 1663. [Cardan ob. 1576.]
Carvello, E. Le Calcul des probabilités et ses applications. 8vo. Pp. ix + 169. Paris, 1912.

Castelnuovo, Guido. Calcolo delle probabilità. Large 8vo. Pp. xxiii + 373. Rome, 1919.

Catalan, E. "Solution d'un problème de probabilité, relatif au jeu de rencontre." Journ. Liouville, vol. ii., 1837.
"Deux problèmes de probabilités." Journ. Liouville, vol. vi.
Problèmes et théorèmes de probabilités. 4to. 1884.
Cauchy. Sur le système de valeurs qu'il faut attribuer à divers éléments déterminés par un grand nombre d'observations. 4to. Paris, 1814.
Cayley, A. "On a Question in the Theory of Probabilities." Phil. Mag., 1853.
Cesiro, E. "Considerazioni sul concetto di probabilità." Periodico di Matematica, vi., 1891.
Charlier, C. V. L. Researches into the Theory of Probability. Publ. in Engl. in Meddelanden from Lund's Astronom. Observatorium, Series ii., No. 24. 4to. 51 pp . Lund, 1906.
"Contributions to the Mathematical Theory of Statistics," Arkiv för matematik, astronomi och fysik, vols. 7, 8, 9, passim.

Vorlesungen über die Grundzüge der mathematischen Statistik. Sm. 4to. Pp. 125. Lund, 1920.
 Comptes rendus de l'Acad. des Sciences morales, vol. i. p. 103, 1875.
"La Logique du probable." Rev. phil. vol. vi. pp. 23-38, 146-163, 1878. Chrystal, G. On some Fundamental Principles in the Theory of Probability. London, 1891.
Clark, Samuel. The Laws of Chance: or a Mathematical Investigation of the Probability arising from any proposed Circumstance of Play, etc. Pp. ii \(+204,1758\).
Cones, J. Chance : A Comparison of 4 Facts with the Theory of Probabilities. Pp. 47. London, 1905.
Condorcet, Marquis de. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. 4to. P'p. exci+304. Paris, 1785. Another edition, 1804.
"Sur les événements futurs." Acad. des Sc., 1803.
Memoir on Probabilities in six parts :
1. "Rétlexions sur la règle générale qui prescrit de prendre pour valeur d'un événement incertain la probabilité de cet événement, multipliće par la valeur de l'événement en lui-même." Hist. de l'Acad. pp. 707-728. Paris, 1781.
2. "Application de l'analyse à cette question: Déterminer la probabilité qu'un arrangement régulier est l'effet d'une intention de le produire." Hist. de l'Acad., Paris, 1781. With Part i.
3. Sur l'évaluation des droits éventuels. 1782, pp. 674-691.
4. Réflexions sur la méthode de déterminer la probabilité des événements futurs, d'après l'observation des événements passés. 1783, pp. 539-5.59.
5. Sur la probabilité des faits extraordinaires. 1783, with Part 4.
6. Application des principes de l'article précédent à quelques questions de critique. 1784, pp. 454-468.
Coover, J. Experiments in Psychical Research at Leland Stanford Junior University. Pp. 641. Stanford University, California, 1917.
[See Psychical Research and Statistical Method by F. Y. Edgeworth, Stat. JL., vol. Ixxxii. (1919), p. 22.2.]
Corbacx, F. Essais métaphysiques et mathématiques sur le hasard. 8vo. Paris, 1812.
Costa. Probabilité du tir. 8vo. Paris, 1825.
"Question de probabilité applicable aux décisions rendues par les

Courcy, Alph. de. Essai sur les lois du hasard suivi d'étendus sur les assurances. 8vo. Paris, 1862.
Cournot, A. Revue de Métaphysique et de Morale, May 1905. Numéro spécialement consacré à Cournot. See especially :
F. Faure: "Les Idées de C'ournot sur la statistique," pp. 395-411.
1). Parodi: "Le Criticisme de Cournot," pp). 451-484.
F. Montré: "Les Racines historiques du probabilisme rationnel de Cournot," pp. 485-508.

Art. "Probatilités." Dictionnaire de Franck.
"Sur la probabilité des jugements et la statistique." Journal de Liouville, \(t\). iii. p. 257.
" Hemoire sur les applications du calcul des chances à la statistique

 Paris. 1493.
(ierman translation by ( \(1 . \mathrm{H}\). Schnuse. 8vo. Braunschweig, 1849.
Contcrat, I. La Lomique de Leibniz d'après des documents inédits. I'p. sis. ? 604 . Parin, 1901.

\section*{Couturat, L.--continued.}
[See especially chap. vi. for references to Leibniz's views on Probability.] Opuscules et fragments inédits de Leibniz. Paris, 1903.
Craig. Theologiae Christianae principia mathematica. 4to. London, 1699. Reprinted Leipzig, 1755.
[Crata (?).] "A Calculation of the Credibility of Human Testimony." Phil. Trans. vol. xxi. pp. 359-365, 1699.
[Also attributed to Halley.]
Crakanthorpe, R. Logica. 1st ed. London, 1622. 2nd ed. London, 1641 (auctior et emendatior). 3rd ed. Oxon., 1677.
[Book v. " De syllogismo probabili."]
Crofton, M. W. "On the Theory of Local Prohability, applied to Straight Lines drawn at random in a Plane." Phil. Trans. vol. 158, pp. 181-199, 1869.
[Summarised in Proc. Lond. Math. Soc. vol. 2, pp. 55-57, 1868.]
"Probability." Encycl. Brit. 9th ed., 1885.
"Geometrical Theorems relating to Mean Values." Proc. Lond. Math. Soc. vol. 8, pp. 304-309, 1877.
Czuber, E. Zum Gesetz der grossen Zahlen. Prag, 1889.
Geometrische Wahrscheinlichkeiten und Mittelwerte. Pp. vii +244 . Leipzig, 1884.

Theorie der Beobachtungsfehler. Pp. xiv +418 . Leipzig, 1891.
Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. Pp. viii +279 . Leipzig, 1899.

Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung. Leipzig, 1903.

Ditto. 2 vols, 8 vo. \(\mathrm{x}+410+\mathrm{x}+470\). Leipzig, 1908-10. Second edition, revised and enlarged. Vol. i. Warscheinlichleitstheorie, Fehlerausgleichung, Kollektivmasslehre, 1908. Vol. ii. Mathematische Statistik, mathematische Grundlagen der Lebensversicherung, 1910.

D'Alembert. Opuscules mathématiques: Paris, 1761-1780.
[Réflexions sur le calcul des probabilités, ii. pp. 1-25, 1761.
Sur l'application du c. des p. à l'inoculation, ii. pp. 26-95.
Sur le calcul des probabilités, etc., iv. pp. 73-105; iv. pp. 283-341; v. pp. 228-231 ; v. pp. 508-510 ; vii. pp. 39-60.]

Mélanges de littérature, d'histoire et de philosophie. Amsterdam, 1770.
[Doutes et questions sur le calcul des probabilités, vol. v. pp. 223-246.
Réflexions sur l'inoculation. Vol. v. (These two papers were reprinted in the first volume of D'Alembert's collected works published at Paris in 1821 (pp. 451-514).)]

Articles in Encyclopédie ou Dictionnaire raisonné :
"Croix ou Pile," 1754.
" Gageure," 1757.
Article in Encyclopédie méthodique: "Cartes."
D'Anières. "Réflexions sur les jeux de hasard." Mém. de l'Acad. pp. 391398. Berlin, 1784.

Dantec, Félix le. "Le Hasard et la question d'échelle." Revue du Mois, vol. 4, pp. 257-288, 1907.

Le Chaos et l'harmonie universelle. Paris, 1911.
Darbishire, A. D. Some Talks illustrating Statistical Correlation. (Reprinted from Memoirs of the Manchester Literary and Philosophical Society.) 21 pp . and plates. 8vo. 1907.
Darbon, A. Le Concept du hasard dans la philosophie de Cournot. Étude critique. Pp. 60. Paris, 1911.
Davenport, C. B. Statistical Methods. 1904.
 a. casu fortuito pendentibus." Phil. Trans. vol. xxvii. pp. 213-264, 1711.

Doctrine of Chances, or A Method of Calculating the Probabilities of Events in Play. 1st ed. 4to. P'p. xiv +175 . 1718. 2nd ed. Large 4to. Pp. xiv +258.1738 . 3rd ed. Large 4to. Pp. xii +348 . 1756. La dottrina d. azzardi applic. ai problemi d. probabilità di vita, di pensi, ecc., trad. da R. Gaeta e G. Fontana. Milan, 1776.

Miscellanea analytica de seriebus et quadraturis. 4to. Pp. \(250+22\). London, 1730.
De Morgan, A. Essay on Probabilities and their Application to Life ('ontingencies and Insurance Offices. 1838.

Formal Logic: or the Calculus of Inference Necessary and Probable. 1847.

Theory of Probabilities. 4to. 1849.
[From the Encyclopaedia Metropolitana.]
On the Structure of the Syllogism and on the Application of the Theory of 'robabilities to Questions of Argument and Authority. 4to. Camb. Phil. Soc. pp. 393-405, 1847 (read Nov. 9, 1846).

On the Symbols of Logic, the Theory of the Syllogism, and in particular of the Copula, and the Application of the Theory of Probabilities to some Questions of Evidence. 4to. Camb. Phil. Soc. vol. ix. pp. 116-125, 185l.
De Witt, John. De vardye van de lif-renten na proportio van de los-renten. La Haye, 1671.

English transl. : Contributions to the History of Insurance, by Frederick Hendriks in the Assurance Magazine, vol. 2, p. 231 (1852).
[For an abstract see \(\mathbf{N}\). Struyck, Inleiding tot het algemeine geography, etc. 4to. Amsterdam, 1740. P. 345.]
Dedekind, R. Bemerkungen zu einer Aufgabe der Wahrscheinlichkeitsrechnung. Pp. 268-271. Crelle J. vol. 1., 1855.
DEGEN, C. F. Tabularum ad faciliorem probabilitatis computationem utilem Enneas. Kiobenhavn, 1824.

1)idion, J. Calcul des probabilités appliqué au tir des projectiles. 8vo. 1858.

Iodson, James. Mathematical Repository. 3 vols. 1753. Vol. ii. pp. 82-136.
1)onkis, W. F. "Sur la théorie de la combinaison desobservations." Linuv. J. (1), vol. xv. 1850 ).
" On Certain Questions relating to the Theory of Probabilities." Phil. Mag., May 18.51.
Dommox, E. Theorie mathématique des assurances sur la vie. 2 vols. Paris, 1878.

 schen Gesellschaft der Wissenschaften mathem.-physik. 1880.
Drobrsch, M. W. Neue Darstellung der Logik. and ed. Leipzig, 185.). Brd ed. 1863. 4th ed. 1875. 5th ed. 1887.
[Probability, pp. 181-209, §§ 145-157 (references to 4th ed). |
Edaewonth, F. Y. "('alculus of Probability applied to Psychical Research." Proceedings of Soce for Psych. Res. Parts viii. and x.
" ()n the Method of ascertaining a Change in the Value of Gold." Roy: Stat. Soe. J. xlvi. pp, 714-718. 1883.
"Law of Error." Phil. Mag. (5) vol. xvi. pp, 300.309, 1883.
"Mcthod of least Squares." Phil. Mag. (5) vol. xvi. pp. 360-37.5, 1883.
"Physical Basis of Probability." Phil. Mag. vol. xvi. ['], 433-43:5, 188:3.

Edgeworth, F. Y.-continued.
"Chance and Law." Hermathene (Dublin), 1884.
"On the Reduction of Observations." Phil. Mag. (5) vol. xvii. pp. 135-141, 1884.
"Philosophy of Chance." Mind, April 1884.
"A priori Probabilities." Phil. Mag. (5) vol. xviii. pp. 209-210, 1884.
"On Methods of Statistics." Stat. Journ. Jub. vol. pp. 181-217, 1885.
[Criticised by Bortkiewicz and defended by Edgeworth, Jahrb. f. nat. Ök. u. Stat. (3), vol. 10, pp. 343-347; vol. 11, pp. 274-277, 701-705, 1896.]
"Observations and Statistics." Phil. Soc. 1885.
"Law of Error and Elimination of Chance." Phil. Mag., 1886, vol. xxi. pp. 308-324.
"Problems in Probabilities." Phil. Mag., 1886, vol. xxii. pp. 371-384, and 1890, vol. xxx. pp. 171-188.

Metretike : or the Method of Measuring Probability and Utility. 8vo. 1887.
"On Discordant Observations." Phil. Mag. (5) vol. xxiii. pp. 1887.
"The Empirical Proof of the Law of Error." Phil. Mag. (5) vol. xxiv. pp. 330-342, 1887.
"The Element of Chance in Competitive Examinations." Roy. Stat. Soc. Journ. liii. pp. 460-475 and 644-663, 1890.
"The Law of Error and Correlated Averages." Phil. Mag. (5) vol. xxxv. pp. 63-64, 1893.
"Statistical Correlation between Social Phenomena." Roy. Stat. Soc. Journ. Ivi. pp. 670-675, 1893.
"The Asymmetrical Probability-Curve." 1896. Phil. Mag. vol. xli. pp. 90-99.
"Miscellaneous Applications of the Calculus of Probabilities." Roy. Stat. Soc. Journ. 1x. pp. 681-698, 1897; 1xi. pp. 119-131 and 534-544, 1898.
"Law of Error." Phil. Trans. vol. xx.
"The Generalised Law of Error." Stat. Journ. vol. 1xix., 1906.
"On the Probable Errors of Frequency-Constants." Stat. Journ. vol. lxxi. pp. 381-397, 499-512, 651-678, 1908; and vol. lxxii. pp. 81-90, 1909.
"On the Application of the Calculus of Probabilities to Statistics." Bulletin xviii. of the International Statistical Institute, Paris, 1910, 32 pp .
"Applications of Probabilities to Economics." Economic Journal, vol. xx. pp. 284-304, 441-465, 1910.
"Probability." Encyclopaedia Britannica, 11th ed. vol. 22, pp. 376403, 1911.
"On the Application of Probabilities to the Movement of Gas-Molecules." Phil. Mag., vol. xl., pp. 249-272, 1920.
"Molecular Statistics." Roy Stat. Soc. Journ., vol. Ixxxiv. pp. 71-89, 1921.

Eggenberger, J. "Beiträge zur Darstellung des bernoullischen Theorems." Berner Mitth. vol. 50 (1894); and Zeitschr. f. Math. u. Ph. 45 (1900), p. 43.
Elderton, W. P. Frequency-Curves and Correlation. 8vo. Lundon, 1907. xiii +172 .
[Contains a useful list of papers on Correlation, p. 163.]
Ellis, R. L. "On the Foundations of the Theory of Probability." 4to. Camb. Phil. Soc. vol. viii., 1843.
[Reprinted in " Mathematical and other Writings," 1863.]
"On a Question in the Theory of Probabilities." Camb. Math. Journ. No. xxi. vol. iv., 1844.
[Reprinted in " Mathematical and other Writings," 1863.]

Ellis, R. L.--continued.
"On the Method of Least Squares." Trans. C'amb. Phil. Soc. vol. viii., 1844.
[Reprinted in "Mathematical and other Writings," 186.3.]
"Remarks on an alleged Proof of the 'Method of Least Squares.'" Phil. Mag. (3) vol. xxxvii., 1850.
[Reprinted in "Mathematical and other Writings," 1863.]
"Remarks on the Fundamental Principle of the Theory of Probabilities." Trans. Camb. Phil. Soc. vol. ix., 1854.
[Reprinted in " Mathematical and other Writings," 1863.]
Elsas, A. "Kritische Betrachtungen über die Wahrscheinlichkeitsrechnung." Philos. Monatssch. vol. xxv. pp. 557-584, 1889.
Emerson, William. Miscellanies, 1776 . [See espec, pp. 1-48.]
Encee, J. F. Methode der kleinsten Quadrate. Fehler theoret. Untersuchungen. Berlin, 1888.
Evgel, G. "Über Möglichkeit und Wirklichkeit." Philos. Monatssch. vol. v. pp. 241-271, 1875.

Ermakoff, W. P. Wahrscheinlichkeitslehre (in Russian).
Euler. "Calcul de la probabilité dans le jeu de rencontre." Hist. Ac. Berl. (1751), pp. 255-270, 1753.
"Sur l'avantage du banquier au jeu de pharaon." Hist. Ac. Berl. (1764), pp. 144-164, 1766.
"Sur la probabilité des séquences dans la loterie génoise." Hist. Ac. Berl. (1765), pp. 191-230, 1767.
"Solution d'une question très difficile dans le calcul des probabilités." Hist. Ac. Berl. (1769), pp. 285-302, 1771.
"Solutio quarundam quaestionum difficiliorum in calculo probabilium." Opuscula analytica, vol. ii. pp. 331-346, 1785.
"Solutio quaestionis ad calculum probabilitatis pertinentis: Quantum duo conjuges persolvere debeant, ut suis haeredibus post utriusque mortem certa argenti summa persolvatur." Opuscula analytica, vol. ii., pp. 31533:3, 17n.
"Wahrscheinlichkeitsrechnung." Opera omnia, ser. 1, A, vol. iv. Leipzig.
 et leurs variations." Journ. Soc. Stat. de Paris, pp. 188-200, 1900).
Fechner, G. Th. Kollektivmasslehre. (Edited by G. F. Lipps.) 1897.
Fick, A. Philosophischer Versuch über die Wahrscheinlichkeiten. P'p. 46. Würzburg, 1883.
Fismer, A. The Mathematical Theory of Probabilities. Translated from the Danish. Pp. \(x x+171\). New York, 1915.
Forbes, J. D. " On the alleged Evidence for a Physical (ommexion between Stars forming Binary or Multiple Groups, deduced from the Doctrine of ('hances." Phil. Mag., Dee. 1850. (See also Phil. Mag., Aug. 1849.)
Fornery. The Lagic of Probabilities. Transl. from the French. 8vo. London, n.d. (? 1760.)



 1773.

Fuss, N. "Recherches sur un problème du calcul des probabilités." Act. Ac. P'etr. (1779), pars posterior, pp, 81-92, 1783.
\(\cdots\) supfiment an memoire sur un probleme du atoul des probabilitos. Act. Ac. Petr. (1780), pars posterior, pp. 91-96, 1784.

Galileo, G. "Considerazioni sopra il givoci dei dadi." Opere, vol. iii. pp. 119-121, 1718. Also, Opere. vol. xiv. pp. 293-296. Firenze, 18 õ5.
"Lettere intorno le stima di un cavallo." Opere, vol. xiv. pp. 231-284. Firenze, 1855.
Galloway, T. A Treatise on Probability. 8vo. Edinburgh, 1839. (From the 7th edition of the Encyclopaedia Britannica.)
Galton, F. "Correlations and thrir Measurement." Proc. Roy. Soc., vol. xlv. pp. 136-145.

Probability, the Foundation of Eugenics. Herbert Spencer Lecture, 1907. (Reprinted-Essays in Eugenics. 8vo. ii +109 pp. London, 1909.)

Gardon, C. Antipathies des 90 nombres, jrobabilités, et observations comparatives, sur les loteries de France et de Bruxelles. 8vo. Paris, 1801.

Traité élémentaire des probabilités, etc. Paris, 1805.
L'investigateur des chances . . . pour obtenir souvent des succès aux loteries impériales de France. Paris.
Garve, C. De nonnullis quae pertinent ad logicam probabilium. 4to. Halae, 1766.

Gataker, T. On the Nature and Use of Lots. 4to. 1619.
Gauss, C. F. Theoria motus corporum coelestium. 4to. Hamburg, 1809.
"Theoria combinationis observationum errombus minin!is obnoxiae." Comm. Soc. Göttingen, vol. v. pp. 33-90. 1823.

Méthode des moindres carrés. Traduit en français par J. Bertrand. 8 vo. 1855.
[A translation of part of the above.]
Wahrscheinlichkeitsrechnung. Werke, vol. iv. pp. 1-53. 4to. Göttingen, 1873.

Getsenheimer, L. Über Wahrscheinlichkeitsrechnung. 8vo. Berlin, 1880.
Gilman, B. I. "Operations in Relative Number with Applications to Theory of Probability." Johns Hopkins Studies in Logic, 1883.
Gladstone, W. E. "Probability as a Guide to Conduct." Nineteenth Cent. vol. v. pp. 908-934, 1879; and in "Gleanings," vol. ii. pp. 153-200.
Glatsher, J. W.L. "On the Rejection of Discordant Observations." Monthly Notices R. Astr. S. vol. xxiii., 1873.
"On the Law of Facility of Errors of Observation, and on the Method of Least Squares." Mem. R. Astr. S. vol. xxxix., 1872.
Goldschmidt, L. "Wahrscheinlichkeit und Versicherung." Bull. du Comité permanent des Congrès Internationaux d'Actuaires, 1897.
Die Wahrscheinlichkeitsrechnung: Versuch einer Kritik. Pp. 279. Hamb., 1897.
[Cf. Zeitschr. f. Philos. u. phil. Kr., cxiv., pp. 116-119.]
Gonzalez, T. Fundamentum theologiae moralis, id est tractatus theologicus de recto usu opinionum probabilium. 4to. Dillingen, 1689. Naples, 1694.
[An abridgement entitled: Synopsis tract. theol. de recto usu opin. prob., concinnata a theologo quodam Soc. Jesu: cui accessit logistica probabilitatum. 3rd ed. 8vo. Venice, 1696. See Migne, Theol. Cur. Compl., vol. xi., p. 1397.]
Gourand, Ch. Histoire du calcul des probabilités depuis ses origines jusqu'à nos jours. 8vo. Paris, 1848, 148 pp .
[His history seems to be a portion of a very extensive essay in 3 folio volumes containing 1929 pp ., written when he was very young, in competition for a prize proposed by the Fr. Acad. on a subject entitled "Théorie de la certitude"; see Séances et Travaux de l'Académie des Sciences morales et politiques, vol. x. pp. 372, 382, vol. xi. p. 137. See TodHeveris.]
Gravesande, W. J. 'S. Introductio ad philosophiam, metaphysicam et logicam continems. Sro. Vimetios, 1737.

Gravesande:, W. J. is cunlinued.
(Euvres philosophiques et mathématiques. 4to. Amsterdam, 1774, 2 vols. 4to. ii. pp, 82-93, 2:21-248.
 rechnung." Abhandlungen der Frisschen schule, N.F., vol. iii., 1910.

 philosophische Kritik. Band 118, pp. 154-167. Leipzig, 1901.
[See also Brümse, Marbe, and v. Bortriewicz.]
Grolots. "Sur une question de probabilité appliquée à lá théorie des nombres." Journal de l'Institut, 1872.
Giroschics, J. A. Logica probabilium in artium practicarum subsidium adornata. Sm. 8vo. Halae, 1764. Y'p. xvi +352 .
Grünbaum, H. Isolierte und reine Gruppen und die Marbesche Zahl "p." Würzburg, 1904.
Guibert, A. "Solution d'une question relative à la probabilité des jugements rendus à une majorité quelconque." Liouv. J. (1) vol. iii., 1838.

Hack. Wahrscheinlichkeitsrechnung. Leipzig, 1911.
Hagen, G. F. Meditationes philosophicae do methodo mathematico. Norimbergae, 1734.

Fortoetzonse einigur aus der Mathomatio abenommenen Rewerin, nach wehben swh der mensehliche Verstand bei Eramenng der Waherheiten richtet. Halle, 1737.
Hagen, G. Grundzüge der Wahrscheinlichkeitsrechnung. Berlin, 1837. (2nd ed. 1867, 3rd ed. 1882.)
 Grundzüge der Wahrscheinlichkeitsrechnung. 38 pp. Berlin, 1884.
Halley. See Craig.
Hans, John. See J. Arbuthnot.
 vol. 53, pp. 152-178, 1901.

Hawsen, P. A. "Ưber die Anwendung der Wahrscheinlichkeitsrechnung auf geodätische Vermessungen." Astr. N. vol. ix. 1831.
 Vierteljahrsschr. f. wiss. Phil. u. Soz., vol. xxviii., 1904.
 1834.
 1840.

 lichkeitsrechnung." Zeitschr. f. Math. u. 'hyys., vol. 38, pp. 374-376. Leipzig, 1893.
 Annalen der Naturphilosophie, vol. 1.
Henry, Charles. La Loi des petits nombres. Recherches sur lo sens de l'écart probable dans les chances simples à la roulette, au trente-et-quarante ete., en ameal dans les phemomenes dipendant de consem purement accidentales. 72 pp . 8vo. Paris, 1908.
Ifenschel, W. "On the Theory of Probabilities." Journal of Actuaries, 1869.
"(Lumbly on I'mimblities." Edan. Rev... Ision.

"On an Application of the Rule of Succession." Edin. Kev., 18500.

Herz, N. Wahrscheinlichkeits- und Ausgleichungsrechnung. Pp. iv + 381. Leipzig, 1900.
Hibben, J. G. Inductive Logic. London, 1896.
[See chaps. xv., xvi.]
Hobhouse, L. T. Theory of Knowledge.
[See Part II., chaps. x., xi.]
Hoyle. An Essay towards making the Doctrine of Chances easy to those who understand vulgar Arithmetic only. Pp. viii \(+73,1754,1758,1764\).
Huberdt, A. Die Principien der Wahrscheinlichkeitsrechnung. 4to. Berlin, 1845.

Hume, David. Treatise on Human Nature. 1st ed. 1739.
[See especially Part III.]
An Enquiry concerning Human Understanding.
[See specially Section vi.]
Essays, Part I., XIV. On the Rise and Progress of the Arts and Sciences, pp. 115, 116, 1742.
Huygens, Ch. "De ratiociniis in ludo aleae." Schooten's Exercitat. math. pp. 519-534. 4to. Lugd. Bat., 1657.
[Written by Huygens in Dutch and translated into Latin by Schooten.] Engl. transl. by W. Browne. Sm. 8vo, pp. 24. London, 1714.
[See also Jac. Bernoullt, Arbuthnot (Engl. Transl.), and Vastel (Fr. Transl.).]

Jahn, G. A. Die Wahrscheinlichkeitsrechnung und ihre Anwendung auf das wissenschaftliche und praktische Leben. Leipzig, 1839.
Janet. La Morale. Paris, 1874. [See Bk. iii. chap. 3 for Probabilism.]
Engl. transl. The Theory of Morals. New York, 1883, pp. 292-308.
Jevons, W. S. Principles of Science. 2 vols. 1874.
Jordan, C. "De quelques formules de probabilité (sur les causes)." Comptes rendus, 1867.
Jourdain, P. E. B. "Causality, Induction, and Probability." Mind, vol. xxviii. pp. 162-179, 1919.

Kahle, L. M. Elementa logicao probabilium methodo mathematica, in usu scientiarum et vitae adornata. Pp. \(10+\mathrm{xxii}+245\). Sm. 8vo. Halae, 1735.

Kanner, M. "Allgemeine Probleme der Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fragen der Statistik." Journ. des Collegiums für Lebens-Versicherungs-Wissenschaft. Berlin, 1870.
Kaufmann, Al. Theorie und Methoden der Statistik. [Translated from the Russian.] Pp. xii +540. Tübingen, 1913.
Kepler, J. "De stella nova in pede serpentarii." 1606. See J. Kepler's Astr. Op. Omn. edidit Frisch, ii. pp. 714-716.
Kirchmann, J. H. von. Über die Wahrscheinlichkeit. Pp. 60. Leipzig, 1878.

Knapp. "Quetelet als Theoretiker." Jahrb. f. nat. Ök. und Stat. (New Series), vol. xviii.
Kozák, Josef. Grundlehren der Wahrscheinlichkeitsrechnung als Vorstufe für das Studium der Fehlerausgleichung, Schiesstheorie, und Statistik. Vienna, 1912.

Theorie des Schiesswesens auf Grundlage der Wahrscheinlichkeitsrechnung und Fehlertheorie. Vienna, 1908.
Kries, J. von. Die Principien der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung. Pp. 298. 8vo. Freiburg, 1886.
[See also Lexis, Meinong and Siqwart.]

Lacrorx, S. F. Traité d́fomentaire du caleul des probabilités. P'p. viii : 299. swo. Paris, Islti.
[2nde éd., revue et augmentée, 1822; 4th ed. 1864.]
[Translated into German: E. S. Unger, Erfurt, 1818.]
Lagrange. "Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations, dans lequel on examine les avantages de cette méthode par le calcul des probabilités, et où l'on résout différents problèmes relatifs ì cette matière." Misc. Taurinensia, vol. 5, pp. 167-232, 1770-1773. Euvres complètes, vol. 2, Paris, 1867-1877.
" Recherches sur les suites recurrentes . . . et sur l'usage de ces équations dans la théorie des hasards." Nouv. Mém. Ac. Berl. (1775), pp. 183-272, 1777. (Euvres complètes, vol. 4. Paris, 1867-1877.

Latsant, C. A. Algèbre. Théorie des nombres, probabilités, géométrie de situation. Paris, 1895.
Lambert, J. H. "Examen d'une espèce de superstition ramenée au calcul des probabilités." Nouv. Mém. Ac. Berl, 1771, pp. 411-420.
Lamman, li. I'nuren huncen über die Ermitteluner bon Wahrscheinlichkeiten. (Inaug.-Dissert.) P'p. 80. Zürich, 1904.
 Arch., vol. 70, 1884.
Lange, F. A. Logische Studien.
Laplace. Essai philosophique sur les probabilités. (Printed as introduction to Théorie analytique des probabilités, from 2nd ed. of the latter onwards.) 4to. Paris, 1814.
(ierman translation by Tönnies. Heidelberg, 1819. German translation by N. Schwaiger. Leipzig, 1886.

A Philosophical Essay on Probabilities: transl. from the 6th French ed. by E. W. Truscott and F. L. Emory. 8vo. New York, 1902, 196 pp.

Théorie analytique des probabilités.
1st ed. 4to. Paris, 1812. 1st and 2nd Suppl., 1812-1820. 2nd ed. 4to. exi \(+506+2\), Paris, 1814. 3rd Suppl. 1820. 3rd ed. Paris, 182 (). 4th Suppl. after 18:0. Euvres completes, vol. 7, 1p. excv +691, Paris, 1847. (Euvres complètes, vol. 7, pp. 832, Paris, 1886.
"Rowherehes sur l'intigration des iqpuations differentielles aus dime rences finies, et sur leur usage dans la théorie des hasards." Mém. prés. ì 1'Acad. des Sc., pp. 113-163, 1773.
" Mémoire sur les suites récurro-récurrentes et sur leurs usages dans la théorie des hasards." Mém. prés. à l'Acad. des Sc., vol. 6, pp. 353-371, 1774.
"Mémoire sur la probabilité des causes par les événements." Mém. prés. à l'Acad. des Sc., voll. 6, pp. 621-656, 1774.
"Mémoire sur les probabilités." Mém. prés. à l'Acad. des Sc., pp. 227. 332, 1780 .
" Mémoire sur les approximations des formules qui sont fonctions de
 de l'Inst., pp. 353-415, 539-5665, 1810.
"Mémoire sur les intégrales définies, et leur application aux probabilités." Mém. de l'Inst., pp). 279-347, 1810.
[The above memoirs are reprinted in (Euvres completes, vols. 8, 9, and


Sur l'application du calcul des probabilités appliqué à la philosophic naturelle. C'onn. des temps. Euvres completes, vol. 13. Paris, 194 .
"Applications du calcul des probabilités aux observations et spécialement aux operations du nivellement." Annales de Chimie. CEuvres completes, vol. 14, Paris, 1913.


Laurent, H. Traité du calcul des probabilités. Paris, 1873.
[A la fin une liste des principaux ouvrages (320) ou mémoires publiés sur le calcul des probabilités.]
"Application du calcul des probabilités à la vérification des répartitions." Journ. des Actuaires français, vol. i.
"Sur le théorème de J. Bernoulli." Journ. des Actuaires français, vol. i.
Lechalas, G. "Le Hasard." Rev. Néo-scolastique, 1903.
"A propos de Cournot: hasard et déterminisme." Rev. de M't. et de Mor., 1906.
Legendre. "Méthode des moindres carrés." Mém. de l'Inst., 1810, 1811.
Nouvelles méthodes pour la détermination des orbites des comètes. Paris, 1805-6.
Lehr. "Zur Frage der Wahrscheinlichkeit von weiblichen Geburten und von Totgeburten." Zeitschrift f. des ges. Staatsw., vol. 45, p. 172, and p. 524, 1889.
Leibniz. Nouveaux Essais. Liv. ii. chap. xxi. ; liv. iv. chaps. ii. § 14, xv., xvi., \(x\) viii., \(x x\).

Opera omnia, ed. Dutens, v. 17, 22, 28, 29, 203, 206; vi. pt. i., 271, 304, 36, 217 ; iv. pt. iii. 264.

Correspondence between Leibnitz and Jac. Bernoulli. L.'s Gesammelte Werke (ed. Pertz and Gerhardt), vol. 3, pp. 71-97, passim. Halle, 1855. [These letters were written between 1703 and 1705.]
See also s.v. Couturat.
Lemoine, E. "Solution d'un problème sur les probabilités." Bulletin de la Soc. math. de Paris, 1873.

Questions de probabilités et valcurs relatives des pièces du jeu des échecs. 8vo. 1880.
"Quelques questions de probabilités résolues géométriquement." Bull. de la Soc. math. de France, 1883.
"Divers problèmes de probabilité." Ass. française pour l'Avancement des Sciences, 1885.
Lexis, W. Abhandlungen zur Theorie der Bevölkerungs- und Moral-statistik. Pp. 253. Jena, 1903.

Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft. Pp. 95. Freiburg, 1877.
"UUber die Wahrscheinlichkeitsrechnung und deren Anwendung auf die Statistik. Jahrb. f. nat. Öl. u. Stat. (2), vol. 13, pp. 433-450, 1886.
[Contains a review of v. Kries's "Principien."]
"Über die Theorie der Stabilität statistischer Reihen." Jahrb. f. nat. Ök. u. Stat. (1), vol. 32, p. 604, 1879.
[Reprinted in Abhandlungen.]
"Das Geschlechtsverhältnis der Geborenen und dic Wahrscheinlichkeitsrechnung." Jahrb. f. nat. Ök. u. Stat. (1), vol. 27, p. 209, 1876.
[Reprinted in Abhandlungen.]
Einleitung in die Theorie der Bevölkerungsstatistik. Strassburg, 1875.
Litagre, J. B. J. Calcul des probabilités et théorie des erreurs avec des applications aux sciences d'observation en général et à la géodésie en particulier. 416 pp . Brussels, 1852. 2nd ed. 8vo. 1879.
"Sur la probabilité d'une cause d'erreur régulière, etc." Bull. de l'Acad. de Belgique, 1855.
Liapounoff, A. "Sur une proposition de la théorie des probabilités." Bull. de l'Acad. des Sc. de Saint-Pét., v. série, vol. xiii.
"Nouvelle Forme du théorème sur la limite de probabilité." Mém. de l'Acad. des Sc. de Saint-Pét., viii. série, vol. xiii. (1901).
Lifbermeister, C. "Über Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik." Sammlung klinische Vorträge, Nr. 110. 1877.

 (кхх, pp. 5 -65, 1902.
Lipps, (i. F. Kollectiomasslehre. 1897.


 répétées." Crelle J. 18:4.

Reprinted. Liouv. J. vol. 24. 1842.
Lotrin, J. Le Calcul des probabilités et les régularités statistiques. 32 pp . 8vo. Louvain, 1910. (Originally published in the Revue Néo-scolastique, Feb. 1910.)

Quetelet, statisticien et sociologue. Louvain, 1912. P'p. xxx +564 . [Contains a very full discussion of Quetelet's Work on Probability.]
Lotze, 11. Lowik. Ist ed. 1874, 2nd ed. 1880.

(See Bk. ii. chap). ix. : Determination of Single Facts and Calculus of Chances. 1

Lubbock, J. W., and Drinkwater. Treatise on Probability. [Library of Lseful Kinowledge.]
[Often wrongly ascribed to De Morgan.]

Macalistler, Dosald. The Law of the Geometric Mean. Phil. Trans., 1879.

The Calculus of Equivalent Statements. Proc. Lond. Math. Soc. Six papers.
LSee particularly 1877, vol. ix. pp. 9-20; 1880, xi. 113-121, 4th paper; 14!
"Growth and Use of a Symbolical Language." Memoirs Manchester Lit. Phil. Soce series iii. vol. 7, 1881 .
"Symbolical or Abbreviated Language with an Application to Mathematical IProbability:" Math. Questions, vol. 28, pp. 20-23.

Various Papers in Mathematical Questions from the Journal of Education, vols. 29,33 , cte.
 Lond. Math. Soc. vol xii. p. 102.
Macfarlaye, Alexander. P'rinciples of the Algebra of Logic.
[See espercially chaps. ii., iii., v., xx., xxi., xxii., xxiii., and the examples.]
Various Papers in Mathematical Questions from the Journal of Education, vols. 32, 36, cte.
MacManos, P'. A. "On the Probability that the Successful (andidate at an Election by Ballot may never at any time have fewer lites than the one who is unsuccessful, etc." Phil. Trans. (A), vol. 209, pp. 153-175, 1909.


 di Maternatica . Fisica della Societia Italiana, vol. 1, pp. 768 -824, \(178 \%\).
 vii. pp, 133-1633.

Manstoss. P. "Sur la portée objective du caleul des probabilités." Bulletin de: 1’Académie de Belgique (Classe des sciences), pp. 12:35-12394, 1903.3.
 keitslehre. 50 pup. Laipzig, 1899.

Die Gileichförmigkeit in der Welt. Munich, 1916t.

Markoff, A. A. "Über die Wahrscheinlichkeit à posteriori" (in Kussian). Mitteilungen der Charkowv Math. Gesell. 2 Serie, vol. iii. 1900.
"Untersuchung eines wichtigen Falles abhängiger Proben" (in Russian). Abh. der K. Russ. Ak. d. W., 1907.
"Über einige Fälle der Theoreme vom Grenzwert der mathematischen Hoffnungen und vom Grenzwert der Wahrscheinlichkeiten " (in Russian). Abh. der K. Russ. Ak. d. W., 1907.
"Erweiterung des Gesetzes der grossen Zahlen auf von einander abhängige Grössen " (in Russian). Mitt. d. phys.-math. Ges. Kazan, 1907.
"Über einige Fälle des Theorems vom Grenzwert der Wahrscheinlich. keiten" (in Russian). Abh. der K. Russ. Ak. d. W., 1908.
" Erweiterung gewisser Sätze der Wahrscheinlichkeitsrechnung auf eine Summe verketteter Grössen " (in Russian). Abh. der K. Russ. Ak. d. W., 1908.
"Untersuchung des allgemeinen Falles verketteter Ereignisse" (in Russian). Abh. der K. Russ. Ak. d. W., 1910.
"Über einen Fall von Versuchen, die cine komplizierte zusammen hängendes Kette bilden," and "Über zusammenhängende Grössen, die keine echte Kette bilden " (both in Russian). Bull. de l'Acad des Sciences. Petersburg, 1911

Wahrscheinlichkeitsrechnung. Transl. from 2nd Russian edition by H. Liebmann. Leipzig, 1912. Pp. vii +318 .

Démonstration du second théorème-limite du calcul des probabilités par la méthode des moments. Saint-Pétersbourg, 1913. Pp. 66.
[Supplement to the 3rd Russian edition of Wahrscheinlichkeitsrechnung, in honour of the bicentenary of the Law of Great Numbers, with a Portrait of Jacques Bernoulli.]
Masaryk, 'T. G. David Hume's Skepsis und die Wahrscheinlichkeitsrechnung. Wien, 1884.
Maseres, F. The Doctrine of Permutations and Combinations, being an Essential and Fundamental Part of the Doctrine of Chances: As it is delivered by Mr. James Bernoulli, in his excellent Treatise on the Doctrine of Chances, intitled, Ars conjectandi . . . 8vo. London, 1795.
Metnoxis, A. Review of V'on Kries's" Die Principien der Wahrscheinlichkeitsrechnung." Göttingische Gelehrte Anzeigen, vol. 2, pp. 56-75, 1890.

Über Möglichkeit und Wahrscheinlichkeit: Beiträge zur Gegenstandstheorie und Erkenntnistheorie. Pp. xvi. +760 . Leipzig, 1915.
Meissner (Отto). Wahrscheinlichkeitsrechnung: I. Grundlehren ; II. Anwendungen. Leipzig, 1912; 2nd ed., 1919. Pp. \(56+52\).
[An elementary primer.]
Mendelssohy, Moses. Philos. Schriften, 2 Tle. \(12 \mathrm{mo}\). Pp. xxii \(+278+283\). Berlin, 1771. (Vide especially vol. ii. pp. 243-283, entitled "Ueber die Wahrscheinlichkeit.")
Mentrè, F. "Rôle du hasard dans les inventions et découvertes." Rev. de Phil., 1904.
"Les Racines historiques du probabilisme rationnel de Cournot." Rev. de Métaphysique et de Morale, pp. 485-508, May 1905.

Cournot et la renaissance du probabilisme au xixe siècle. Paris, 1908.
Merriman, M. A Text-book of the Method of Least Squares. New York, 1884. Pp. vii + 198. 6th ed., 1894.
" List of Writings relating to the Method of Least Squares, with Historical and Critical Notes." Trans. Connecticut Acad. vol. 4, pp. 151-232, 1877.
Mertz. Die Wahrscheinlichkeitsechnung und ihre Anwendung, ete. Frankfort, 1 s.i.t.
Messina, I. " Intorno a un nuovo teorema di calcolo delle probabilità." 20 pp . 4to. Giornale di Matematiche di Battaglini, vol. 1vi. (1918). Naples.

\section*{Missons．I．－rombinaud．}
［Described Stat．Jl．vol．1xxxii．（1919），p．（112．］
 Bernoulli e sui postulati empirici per la loro applicazione．＂Boll．del Lavoro et della Presidenza，vol．xxxiii．（1920）．
Meyer，A．Essai sur une exposition nouvelle de la théorie analytique des


Cours de calcul des probabilités fait à l＇université de Liége de 1849 ì 1857. Publié sur les mss．de l＇auteur par F．Folie．Bruxelles， 1874.

Vorlesungen über Wahrscheinlichkeitsrechnung．（Translation of the above by E．（zuber．）Pp．xii +554 ．I 1 ipzig， 1879.
Mrchele．＂An Inquiry into the Probable Parallax and Magnitude of the Fixed Stars，from the Quantity of Light which they afford us，and the particular Circumstances of their Situation．＂Phil．Trans．vol．57，pp．234－264， 176．）．
 et de Mor．vol．x．pp．667－681， 1902.
Mifle，J．ふ．Srstem of Logic．Bk．iii．chaps．18， 23.
Moxdesir．＂＇solution d＇une question qui se présente dans le calcul des pro－ bahilités．＂Liouville Journ．wol．ii．
Moxro，C．J．＂Note on the Inversion of Bernuulli＇s Theorem in Probabilities．＂ Proc．Lond．Math．Soc．vol．5，pp．74－78 and 145， 1874.
Montessus，R．de．Leçons élémentaires sur le calcul des probabilités．Pp． 191．Paris，1908．（Reviewed Stat．Journ．，1909，p．113．）
＂Le Hasard．＂Rev．du Mois，March \(190{ }^{\circ}\) ．
Moxtesses，R．de，and Lechalas，G．＂L＇n Paradoxe du calcul des proba－ bilités．＂Nouv．Ann．iv．（3）， 1903.
Mostmort，P．de．Essai d＇analyse sur les jeux de hasard．4to．Pp．


Ditto．4to．pp．414．Paris，1714．（The 2nd ed．is increased by a treatise on Combinations，and the correspondence between M．and Nicholas Pammolli．）
Mostucla，J．T．Histoire des mathématiques． 4 vols．4to．Paris，1799－ ノいに。

Vol．iii．pp．3sin－426．
Niawcomb，simon．A Statistical Inquiry into the Probability of Causes of the Production of Sex in Human Offspring．（Published by the Carnegie Institution of Washington．）P1，34．8vo．Washington， 1904.
Nicole，F．＂Examen et résolution de quelques questions sur les jeux．＂Hist． Ac．Par．pp．45－56，331－344，1730．
Niecpost，C．F．de．Un puletort ou amusemens d＇un sexagenaire．8vo． Bruxelles，1818．Containing＂（onversations sur la théorie des pro－ l，abhatis．＂

 1～！
Nixos，J．W：＂An Experimental Test of the Normal Law of Firror．＂Sitat． Journ．vol．76，pp．702－706， 1913.

Oettinger，L．Die Wahrscheinlichkeitslehre．4to．Berlin，18io．
［Reprinted from Crelle，J．，vols．26，30，34，36，under the title，Unter－ wuchungen über Wahrscheinlichkeitsechnung．］
 ＂＇ur la probabilité des hypotheses．＂Malanges math．et astr．， 1859.

Pagano, F. Logica dei probabili. Napoli, 1806.
Parisot, S. A. Traité du calcul conjectural ou l'art de raisonner sur les choses futures et inconnues. 4to. Paris, 1810.
Pascal, B. "Letters to Fermat." Varia opera mathematica D. Petri de Fermat. pp. 179-188, Toulouse, 1678.

Euvres, vol. 4, pp. 360-388, Paris, 1819.
Patavio. Probabilismus methodo mathematico demonstratus. 1840.
Paulian, Fr. "L'erreur et la sélection." Rev. Philos. vol. viii. pp. 72-86, 179-190, 290-306, 1879.
Peabody, A. P. "Religious Aspect of the Logic of Chance and Probability." Princeton Rev. vol. v. pp. 303-320, 1880.
Pearson, K. "On a Form of Spurious Correlation which may arise when Indices are used, etc." Proc. Roy. Soc. vol. lx. pp. 489-498.
"On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling." Phil. Mag. (5), vol. \(50, \mathrm{pp} .157-160,1900\).
"On some Applications of the Theory of Chance to Racial Differentiation." Phil. Mag. (6), vol. 1, pp. 110-124, 1901.

Contributions to the Mathematical Theory of Evolution.
[The main interest of the twelve elaborate memoirs published in the Phil. Trans. under the above title is in every case statistical. References are given below to those of them which have most reference to the theory of Probability and in which Professor Pearson's general theory is mainly dereloped.]
II. "Skew Variation in Homogeneous Material." Phil. Trans. (A), vol. 186, Part i. pp. 343-414, 1895.
III. "Regression, Heredity, and Panmixia." Phil. Trans. (A), vol. 187, pp. 253-318, 1897.
IV. "On the Probable Errors of Frequency C'onstants and on the Influence of Random Selection on Variation and Correlation." Phil. Trans. (A), vol. 191, pp. 229-311, 1898. (With L. N. G. Filon.)
VII. "On the Correlation of Characters not quantitatively measurable." Phil. Trans. (A), vol. 195, pp. 1-47, 1901.
"Mathematical Contributions to the Theory of Evolution." Roy. Stat. Soc. Journ. lvi., 1893, pp. 675-679 ; lix., 1896, pp. 398-402; lx., 1897, pp. 440-449.
"On the Mathematical Theory of Errors of Judgment, with special reference to the Personal Equation." Phil. Trans. (A), vol. 198, pp. 235 299, 1902.

On the Theory of Contingency and its relation to Association and Normal Correlation. Pp. 35. 4to. London, 1904.

On the Ceneral Theory of Skew Correlation and Non-linear Regression. Pp. 54. 4to. London, 1905.

On further Methods of determining Correlation. London, 1907. (Reviewed by G. U. Yale Journ. Roy. Stat. Soc., Dec. 1907.)
"On the Influence of Past Experience on Future Expectation." Phil. Mag. (6), vol. 13, pp. 365-378, 1907.
"The Fundamental Problem of Practical Statistics." Biometrika, vol. xiii. pp. 1-16, 1920.
[On Inverse Probability.]
"Notes on the History of Correlation." Biometrika, vol. xiii. pp. 25-45, 1920.
"The Chances of Death" and other essays. 2 vols. 8vo, London, 1897.

The Grammar of Science. London, 1892.

Peirce, C. S. "A Ineury of i'rubable lnference." Johns Mophins "studies in Logic," 1883.
"On an Improvement in Boole's Calculus of Logic." Proc. Amer. Acad. Arts and Sci. vol. vii. pp. \(250-261,1867\). Pp. 62. Camb., 1870.
Perozzo. "Nuove applicazioni del calcolo delle probabilità allo studio dei fenomeni statistici." Proceedings of Academia dei Lincei, 1881-82.
 rechnung in der Statistik. P'p. 33. 4to. Dresden, 1883.
 Méta. et de Mor. vol. x. pp. 68:-693, 1902.
Pinard, H. "Siur la Convergence des Probabilités." Rev. Néo Schol. de Phil. No. 84 (1919) and No. 85 (1920).
Pincherle, S. " Il calcolo delle probabilità e l' intuizione." Scientia, vol. xix. pp. 417-426, 1916.

Pizzetti, P. I fondamenti matematici per la critica dei risultati sperimentali. Atti della R. Univ., Genova, 1892.
Plaats, J. 1. van der. Uver de toepassing der waarschijnlijkheidsrekening op medische statistick. 1895.
Plana, (\&. "Mémuire sur divers problèmes de probabilité." Mémoires de l'Académie de Turin for 1811-12, vol. xx. pp. 355-408, 1813.
l'uincare, H. Calcul des probabilités. P'p. 274. Paris, 1896. 2nd edition (with additions). 1'1. 333. l'aris, 1912. S'cience et hypothèse. Paris. Eingl. transl., London, 1905. Science et méthode. Paris. (Includes a chapter on "Lo Hasard.") Eng. transl. (by F. Maitland). Pp. 2s8. London, 1914.
"Le Hasard." Rev. du Mois, March 1907.
Porsson, S. D. Recherches sur la probabilité des jugements en matièe eriminelle et en matière civile, précídées des règles générales du calcul des probabilites. 4to. Pp. ix +415 . Paris, \(18: 37\).

Lehrbuch der Wahrscheinlichkeitsrechnung. German translation of the above by H. Schnuse. Braunschweig. 8vo. \(18 \pm 1\).
"Sur la probabilité des résultats moyens des observations." Comn. des Temps. P'p. 273-302, 1827. I'p. 3-22, 1832.
"Formules relatives aux probahilités qui dépendent de très grand numbres." Compt. Rend., Acad. P'aris, vol. 2, pp. 603-613, 1836.
"Sur le jeu de trente et quarante." Annal. de Ciergonne, xv.
"Solution d'un problème de probabilité." Liouv. J. (1), vol. 2, 1837.
"Mémoire sur la proportion des naissances des filles et des garcons." Mém. Acad. Paris, vol. 9, pp. 239-308, 1830.
Pondra et Hossard. Question de probabilité résolue par la géométrie. Svo. Paris, 1819.
Poretzki, Platon. S. "Solution of the general Problem of the Theory of Probability by means of Nathematical Logic." (In Russian.) Bull. of the physico-mathematical Academy of Kasan, 1887.
Prevost, P . "Sur les principes de la théorie des gains fortuits." Nouv. Mém. Pp. 430-472. Berlin, 1780.
Prevost, P., and Lhuilier, S. A. "Sur les probabilités." Mém. Ac. Berl. (1796), pp. 117-142, 1799.
"Sur l'art d'estimer la probabilité des causes par les effets." Mém. Ac. Berl. (1796), pp. 3-24, 1799.
"Remarques sur l'utilité et l'étendue du principe par lequel on estime la probabilité des causes." Mém. Ac. Berl. (1796), pp. 25-41, 1799.

Note on last. Mem. Ac. Berl. (1797), p. 152, 1800 .
"Mémoire sur l'application du calcul des probabilités à la valeur du témoiguage." Mém. Ac. Berl. (1797), pp. 120-151, 1800.

Price, R. See Bayes.
Pringsheim, A. See Daniel Blirnoulli.
"Wreiteres zm Geschichtedes Detershurger Problems." (irmmert, Ar.hir, 77, 1881.
Proctor, R. A. Chance and Luck. A Discussion of the Laws of Luck, Coincidences, Wagers, Lotteries, and the Fallacies of Cambling, with Notes on Poker and Martingales. Pp. vii +263 . London, 1887.
Protimalethes. Miracle versus Nature : being an Application of Certain Propositions in the Theory of Chances to the Christian Miracles. 8vo. Cambridge, 1847.

Quetelet, A. Instructions populaires sur le calcul des probabilités. 12 mo . Bruxelles, 1828.

Engl. transl. : Popular Instructions on the Calculation of Probabilities, transl. with notes by R. Beamish. 1839.

Dutch transl. by H. Strootman. Breda, 1834.
Lettres sur la théorie des probabilités appliquée aux sciences, morales et politiques. Bruxelles, 1846.

Engl. transl. : Letters on the Theory of Probabilities as applied to the Moral and Political Sciences, transl. by O. G. Downes. 8vo. 1849.
"Sur la possibilité de mesurer l'influence des causes qui modifient les élémens sociaux." Corresp. mathém. et phys. vol. vii. pp. 321-346. Bruxelles, 1832.
"Sur la constance qu'on observe dans le nombre des crimes qui se committent." Corresp. mathém. et phys. vol. vi. pp. 214-217. Brussels, 1830.
"Théorie des probabilités." (In the Encycl. populaire.) Brussels, 1853.
"Sur le calcul des probabilités appliqué à la science de l'homme." Bull. de l'Acad. roy. vol. xxvi. pp. 19-32. Brussels, 1873.
[For a full bibliography and discussion of Quetelet's writings on these topics see Lottin's Quetelet.]

Rayleigh, Lord. "On James Bernouilli's Theorem in Probabilities." Phil. Mag. (5), vol. 47, pp. 246-251, 1899.
Regnault. Calcul des chances et philosophie de la bourse. 8vo. Paris, 1863.

Renouvier, Ch. L'Homme: la raison, la passion, la liberté, la certitude, la probabilité morale. 8vo. 1859.
Revel, P. Camille. Esquisse d'un système de la nature fondé sur la loi du hasard. 1890. 2nd ed. (corrigée), 1892.

Le Hasard, sa loi et scs conséquences dans les sciences et en philosophie. Paris, 1905. 2nd ed. (corrigée et augmentée). Pp. 249. Paris, 1909.
Rizzetti, J. "Ludorum scientia, sive artis conjectandi elementa ad alias applicata." Act. Erud. Suppl. vol. 9, pp. 215-229, 295-307. Leipzig, 1729.

Roberts, Hon. Francis. "An Arithmetical Paradox concerning the Chances of Lotteries." Phil. Trans. vol. xvii. pp. 677-681, 1693.
Roger. "Solution d'un problème de probabilité." Liouv. J. (1), vol. 17, 1852.

Rouse, W. Doctrine of Chances, or the Theory of Gaming made easy to cvery Person-Lotteries, Cards, Horse-Racing, Dice, etc. 1814.
Rudiger, Andreas. De sensu falsi et veri libri iv. [Lib. i. cap. xii. et lib. iii.] Editio Altera. 4to. Lipsiae, 1722.

Ruffini. Critical Reflexions on the Essai philosophique of Laplace (in Italian). Modena, 1821.
das Schiessen und auf die Theorie des Einschiessens. Stuttgart, 1906.
Sawitscif, A. Die Anwendung der Wahrscheinlichkeitstheorie auf die Berechnung der Beobachtungen und geodätischen Messungen oder die Methode der kleinsten Quadrate. (Translated into German from the Russian by Lais.) Leipzig, 1863.

strater, H. Vid. Poram

Scott, Johs. The Doctrine of Chance : the Arithmetic of Gambling. 56 pp . Svo. 190 s.


Smbidon, W. II. "Chance." Journal of Phil., Psych., and Sci. Meth., vol. ix. pp. 281-290. 1912.

Simepeari), W. F. "On the Application of the Theory of Error to Cases of Nurmal Distribution and Ňormal Correlation." Phil. Trans. A. vol. 192, pp. 101-167, 1899.
"On the Calculation of the most Probable Values of the Frequency Cimstants for Data arranged according to Equidistant Divisions of a Scale." Proc. Lond. Math. Soc. vol. xxix. pp. 353-380.
"Normal Correlation." Camb. Phil. Soc. vol. xix.
"Nurmal Distrihution and Correlation. Roy. Soc. Trans., 1898.
su;wart, C. Review of von Kries in Vierteljahrsschr. für Wiss. Mhil. xiv. p. 9).

Logik. Tübingen, 1878.
2nd ed. Freiburg, i. B., 1893. English ed., 1895.
Yol. ii. Part 3, chap. 3, § 85, Die Wahrscheinliehkeitsrechnung; 5, § 102. Die Wahrscheinlichkeit auf statischem Boden.

Roferences in English ed. :
Probability, vol. ii. pp. 216-230, 261-271 (errors of observation), 303 . 309 (induction), 504-507 (statistics).
Smmon: T. (". "A New Theorem in Probability." Proc. Lond. Math. Sowe. wol. \(26, \mathrm{pp} .29(-323,1895\).
"Sur la probabilité des événements composés." Ass. Franc. pour PAvancement des Sciences. 1896.
 Actuaires français, i.
Givison, T. "A Letter to the Rieht Honourable (ienorge, Farl of Macelesfield, President of the Royal Society, as to the Advantage of taking the Mean of a Number of observations in Practical Astromomy." Phil. Trans. vol. Niix. pp.s.2.93, 1755.
" In Attermet to show the Advantage arising by taking the Mean of a Nomber of ()hservations in Practionl Astronomy." (Miscellaneous tracts on come curious subjects, pp. 61.75). London, 4tn, 1757.
[ I reprint of the above with some new matter. The probability, assumine prositive and newative errors to be equally likels, that the mean is nearer to the truth than a single whereration taken at randem, is here incestigated for the first time. 7

Treatise on the Nature and Laws of Chanee. 4to. London, 174!
Anotherentition. 8vo. 1792.
 xxiii. pp). 50-656, 1857.

 wohlf ile Ausg. 4to. Braunschweif, 1840.

Spinoza. "Letter to Jan van der Meer." Opera ed. Van Vloten and Land, vol. ii. pp. 145-149, Ep. 38 (in Latin and Dutch).

See also Spinoza's Briefwechsel in J. H. v. Kirchmann's Philos. Bibliothek, vol. xlvi. pp. 145-147.
Sprague, T. B. On Probability and Chance and their Connexion with the Business of Insurance. 8vo. 1892.
Stamkart, F. J. Over de waarschijnlijkheidsrekening. 8vo.
Sterzinger, O. Zur Logik und Naturphilosuphie der Wahrscheinlichkeitslehre. Leipzig, 1911.
Stemart, Dugald. "On the Calculus of Probabilities, in reference to the Preceding Argument for the Existence of God, from Final Causes." Philosophy of the Moral Powers, vol. ii. pp. 108-119. (Sir W. Hamilton's ed., Edin., 1860.) 1st ed., 1828.
Stieda, L. Über die Anwendung der Wahrscheinlichkeitsrechnung in der anthropologischen Statistik. Arch. f. Anthrop., 1882.

2nd ed. 8vo. Braunschweig, 1892.
Streeter, T. E. The Elements of the Theory of Probabilities. 31 pp. 8vo. 1908.

Strove. Catalogus novus stellarum duplicium et multiplicium. Dorpati, 1827, pp. xxxvii-xlviii.
Stumpr, C. "Bemerkung zur Wahrscheinlichkeitslehre." Jahrb. f. national. Ök. u. Stat. (3), vol. 17, pp. 671, 672, 1899 ; vol. 18, p. 243, 1899.
[In criticism of Bortkiewicz, q.v.]
Stumpf, K. "Über den Begriff der mathematischen Wahrscheinlichkeit." Ber. bayr. Ak. (Phil. Cl.), pp. 37-120, 1892.
"Uber die Anwendung des mathematischen Wahrscheinlichkeitsbegriffes auf Teile eines Continuums." Ber. bayr. Ak. (Phil. Cl.), pp. 681691, 1892.
Suppantscnitscm. Einführung it die Wahrscheinlichkeitsechnung. Leipzig.
Tatr, P. G. "Law of Frequency of Error." Edin. Phil. Trans. vol. 4, 1865.
On a Question of Arrangement and Probabilities. 1873.
Tohebycief, P. L. Essai d'analyse élémentaire de la théorie des probabilités. 4to. Moscow, 1845 (in Russian, degree thesis). Pp. ii \(+61+\) iii.
"Démonstration élémentaire d'une proposition générale de la théorie des probabilités." Crolle J. vol. 33, pp. 259-267, 1846.
"Des valeurs moyennes." Liouv. J. (2), vol. 12, pp. 177-184, 1867.
(Extrait du Recueil des Sciences mathématiques, vol. ii.)
"Sur deux théorèmes relatifs aux probabilités." Petersb. Abh. vol. 55, 1887. (In Russian.) French translation by J. Lyon: Act. Math. Petr. vol. 14, pp. 305-315, 1891.

Euvres. 2 vols. 4to. St-Pétersbourg, 1907.
(The three memoirs preceding are here reprinted in French.)
Terrot, Bishop. "Summation of a Compound Serics and its Application to a Problem in Probabilitics." Edin. Phil. Trans., 1853, vol. xx. pp. 541-545.
"On the Possibility of combining two or more Probabilities of the same Event, so as to form one Definite Probability." Edin. Phil. Trans., 1856, vol. xxi. pp. 369-376.
Thiele, T. N. Theory of Observations. Pp. \(6+143\). 4to. London, 1903.
Thomson, Archbishor. Laws of Thought. § 124, Syllogisms of Chance (13 pp.).
Thubeuf. Élémens et principes de la royale arithmétique aux jettons, etc. 12mo. Paris, 1661.
Tmprding. Die Analyse des Zufalls. Pp. ix + 168. Braunschweig, 1915.
Todilunter, I. "On the Method of Least Squares." Camb. Phil. Trans. vol. ii.
A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace. Lge. 8vo. pp. xvi +624, Camb. and Lond., 1865.

Tozer, J. On the Measure of the Force of Testimony in Cases of Legal Evidence. 4to. Camb. Phil. Soc. vol. viii. Part II. 16 pp. (read Nov. 27, 1843). 1844.

Trombley. "Observations aur le calcul d'un jeu de hasard." Mem. A". Bierl. (1802), pp. 86-102.
"Recherches sur une question relative au calcul des probabilités." Mém. Ac. Berl. (1794-5), pp. 69-108, 1799.
(On Euler's memoir, "Solutio quarundam quaestionum difficiliorum in calculo probabilitatum.")
"Do probabilitate causarum ab effectibus oriunda." Comm. Soc. Reg. Gott. (1795-8), vol. 13, pp. 64-119, 1799.
"Observations sur la méthode de prendre les milieux entre les observations." Mém. Ac. Berl. (1801), pp. 29-58, 1804.
"Disquisitio elementaris circa calculum probabilium." Comm. Soc. Reg. Gott. (1793-4), vol. 12, pp. 99-136, 1796.
Tschuprow, A. A. "Die Aufgaben der Theorie der Statistik." Jahrb. f. gesetzg. Verwalt. u. Volkswirtsch. vol. 29, pp. 421-480, 1905.
"Zur Theorie der Stabilitat statistischer Reihen." Skandinavisk Aktuarietidskrift, pp. 199-256, 1918; pp. 80-133, 1919.
Twardowski, K. "Über sogenannte relative Wahrheiten." Arch. f. syst. Philos. vol. viii. pp. 439-447, 1902.

U'Ress, F. M. " C̈mer den Bexriff der mathematistuen Wahrsheintichkeit "


V'asties, L. G. F. L'Art de conjecturer. Traduit du latin do J. Bernoulli, avec observations, éclaircissemens ct additions. Caen, 1801.
[Translation of Part I. only of Bernoulli's Ars Conjectandi (q.v.) containing a commentary on and reprint of Huygens, De ratiociniis in ludo aleae.]
Vexs, J. The Loric of Chance. 1866. 2nd ed., 1876. 3rd ed., 1888.
"The Foundations of (hance." Princeton Rev. vol. 2, pp. 471-510, 1872.
" On the Nature and Lses of Averages." Stat. Journ. vol. 54, pp. 42941. 1 1 : 41.

Wagsbr, A. Jie (iesetzmässigkeit in den scheinhar willkürlichen Handlum in des Monathen. Hanlant. Astil.
 d. ges. Versicherungswissensehaft. Berlin, 1906.
 Quantities into Probable Rolations and Annuities, etc. I'p. 59. C'am(1010) 1732.

An Essay on the Principles of Human Knowledse. I'p. 244. Cambridue, 1794.
Wrfatos, J. Mannal of Logic. (Probability, vol. ii. pp). 185-185.) Lom1ton. 1-40:
Wistercaakd. Girundziige der Theorie der Statistik.
 vol. x. 191 t.
 actions of the Farulty of Actuaries in Scotland, vol. viii. (1920), pp. \(1633 \cdot 206\).
[Problems of Inverse Probability including the Iaw of Nuccession. This paper is followed ty others on the same subject by various writers.]

Whitwortir, W. A. Choice and Chance, An Elementary 'Ireatise on Permutations, Combinations, and Probability, with 300 Exercises. 1867. 2nd ed., 1870. 3rd ed. pp. viii +244 . Cambridge, 1878.

Expectations of Parts into which a Magnitude is divided at Random. 1898.

Wicksell, S. D. "Some Theorems in the Theory of Probabilitics." Skandinavisk Aktuarietidskrift, p. 196 (1910).
Wignne, H. A. De leer der waarschijnlijkheid in hare toepassing op het dagelijksche leven. 1862.
Wilbraham, H. "On the Theory of Chances developed in Prof. Boole's 'Laws of Thought.'" Phil. Mag., 1854.
Wild, A. Die Grundsätze der Wahrscheinlichkeitsrechnung und ihre Anwendungen. München, 1862.
Windelband, W. Die Lehren vom Zufall. Berlin, 1870.
Wolf, A. "The Philosophy of Probability." Proc. Arist. Soc. vol. xiii. pp. 29. London, 1913.
Wolf, R. "Über eine neue Serie von Würfelversuchen." Vierteljs. Naturforsch. Gesellsch. in Zürich, vol. 26, pp. 126-136 and 201-224, 1881; vol. 27, pp. 241-262, 1882; vol. 28, pp. 118-124, 1883.
"Neue Serie von Würfelsversuchen." Ibid. vol. 38, pp. 10-32, 1893.
"Versuche zur Vergleichung der Erfahrungswahrscheinlichikeit init der mathematischen Wahrscheinlichkeit." Mitth. d. Naturforsch. Gesellsch., Bern, 1849-1851, 1853.
Wolff, Christian. Philosophia rationalis sive logica. Leipzig, 1732.
Woodward, R. S. Higher Mathematies, chap. x. pp. 467, 507. "Probability and Theory of Error." New York, 1900.

Probability and Theory of Errors. New York, 1906.
Wyrouboff, G. "Le Certain et le probable." La Philos. posit. p. 165, 1867.
Young, J. R. Elementary Treatise on Algebra, Theoretical and Practical, with an Appendix on Probabilities and Life Annuities. 4th ed. enlarged, post 8vo. 1844.
Young, Rev. M. "On the Force of Testimony in establishing Facts contrary to Analogy." Trans. Roy. Ir. Acad. vol. vii. pp. 79-118, 1800.
Young, T. "Remarks on the Probabilities of Error in Physical Observations, etc." Phil. Trans., 1819.
Yole, G. U. "On the Theory of Correlation." Journ. Stat. Soc. vol. Ix. p. \(812,1897\).
"On the Association of Attributes in Statistics." Phil. Trans. (A), vol. 194, pp. 257-319, 1900.
"On the Theory of Consistence of Logical Class-frequencies." Phil. Trans. (A), vol. 197, pp. 91-132, 1901.

An Introduction to the Theory of Statistics. Pp. xiii. +376. London, 1911.

Yule and Galton. "The Median." Stat. Journ. pp. 392-398, 1896.

\section*{IN1HE}

Acquaintance，direet，12
Addition，of probabilities， 37,135

T：Imentar of，114，1こ！，111

Analogey，principle of，iss
and induction，218，222
negative， \(219,223,233\) ， \(41 \%\)
positive，220，223， 415
and wencralisation，223
lurical foundation of，2．5
and Bacon，2fs
and Leibniz，ごン
and Jevons， 273
and statistics， \(391,407,415 \mathrm{f}\) ．
A！rilluit，\(\therefore\) ．．．．－
Apprehension，direct，and cthimal juder－ ment，315；
Arcument， 13

and indurtion， 27.4
Arithmetic mean（or averace），205
and laws of arror， 197
Iaplare on， 206
Gatuss on， 206

Asymmetry，and Bernoullis Theorem， 358 f ．

Avorager，20．5 f．


Axioms， 135 f ．
non－self－evident，29！）
```

1:a in|ll . .3:%

```

```

    and Emale of Sucer-vion, 37, % m
    1:=..."H, 24...l.
4.1.1, %, -4.4
and limited varicty, 270

```

Bayes，and Inverse Probability， 174 Theorem of， 379
Brlief，rational， 4 f．，10，16， 307 deforew of， 11
Bentham，measurement of P＇oba－ bility，20
Bernoulli，Daniel，and Inverse Proba－ bility， 174


Bermoulli，Jac．， 15 n．，41，76，81，83， 86，368， 369
weirht of evidence，31：3
second axiom of， \(32 \because\)
and regular frequeney，33：3
and statistical serice， 392
Bernoulli＇s Theorem，luy，314，319 n．， \(3: 33,337 \mathrm{f}\) ．
and asymmetry， 358 f ．
empirical verification of， \(3(6] \mathrm{f}\) ．

1．－1t．mat．\＆－．． 49
on multiplication，l3is
and Maxwell，17ご。
and independence， \(17: 3\)
and Law of Error，20）\(n\) ．
and chance，2st
and P＇etershurg Paradox，317
and Bromoulli＇s＇Theorem，33！
and Rule of Succession，3s：

Boshek and Rule of Fuecession， \(3 x: 3\)
Bude＇s Law，30．4

and（iepman lowicians， 87
and relation of Probabilit y，！n！
aul mymbelie probability，1．5
and approximation， 161
and independence， \(16{ }^{-7}\)
．101 IS thatls．II：
and combination of premisses， 179


Boole (contd.) -
and testimony, 180
and Challenge Problem, 187
and Cournot, \(284 n\).
and Rule of Succession, 382
Borel, 47 n., 48
Bortkiewicz, von, and great numbers, \(333 n\).
and Marbe, 365 n.
method of, 384
and Lexis, 393 f .
and Law of Small Numbers, 401 f .
and Quetelet, 402
Boscovitch and Least Squares, 210
Bowley, 421, 423, 424 f.
Bradley, \(319 n\).
- and relativity of Probability, 91
and Bernoulli's Theorem, \(341 n\).
Broad, C. D., 257 n.
Brömse and Marbe, \(365 n\).
Brünn and lotteries, 364
Bruns and Marbe, 365 n .
Buffon, 317, 322
and coin-tossing, 362
Butler, Bishop, 79, 80, 309, 310
and risk, 321
Calculus of Probability, \(83 n\)., 149, 164, 303, 428
and Psychical Research, 302
and Sociology, 335
' Casual,' 288
Causaiity, 263, 276
and independence, 164
'Cause,' 275
Cause, final, 297
Cayley, and tradition, 185
and Challenge Problem, 187
Certainty, 10, 127, 128
and truth, 15
Kahle and, 90 n.
definition of, 120
relation of, 134
and Bacon, 267
and Leibniz, 272
Chance, objective, 281, 286 f., 295, 418
Couturat on, 283
Poincaré on, 284, 289
Condorcet on, 284
definition of, 287
and planets, 293
and binary stars, 295
Coefficient of Credibility, 183
of Correlation, 421 f .
Combination of premisses, 149, 178

Comte, and ' seven,' 246
and statistics, 335
Condorcet, 83 n., 317
and testimony, 180
and chance, 282, 284
and ethics, 313,316
and gambling, 319
Conduct and Probability, 307
Consistence and group theory, 124
Contradiction, 143
Coover, J., 298 n.
Correlation, 329, 390
and statistical frequency, 330
Quantitative, 391, 426
Inductive, 406
enefficient, 21 i .
Cournot, and frequency theory, 92
and independence, 166
on testimony, 180
and causality, 275
and chance, 282,283
Couturat, \(272 n\)., \(311 n\).
Craig and tradition, 184
Cramer and Petersburg Paradox, 318
Crofton, \(47 n\).
Cumulative Formula, 150
Johnson and, 121
Czuber, 47 n., 78, 82, 86, 339 n., 345 n., 347
and symbolic probability, 156
and 'cause,' 275 n.
and risk, 315 n .
and Bernoulli's Theorem, \(340 n\)
and statistical frequency, 351,394
and Tchebycheff's Theorem, 353 n ., 355 n .
and verification of Bernoulli, 362 n .
and lotteries, 364
and Marbe, 365
and Inverse of Bernoulli's Theorem, 370 n .
and Rule of Succession, 376 n ., 382
J)'Alembert, 82, 170 n. \(321,365 n\)., 369
and chance, 282
and planets, 293
and mathematical expectation, 314
and ethics, 316
and Petersburg Paradox, 317
and Marbe, 365
Darbon, A., and Cournot, 284
Darwin, 108
and Lyell, 161
and Mill, 265
Dedekind and 'Challenge Problem,' 187 n .

\section*{Definitions， 134 f．}
summary of， 120
de la Placette，Jean，and chance， 283
De Morgan，21，74， 83
and inference， 139
and independence， 168
and Inverse Probability， 178
and combination of premisses， 179
and tradition， 184 n ．
and planets， 293
pupil of， 362
and Inverse of Bernoulli＇s Theorem， 370 n ．
and Rule of Succession，375，382
De Witt and arithmetic averages， 206
Dice－tossing， 361 f ．
Diderot on testimony， 183
Discordant observations，rejection of， \(21: 3\)
1）いにin，W．F．．．2い
and Inverse Probability， 176
Dormoy， 394
F．14．3nrth，29 \(2 ., 84,85,362 \mathrm{n}, 339\), 400

and randomness， 290
and Psychieal Research， 298 n ．
and ethics， 316
and German statisticians， 394
Ëgrenberger，340 \(n\) ．
Ellis，Leslie，84，85
and frequency theory， 92
and Least Squares， 207 n ．， \(2(99\)
 \(271 \mathrm{n}, 274 \mathrm{n}\) ．
and Bernoulli＇s Theorem， 341
Empirical School，85， 86
Epistemology，3（1）
and inductive hypothesis， 261
Equiprobability， \(41,63,65\)
Equivalence，definition of，120， 134
axiom of， 135
principle of， 141
Error，probable， 329
Ethics， 307 f．

Event，probability of，\(\overline{5}\)
Fividence，and measuremont of Prob－ ability，7， 35
relevant and irrelevant， 53,54
independent and complementary，is external， 57
addlition of，66，68
weight of， 71
and Induction，2：21

Excluded Middle，Law of， 143
Experience and the Principle of Indifference， 100

Fechner，and median， 201
and law of sensation，208
and lutteries， 364
Fermat，formula of，：242
Forbes，J．I）．， 20 n．，21， \(294 n\).
Frazer，Sir J．， 245
Frequency curves， 199 and statistics， 328
Frequency，statistical，331）
Frequency theory， 92 f ． and randomness， 290 and Bernoulli＇s Theorem，344
and Rule of Succession， 378
F＇resnel and simplicity， 206
Fries， \(15 n\) ．
Galton， 321
and Fechner＇s law，20s
（：ambling，31！）
（iauss，and laws of error， 196 n．， 198
and arithmetic mean， 206
and Least Squares， 210
Generalisation， 389
definition of， \(2: 2.2\)
from statisties， 328
Generator properties， 253
plurality of， \(254,256,257\)
Geometrical probability，47，6：
German logicians， 87
（iibbon，29，32：2， 333
（iilman，B．I．，and symbolic prob）－ ability， 156
Goldschmidt， 29 n ．
（ioodness，organic nature of，310
（ifunat，3！－9？
Great Numbers，Law of，82，330， 333 f ．
（ireville，Fulke， \(46 t ;\)
（irimschl， 248 n ．
and Marbe， 365 n ．
（iroups，of propositions， 117,124
definition of， 120,125
real and hypothetical，1：39


and diseordant observationts， 21411 ．
Halley and mortality statist ios， \(33^{2} 2\)
Herodotus， 307
Herschell and binary stare， 294
Houdin， \(364 n\) ．
Hudson，W．H．，and animism， 2.47 n ．
Hume， \(52,70,80,81,82,83,239,427\) and testimony， 182

Hume (contd.) -
fand Induction, 218, 233, 265, 272
land analogy, 222, 224
and chance, 282
Huyghens, 82
and 'six,' 247
Hypothesis, 7
Hypothetical entities, 299

\section*{Implication, 124}

Impossibility, 15
definition of, 120
relation of, 134
Inconsistency, definition of, 120
Independence, for knowledge, 107, 165
definition of, 120,138
Theorem of, 121, 146
of events, 164
and law of error, 195
and measurement, 204
and averages, 212
and discordant observations, 214
and chance, 283
Index numbers, 211
' Induction,' 274
Induction, 97
Principle of, 68
and frequency theory, 98, 99, 107
and Logic, 217
pure, 218
universal, 220, 406, 417
validity of, 221
and statistics, 327 f .
statistical, 416 f .
Inductive corrclation, 220, 257, 2. 2s, 392, 397, 40 ;
Inductive hypothesis, 260, 264
Inductive method, 260
Inference, 129
necessary, 120, 139
hypothetical and assertoric, 130
statistical, 327 f .
Insurance, 22, 285, 404
Intuition versus experience, 86 and ethical judgment, 312
Inverse Probability, 149, 174
and Venn, 100
and frequency theory, 106
Theorem of, 121
and statistics, \(369,370 \mathrm{n}\).
and Bowlev, 425
Irrelevance, 255
judgments of, 54
definition of, 55, 120, 138
Theorem of, 121, 146

James, W., and spirits, 301
Jesuits, 308
Jevons, \(244 n\) 。
and equiprobability, \(42 n\).
and Inverse Probability, 178
and index numbers, 212
and Induction, 222, 238, 243, 265, 273, 274
and analogy, 246
and coin-tossing, \(360^{2}\)
and Rule of Succession, 382
Johnson, W. E., 116
and propositions, \(11 n\).
and added evidence, 68
and cumulative formula, 121, 150, 153, 155
and groups, 124
and testimony, 183
Judgments, 54
of preference and relevance, 65
direct, 70
disjunctive, 77
Kahle and the Probability relation, 90
Kant, 333
and Hume, 272
Kapteyn, Prof. J. C., and law of error, 199
Knowledge, 10
kinds of, 3, 4
direct and indirect, 12, 26:
incomplete and proper, 13
of logical relations, 14
probable and vague, 17
relativity of, 17
vague and distinct, 53
homologic and ontologic, 276,288
and ignorance, 281
and chance, 289
Fifice, vom, \(4 \because\), it \(\mu, 45 \%, 46 \mu\), 50 , \(67 \mu, 84\)
and equiprobability, 87
and Principle of Indifference, 172
and independence, 173
and Inverse I'mbability, 171;
and knowledge, 276
and Cournot, \(284 n\).
and School of Lexis, 394
Lacroix, 184 n .
Lambert and Least Squares, 210
Lämmel, \(47 n\).
and symbolic probability, 156
Laplace, 15 n., 28 n., 31, 82, 83, 84, 318, 427
school of, 44, 51, 86, 358, 365

Litplue al (in).
and minton a! Powhbillis. al
and ind

and testimony, \(\mathrm{i} 8(1), 182\)
and doctrine of averages, \(2\left(\begin{array}{l}2 \\ \hline\end{array}\right.\)
and arithmetic mean, 206
ath Lenstriquarn..:In
and Induction, 220, 239, 265, 27:3
and chance, 282
and planets, \(293 n\).
and Quetelet, 334
and Bernoulli's Theorem, 340, 341, 370
and Rule of Succession, 3.51 n ,

and birth proportions, 364
and unknown probabilities, 370
an: Is y.... Thimorem. ash
and statistical series, 392
Laurent and gambling, 319
1aw, 311 n .
Law of error, 194 f .
and arithmetic mean, 197
and peometric mean, 198
and median, 200
and mode, 203
normal law, 199, 202, 205

Least squares and Venn, 206
method of, 202, 205, 206, 2019
Lee and tradition, \(184 n\).
Legendre and Least Squares, 210
Leibniz, 24 n., 308, 368, 392, 427 and arithmetic average, 206 and Induction, 272
Lexis, and asymmetry of statistical frequency, 359 \(n\).
and Marbe. 345 n .
method of, \(384,393 \mathrm{f}\)., 397 f .
and Edgeworth, 401
and tatiotical stabilits 11.. 10...
locke. \(76,80,82,83,308,323\)
on tradition, 184
and weight of evidence, 313
logtic, academic, 3
of probability, 8
of implication, 58
and Induction, 217, 24.5
and initial probability, 299
Lagimal printity, 123
1otteries, 333 n., 361,364 f. published results of, 363
\(1 . \cdots\)
and Rule of Sucression, 382
Lucretius, 427

M'Alister, Sir Donald, and laws of etror, 198
Macaulay and Bacon, 266
Mocioll, and symbolic probability, 15.5
and Boole, \(167 n\).
and Inverse Probability, 176
and 'Challenge Problim,' 188 n .
Macfarlane, and independence, 169 n . and tradition, \(18 . \pi\)
and 'C'hallenye P'roblem,' 187 n .
Maclaurin, Theorem of, 207
Marbe, Dr. Karl, and roulette, 36.5
Narginal utility, 318
Markoti, A. A., 1IT ..
and Inverse Probability, 176
and Tchebycheff's Theorem, 3.77
Mathematical Expectation, 311, 315, 316
Mathematicians, and probability, 84
and cumulative formula, 152
and laws of error, 207
and ethics, 316
Maxim, Sir Hiram, \(364 n\).
Maxwell, 172 n .
and theory of gases, 172
Mayer and Least squares, 210
Heans and laws of error, 194 f .
Measurement of Probability, 34, 1.58, 311
and frequency theory, 94
and induction, 259, 388
and psychical researeh, 302

Median and laws of error, 200
Meinong, 78
Meissner, Othe, and dice-throwing. 36.3
Memory, 14
Mendelism and statistics, \(335,419\). 125
Merriman, Mansficld, and Least Squares 209
Metaphysics and certainty, 2339
Nothod of Difference, 246
Michell, 302

and linary stars, 294
Middle Term, Fallacy of, 18:, 15.5
Mill, and inductive corrolations, a2. and induction, 26.5 f .
and plurality of causes, 267 n .

and pure induction, 209
methods of, 270
and limited variety, 271
Moulality and probability, \(16 n\).

Modality (contd.) --
Venn and, 98
Mode, and law of error, 203
asymmetry about, 361
Monte Carlo, 364
Moore, G. E., 19, 240 n., 309
Morgan, vide De Morgan
Multiplication, 135
definition of, 120
thenrems of, 121, 148, 342
of instances, 233 f .
Munro, 370 n .
Necessary connection, law of, 251
Newton, and induction, 244
and 'seven,' 247
and Bacon, 265
Nitsche, A., 45 n., \(50 n ., 78,172 n\).
Occurrences, remarkable, 302
Pascal, 82
Pearson, Karl, 84, 351 n. and frequency theory, 100
and arithmetic mean, 208
and stars, 297
and asymmetry, 347, \(359 n\).
and generalised Probability curves, 347
and roulette, 364
and Rule of Succession, 379, 382
Peirce, \(50 n\)., 304
and randomness, 290
Petersburg Paradox, 316
psychology of, 318
and Buffon, 362
Peterson and tradition, 184
Physics and initial probability, 299
Planets, movements of, 293
Playfair, Dr. Lyon, 305
Plurality of causes and Mill, 267
Poetry and statistics, 401
Poincaré, Henri, 48, 84
and independence, 173
and chance, 284, 289
Poisson, 51 n., 362 n.
on testimony, 180
and least errors, 207
and Petersburg Paradox, 317
and gambling, 319
and great numbers, 333, 336
Theorem of, 344
and statistical frequency, 348
and Tchebycheff, 357
and inverse of Bernoulli's Theorem, 370

Poretzki, Platon S., and symbolic probability, 157
Port Royal logic, 70, 80, 321
and probabilism, 308
Prediction, value of, 305
Price and Bayes, \(174 n\).
Primitive people and rational belief, 245
Principle of compelling reason, 86
Principle of Indifference, 42, 81, 83 f., 87, 104, 107, 171
analysis of, 53
modification of, 55, 58
and induction, 99
and measurement, 160
and Psychical Research, 302
and ethics, 310
and statistics, 367
and Laplace, 372, 374
and Rule of Succession, 377
Principle of Non-Sufficient Reason, 41, 85
Principle of superposition of small effects, 249
Probabilism, 308
'Probability,' 8 Venn's use of, 95
Edgeworth's use of, \(96 n\).
Probability, and relevant knowledge, 4
objective relation of, \(5,8,281\)
mathematical, 6
dependent on evidence, 7
philosophical definition of, 8
three senses of, 11
measurement of, 20 f., 37 and law, 24
and similarity, 28, 36
comparison of, 34, 66, 160
series of, 35, 38
'geometrical,' 47, 48, 62
and rational belief, 97
and statistical frequency, 98
and truth frequency, 101 f ., 337 f .
Inverse, 106, 149
and truth, 116, 322
negative, 139
finite, 237
and randomness, 291
and planetary orbits, 293
and binary stars, 294
and star drifts, 296
and final causes, 297
and spirits, 300,301
and telepathy, 300
and ethics, 307

Probability (romed.)
from statistics, 367 f .
'unknown,' and Laplace, 372
Probability relation, 4, 8, 13, 134 intuition of, 52
Probable error, 74
Prowtor, 3it!".
Proposition, characterisation of, 3, 4 primary and secondary, 11, 13
knowledge of, 12
self-evident, 17
classes of, 101 f .
groups of, 117,124
sub-groups of, 126,129
disjunction and conjunction of, 134 synthetic, 26;3
existential, 276

and induction, \(2 \because 2\)
and randomness, 291
P'schical Research, 278 f.
Puychology and probability, 52
Pythagoras and 'seven,' 246
Quctelet, 333 n., \(334,335,401,415\), 427,428
and arithmetic man, 2018
and balls, 362
and statistical stability, 393
Randomness, 2s1, 290,412
Pearson's use of, 297
Relation, of probability, 6 of 'between,' 35, 39
Relativity, of knowledge, 17 of probabilities, luz
doctrine of, and the Law of Cniformity, -48 n .
Relevancr, judgments of, 54

theoroms of, 147
Remarkableness, 30:
Requirement, 1:9
Risk, 315
and ethies, 313
and Petersburg Paradsx, 319

' pherim!. 3.2.
Roulette, 361, 3654
publishod results of, \(36: 3 \mathrm{n}\).
Rule of Succession, 359 n., 368, 372, 374
proof of, 375
and frequency theory, 378
and Pearson, 3811 n .


\section*{Russell, Bertrand (contd.)}
and inference, 117
and implication, 1:4
Schematisation, 67
Schröder and symbolic probability, 157
Selection, random, 292
series of probabilities, 35,38
and frequency theory, 93
independent, 283, 420
organic, 399, 420
Gaussian, \(421 n\).
Sigwart, 88
and inverse probability, 178
and induction, 273
Simmons and asymmetry in Pernoulli's Theorem, 359
Simpson and Least Syuares, 210
Small Numbers, Law of, 401 f .
Soeiety for P'sychical Research, 298 n .
Space. 255
and uniformity, 226
irrelevance of, 301
Spedding and Ellis and Bacon, 265 n ., \(266 n\).
Spielräume, doctrine of, 88
Spinoza, 116 n., \(282 n\).
Spirits, probability of, 300
Star drifte, 296
Stars, binary, 294
Statistical frequency, theory uf, 93 f .
generalisation of, 101
criticism of, \(10: 3\)
stability of, \(336,392-415\)
fluctuation of, 392
Statistical inference, 327 f .
induction, 406 f .
Statisties, and predietion. 306
descriptive and inductive, \(3: 27\)
Stumpf, 41 r. . if \(n\)., 172 .
Sub-analogies, \(223,2: 29\)
Sub-groups of propositions, \(126,12!9\)
Succession, Law of 82
See Rule of
Süssmikch and regular frequencies, 333

Taylor, Jeremy, 308 n.
Tchehycheff, Theorem of, 353, 355
and Poisson's Theorem, 357
Telepathy, probability of, 30 )
Terrot, Bishop, 43 n .
and Whately, 179 n .
and combination of premisses, 17!
Testimony, theory of, 180

Time, 255
and uniformity, 226
irrelevance of, 301
Todhunter, 29.4n., 318 n., 370 n.
and Bayes, 175
and Craig, 184
and Petersburg Paradox, 316
and Bernoulli's Theorem, 340 n .
Truth and probability, 116 n., 322
Truth frequency, 101, 406
Tschuprow, 358, \(399 n\).
and statistical frequency, \(348,394 n\). method of, 384

Uniformity of Nature, Law of, 226, 248, 255, 263, 276
and Mill, 270
Universal Causation, Law of, 248
Universal Induction and statistical methods, 389, 406-417
Universe of reference, 117, 129, 130
Unknown probabilities, 372, 373, 375

Variables in Probalility, \(2 \mathrm{~s}, 12:, 412 \ldots\). Variety, 234
and induction, 219
limitation of, 258, 260, 427
Venn, 84, 106 n., \(294 n\).

Venn (contd.) -
and experience, 85
and Bernoulli, 86, 341
and frequency theory, 93 f.
and inverse probability, 100
and Least Squares, 206 n .
and induction, 273
and chance, 288
and ' random,' 290
and Pule of Succession, 372, 378,382
Weight, of evidence, 312
and ethies, 315
Weighting of averages, 211
Weldon and dice, 362
Whately and combination of premisses, 178
Whitehead, and frequency theory, 101 and invalid inference, \(329 n\).
Whittaker, I. T., and Rule of Succession, \(376 n\).
Wilbraham, H., and Boole, 167 n .
Wolf and dice, 362
Yule, \(349 n\)., \(361 n\).
and approximation, 161
and independence, 166
and 'statistics,' 327
and coin-tossing, \(346 n\)., 361 n .
and correlation, 421, 424

O False and treacherous Probability, Enemy of truth, and friend to wickednesse; With whose bleare eyes Opinion learnes to see, Truth's fecble party here, and barrennesse.

THE ENI)

\section*{THE ECONOMIC}

\section*{CONSEQUENCES OF THE PEACE}
\(1 \% / / \%\) - - Mr. Keyness work on the Peare Cinlothere is one of a calibre quite different from any of those others which we have hitherto received. Mr. Keynes writes with knowledge; he was himself one of the chief actors in the Conterence, and his book is an important political event. . . Mr. Keynes brings great literary abiiity, a broad view, a clear grasp of general principles, to bear upon the very complicated matters with which he is occupied, and in his hands these questions of coal, exchange, and reparation can be read with pleasure by the non-technical student."
 markable book betrays a grasp of the subject which could only have been derived from personal experience at the (ounference itself."
 and arguments, to which every one will resort for years to come who wishes to strike a blow against the forces of prejudice, delusion and stup,idity. . . . Never was the case for reasonableness more power fully put. It is enforced with extraordinary art. What might easily have been a difficult treatise, semi-official or academic, proves to be as fascinating as a good novel: it has all the merits-the accuracy, the method, the well-consitered arrangement of the best kind of State Paper, with none of the shortcomings."

\author{
Sion. 7s. 6it. net
}

\section*{INDIAN \\ CURRENCY AND FINANCE}

ECONOMIC JOURNAL.—"The book is, and is likely long to remain, the standard work on its subject. . . . While academic students will be grateful for this acute and informing work, it will be read with as much interest, and perhaps even greater appreciation, by men of business and affairs."

SPECTATOR.-" Mr. Keynes's careful and disinterested study of the monetary facts of twenty years, and his methodical marshalling of facts and figures, will be useful even to those--and they will probably be few -who are not convinced by his reasoning."

CLARE MARKET REVIEW.-"By his really masterly treatment of the Indian currency system, the author has made a very valuable addition to our economic literature. . . . Mr. Keynes has succeeded admirably in both of the chief tasks which the writing of his book involved-those, namely, of explaining, and of upholding, the system."

ECONO MIST. -" A searching, well-informed, and admirably lucid survey."

BANLERS' MAGAZINE.-"Written in an attractive manner, without undue repetition or employment of charts or tables, the work is of value to all students of currency matters."
LONDON: MACMHLLAN \& CO., LTd.

```


[^0]:    ${ }_{1}$ This will be written $a / h=a$.
    ${ }^{2}$ See also Chapter II. §5.

[^1]:    1 This classification of "primary" and "secondary" propositions was sumeratiol th the hy Mr. W. F. Johment.

[^2]:    ${ }_{1}$ This view has often been taken, e.g., by Bernoulli and, incidentally, by Laplace; also by Fries (see Czuber, Entwichlung, p. 12). The view, occasionally held, that probability is concerned with degrees of truth, arises out of a confusion between certainty and truth. Perhaps the Aristotelian doctrine that future events are neither true nor false arose in this way.

[^3]:    1 Necessity and Impossibility, in the senses in which these terms are used in the theory of Modality, seem to correspond to the relations of Certainty and Impossibility in the theory of probability, the other modals, which comprise the intermediate degrees of possibility, corresponding to the intermediate dearees of probability. Almost up to the end of the seventeenth century the traditional treatment of modals is, in fact, a primitive attempt to bring the relations of probability within the scope of formal logic.

[^4]:    ${ }^{1}$ I do not mean to imply, however, at any rate at present, that the ultimate premata of an atamon mon alsays be promary properitions.

[^5]:    1 "Whenever the terms greater and less can be applied, there twice, thrice, ete., can be conceived, though not perhaps measured by us." --" Theory of Probabilities," Encyclopaedia Metropolitana, p. 395. He is a little more guarded in
     probability is concerned.

[^6]:    ${ }^{1}$ Leibniz notes the subtle distinctions made by Jurisconsults between degrees of probability ; and in the preface to a work, projected but unfinished. which was to have been entitled $A d$ stateram juris de gradibus probationum et probabilitutum he recommends them as models of logne in contingent questions (Couturat, Logique de Leibniz, p. 240).
    ${ }^{2}$ I have considerably compressed the original report (Sapwell v. Bass).

[^7]:    1 (haplin $\because$ Hirhs ( 1911 ).
     the probability of subsequent marriage into the peerage.

[^8]:     so subtly as this; for the average value of a prize (I have omitted the details bearingen their value) could not have been fairly eatimated wo high as E foot.

[^9]:    ${ }^{1}$ Not altogether; for it would be natural to select the set to which the relation of certainty belongs.

[^10]:    ${ }^{2}$ Quoted by Mr. Binatnquet with reference to the Principle of Non Sinficient Reason.
    ${ }^{2}$ See also Chap. VII.

[^11]:    ${ }^{1}$ Published in 1886. A brief account of Von Kries's principal conclusions will be given on p. 87. A useful summary of his book will be found in a review by Meinong, published in the Göttingische gelehrte Anzeigen for 1890 (pp. 56-75).
    ${ }^{2}$ (f. (e.g.) the well-known passage in Jevons's I'rinciples of S'cience, vol. i. p. 243, in which he assigns the probability $\frac{1}{2}$ to the proposition "A Platythliptic Coefficient is positive." Jevons points out, by way of proof, that no other

[^12]:    probability could reasonably be given. This, of course, involves the assumption that every proposition must have some numerical probability. Such a contention was first criticised, so far as I am aware, by Bishop Terrot in the E'din. Phil. Trans. for 1856. It was deliberately rejected by Boole in his last pub lished work on probability: "It is a plain consequence," he says (Edin. I'hil. Trans. vol. xxi. p. 624), " of the logical theory of probabilities, that the state of expectation which accompanies entire ignorance of an event is properly represented, not by the fraction $\frac{1}{2}$, but by the indefinite form g." Jevons s particular example, however, is also open to the objection that we do not even know the meaning of the subject of the proposition. Would he maintain that there is any sense in saying that for those who know no Arabic the probability of every statement expressed in Arabic is even? How far has he been influenced in the choice of his example by known characteristics of the predicate 'positive'? Would he have assigned the probability $\frac{1}{2}$ to the proposition 'A Platythliptic Coefficient is a perfect cube'? What about the proposition A Platythliptic Coefficient is allogeneous ' ?

[^13]:    ${ }^{1}$ This example is taken from Von Kries, op. cit. p. 24. Von Kries does not seem to me to explain correctly how the contradiction arises.
    *A. Nitache (" Die Dimenwionen der Wahrscheinlichleeit und die Eviden/ der ('ngewissheit," V'ierte!juhrsschr. S. wissensch. D'hilus. vul. avi. 1. 29, 1892), in

[^14]:    criticising Von Kries, argues that the alternatives to which the principle must be applied are the smallest physically distinguishable intervals, and that the probability of the specific volume's lying within a certain range of values turns on the number of such distinguishable intervals in the range. This procedure might conceivably provide the correct method of computation, but it does not therefore restore the credit of the Principle of Indifference. For it is argued, not that the results of applying the principle are always wrong, but that it does not lead unambiguously to the correct procedure. If we do not know the number of distinguishable intervals we have no reason for supposing that the specific volume lies between 1 and 2 rather than 2 and 3 , and the principle can therefore be applied as it has been applied above. And even if we do know the number and reckon intervals as equal which contain an equal number of 'physically distinguishable' parts, is it certain that this does not simply provide us with a new system of measurement, which has the same conventional basis as the methods of specific volume and specific density, and is no more the one correct measure than these are ?

[^15]:    1 The best accounts of this subject are to be found in Czuber, Geometrische
     vol. i. Pp. 75-109; Crofton, Encycl. Brit. (9th edit.), article 'Probability'; Borel, Eléments de la théorie des probabilités, chaps. vi.-viii.; a few other references are given in the following pages, and a number of discussions of individual problems will be found in the mathematical volumes of the Kducational Times. The interest of the subject is primarily mathematical, and no discussion of its principal problems will be attempted here.
    ${ }^{2}$ As Czuber points out (Wahrscheinlichkeitsrechnung, vol. i. p. 84), all problems, whether geometrical or arithmetical, which deal with a continuum and with non-thumerable aggregates, are commonly diacoussed under the name of 'Lesmetrical probability:' See also Lammel, U'ntorsuchungen.

[^16]:    ${ }^{1}$ P'oinearé, C'alcul des probubilitis, Pp . 126 et seq.
    2Burtranl, Caleal des probubilites, p. 4: " L’infini n’est pas un nombre ; on ne doit pas, sans explication, l'introduire dans les raisonnements. La procison illuanore d.s mots pourait faire naitme des contradietions. Choisir and hasard, entre un nombre infini do cas possibles, n'est pas une indication sullisante. ${ }^{\text {. }}$

[^17]:    1 The difficulty in question was first pointed out by Boole, Laws of Thought, pp. 369-370. After discussing the Law of Succession, Boole proceeds to show that "there are other hypotheses, as strictly involving the principle of the 'equal distribution of knowledge or ignorance' which would also conduct to conflicting results." See also Von Kries, op. cit. pp. 31-34, 59, and Stumpf, Über den Begriff der mathematischen Wahrscheinlichkeit, Bavarian Academy, 1892, pp. 64-68.
    ${ }^{2}$ If A and B are two balls, A white, B black, and A black, B white, are different 'constitutions.' But if we consider different numerical ratios, these two cases are indistinguishable, and count as one only.
    ${ }^{3}$ C. S. Pcirce in his Theory of Probable Inference (Johns Hopkins Studies in Logic), pp. 172, 173, argues that the 'constitution' hypothesis is alone valid, on the ground that, of the two hypotheses, only this one is consistent with itself. I agree with his conclusion, and shall give at the close of the chapter the fundamental considerations which lead to the rejection of the 'ratio' hypothesis. Stumpf points out that the probability of drawing a white ball is, in any case, $\frac{1}{2}$. This is true; but the probability of a second white clearly depends upon which of the two hypotheses has been preferred. Nitsche (loc. cil. p. 31) seems to miss the point of the difficulty in the same way.

[^18]:    ${ }^{1}$ This is Poisson's sulution, Recherches, P. 96.

[^19]:    ${ }^{1}$ As it is the aim of trigonometry to determine the position of an object, which is in a sense visible, not by a direct observation of it, but by observing some other object tosether with certain relations, so an indiecet method of this kind is the aim of all logical system. If the truth of some propositions, and the validity of sume arebuments, combld not be reconenised directly, we could make mo progress. We may have, moreover, some power of direct recognition where it is not necessary in our logical system that we should make use of it. In these cases the method of logical proof increases the certainty of knowledge, which we might be able to possess in a more doubtful manner without it. In other cases, that, for instance, of a complicated mathematical theorem, it enables us to know propo-itions to be true, which are altweether beyond the reach of our direct insight ; just as we can often obtain knowledge about the position of a partially visible or even invisible ohject hy starting with otmervations of other objects.

[^20]:    ${ }_{1}$ That is to say, $h_{1}$ is irrelevant to $x / h$ if $x / h_{1} h=x / h$.
    ${ }^{2}$ That is to say, $h_{1}$ is irrelevant to $x / h$, if there is no proposition $h_{1}^{\prime}$ such that $h_{1}^{\prime} h_{1} h \quad 1, h_{1} h_{1}=1$, anl $r h_{1} h_{1}: s_{i} h$.
    ${ }^{3}$ Where no misunderstanding can arise, the qualification 'as a whole' will i, somnti:nes omittal.

    4 I.e (in symbolism) $h_{1}$ and $h_{2}$ are independent and complementary parts of $h$ if $h_{1} h_{2}, h_{1} h_{1} h_{2} ; 1$, and $h_{2} / h_{1} \neq 1$. Alsog $h_{1}$ is relevatut if $x_{1} h_{1}: x_{1} h_{2}$.

[^21]:    ${ }^{1}$ If $\phi(a), \phi(b)$, etc., are propositions, and $x$ is a variable, capable of taking the values $a, b$, etc., then $\phi(x)$ is a propositional function.

[^22]:    i The more complicated cases in which the propositional function, of which the alternatives are instances, involves more than one variable (see s 16), can be dealt with in a similar manner mutatis mutandis.

[^23]:    ${ }^{1}$ See Chap. IV. § 14 for the meaning of these terms.

[^24]:    1 This phrase is used by Von Kries, op. cil. p. 179, in a somewhat similar eonnection.
     n., (a) ith; andl, , (l, (b) $1, h$.

[^25]:    ' $h_{1}$ i- $n e x$ evidence so long as $h_{1} / h 1$.

[^26]:    ${ }^{1}$ See also Chapter XXVI. § 7.
    ${ }^{2}$ ('f. Locke, Eissay conctining II uman Üuderstunding Fook ii. chap). xxi. है (67: "He that judges without informing himself to the utmost that he is capable, cannot acquit himself of judging amiss."

[^27]:    ${ }^{1}$ There are also some remarks by Czuber (Wahrscheinlichkeitsrechnung, vol. i. p. 202) on the Erkenntnisswert of probabilities obtained by different methods, which may have been intended to have some bearing on it.

[^28]:    1 " It is not my design to inquire furt her into the nature, the foundation and measure of probahility ; or whence it proceeds that likeness should beget that presumption, opinion and fall ronviction, which the human mind is formed to receive from it, and which it does necessarily produce in every one; or to guard against the errors to which reasoning from analogy is liable. 'This belongs to the subject of logic, and is a part of that subject which has not yet been thoroughly considered."

[^29]:    ${ }^{1}$ Eng. Trans., p. 353.
    ${ }^{2}$ An Essay concerning Human Understanding, book iv. "Of Knowledge and Opinion."
    ${ }^{3}$ Introduction to the Analogy.

[^30]:    

[^31]:    ${ }^{1}$ French philosophy of the latter half of the eighteenth century was profoundly affected by the supposed conquests of the Calculus of Probability in all fields of thought. Nothing seemed beyond its powers of prediction, and it almost succeeded in men's minds to the place previously occupied by Revelation. It was under these influences that Condorcet evolved his doctrine of the perfectibility of the human race. The continuity and oneness of modern European thought may be illustrated, if such things amuse the reader, by the reffection that Condorcet derived from Bernoulli, that (iodwin was inspired by Condorcet, that Malthus was stimulated by Godwin's folly into stating his famous doctrine, and that from the reading of Malthus on Population Darwin received his earliest impulse.
    
    3 Up. 'at. p. 42. .

[^32]:    ${ }_{1}$ Poincaré's opinions on Probability are to be found in his Calcul des Probabilités and in his Science et Hypothese. Neither of these books appears to me to be in all respects a considered work, but his view is sufficiently novel to be worth a reference. Briefly, he shows that the current mathematical definition is circular, and argues from this that the choice of the particular probabilities, which we are to regard as initially equal before the application of our mathematics, is entirely a matter of 'convention.' Much epigram is, therefore, expended in pointing out that the study of probability is no more than a polite exercise, and he concludes: "Le calcul des probabilités offre une contradiction dans les termes mêmes qui servent à le désigner, et, si je ne craignais de rappeler ici un mot trop souvent répété, je dirais qu'il nous enseigne surtout une chose ; c'est de savoir que nous ne savons rien." On the other hand, the greater part of his book is devoted to working out instances of practical application, and he speaks of ' metaphysics' legitimising particular conventions. How this comes about is not explained. He seems to endeavour to save his reputation as a philosopher by the surrender of probability as a valid conception, without at the same time forfeiting his claim as a mathematician to work out probable formulae of practical importance.

[^33]:    ' On the Foumblations of the Theory of I'ralutialitu...
    
    
    1 1.1. 1.1. . 1. 1.

[^34]:    
    

[^35]:    ${ }^{1}$ Sigwart, Logic (Eng. edition), vol. ii. p. 220.

[^36]:     Of his four fundamental rules of probability, for instance. three are, as he states them, certainly falso.

[^37]:    ${ }_{1}$ These essays were published in the Transactions of the Camb. Phil. Son., the first in 1843 (vol. viii.), and the second in 1854 (vol. ix.). Both were repmutid in Mathematical and other Writings (1863), together with three other howf papers on Probability and the Method of Least Squares. All five are full of spirit and originality, and are not now so well known as they deserve to be.
    ${ }^{2}$ The first edition appeared in 1866 . Revised editions were issued it 1876 and 1888 . References are given to the third colition of 1888 .

[^38]:    ${ }^{1}$ Edgeworth uses the term 'probability' widely, as I do; but he makes a distinction corresponding to Venn's by limiting the subject-matter of the Calculus of Probabilities. He writes ('Philosophy of Chance,' Mind, 1884, p. 223): "The Calculus of Probabilities is concerned with the estimation of degrees of probability; not every species of estimate, but that which is founded

[^39]:    ${ }^{1}$ Let the reader, who is acquainted with this chapter, consider what precise assumption Venn's reasoning requires on p. 187 in the example which seeks to show the efficacy of Lord Lister's antiseptic treatment à posteriori. What is the 'inevitable assumption about the bags' when it is translated into the language of this example?

[^40]:    ${ }^{1}$ In what follow: 1 an muchindebsed for some sugeretions in favour of the frequency theory communicated to me by Dr. Whitehead; but it is not tw be. supposed that the exposition which follows represents his own opinion.
    -This is 1)r. Whitehead's phrase.

[^41]:    i The question, previunsly at issue, as to hou the class of reference is deter-
    

[^42]:    1 In the course of the present discussion the disjunctive $a+b$ is never interpreted so as to exclude the conjunctive $a b$.
    ${ }^{2}$ For a discussion of this term see Chapter XVI. § 2.
    ${ }^{3}$ Venn argues (Logic of Chance, pp. 173, 174) that there is an inductive ground for making this inferenee. The question of extending the fundamental theorems of a frequency theory of probability by means of induction is diseussed in $\S 14$ below.

    4 lide Chapter XII. § 6, and Chapter XIV. § 4.
    5 Vide Chapter XIV.§ 5.

[^43]:    
     reducing unnecessary organs to a rudimentary condition.

[^44]:    ${ }^{1}$ Spinoza had in mind, I think, the distinction between Truth and Probability in his treatment of Necessity, Contingence, and Possibility. Res enim omnes ex data Dei natura necessario sequutae sunt, et ex necessitute naturue Jei delerminatae sunt ad certo modo existendum el operandum (lithices i. 33). That is to say, everything is, without qualification, true or false. At res

[^45]:    
     term it, Prohability, solely arises out of the limitations of our knowledge Contingence in this wide sense, which includes every proposition which, in relation to our knowledge, is only probable (this term covering all intormediate degrees of probability), may be further divided into Contingence in the strict senwe, which corresponds to an it priori or formal probability exceeding zero, and P'ossibility ; that is to say, into formal possibility and empirical possibility.
    
    
    
    
    

[^46]:    A Or more strictly, "perception of which, together with knowledge of the first set, justifies an appropriate degree of rational belief about the second."

[^47]:    ${ }^{1}$ ' $a$ can be inferred from $b$, ' $a$ followe from $b$,' ' $a$ is certain in relation to $b,{ }^{\prime} \quad a$ is logically involved in $b$,' I regard as equivalent expressions, the precise meaning of which will be defined in succeeding paragraphs. ' $a$ is implied by $b$,' I use in a different sense, namely, in Mr. Russell's sense, as the equivalent of ' $b$ or not-a.'

    2 For the conception of a group, and for many other notions and definitions in the course of this chapter-those, for example, of a real group and of logical priority-I am largely indebted to Mr. W. E. Johnson. The origination of the theory of groups is due to him.

[^48]:    
    

[^49]:    

[^50]:    

