

International Encyclopedia of Unified Science
Volumes I and II: Foundations of the Unity of Science
Volume I · Number 6

Principles of the Theory of Probability

Ernst Nagel

University of Chicago Press

International Encyclopedia of Unified Science
Foundations of the Unity of Science

International Encyclopedia of Unified Science

Editor-in-chief Otto Neurath

Associate Editors Rudolf Carnap Charles W. Morris

**Foundations
of the Unity of Science**

(Volumes I-II of the Encyclopedia)

The University of Chicago Press. Chicago, Illinois

International Encyclopedia of Unified Science

Volume I · Number 6

Principles of the Theory of Probability

Ernest Nagel

The University of Chicago Press. Chicago, Illinois

Committee of Organization

RUDOLF CARNAP
PHILIPP FRANK
JOERGEN JOERGENSEN

CHARLES W. MORRIS
OTTO NEURATH
LOUIS ROUGIER

Advisory Committee

NIELS BOHR
EGON BRUNSWIK
J. CLAY
JOHN DEWEY
FEDERIGO ENRIQUES
HERBERT FEIGL
CLARK L. HULL
WALDEMAR KAEMPFERT
VICTOR F. LENZEN
JAN LUKASIEWICZ
WILLIAM M. MALISOFF

R. VON MISES
G. MANNOURY
ERNEST NAGEL
ARNE NESS
HANS REICHENBACH
ABEL REY
BERTRAND RUSSELL
L. SUSAN STEBBING
ALFRED TARSKI
EDWARD C. TOLMAN
JOSEPH H. WOODGER

International Encyclopedia of Unified Science
Volume I · Number 6
Principles of the Theory of Probability

Ernest Nagel

Contents:

	PAGE
I. THE MATERIALS FOR THE STUDY OF PROBABILITY	
1. Introduction	1
2. Development and Applications of the Theory of Probability	4
II. THE CALCULUS OF PROBABILITY AND ITS INTERPRETATIONS	
3. Preliminary Distinctions	17
4. Fundamental Ideas of the Frequency Interpretation of Probability	19
5. Fundamental Theorems in the Calculus of Probability	26
6. Nonfrequency Interpretations of Probability Statements	44
III. UNSETTLED PROBLEMS OF GENERAL METHODOLOGY	
7. Logical Problems of the Frequency Interpretation of the Probability Calculus	51
8. Probability and Degree of Confirmation or Weight of Evidence	60
9. Concluding Remarks	75
SELECTED BIBLIOGRAPHY	79

Vol. I, No. 6

vii

Principles of the Theory of Probability

Ernest Nagel

I. The Materials for the Study of Probability

1. Introduction

The daily affairs of men are carried on within a framework of steady habits and confident beliefs, on the one hand, and of unpredictable strokes of fortune and precarious judgments, on the other. Our lives are not filled with constant surprises, and not all our beliefs are betrayed by the course of events; nevertheless, when we examine the grounds even of our most considered actions and beliefs, we do not usually find conclusive evidence for their correctness. We undertake commercial or scientific projects, although we do not know whether illness or death will prevent us from completing them; we plan tomorrow's holiday, although we are uncertain what weather tomorrow will bring; we estimate our budget for next year, although we are not sure whether the consequences of floods, droughts, or wars will not seriously throw it out of balance. In spite of such uncertainties, we manage to order our lives with some measure of satisfaction; and we learn, though not always easily, that, even when the grounds for our beliefs are not conclusive, some beliefs can be better grounded than others. Our claims to knowledge may not be established beyond every possibility of error, but our general experience is warrant for the fact that even inconclusive arguments may differ in their adequacy.

These observations are commonplaces. But they immediately lose their triviality if, by setting them in the context of a penetrating comment of Charles Peirce, we extend them to the procedures and conclusions of the various special sciences. The American logician once remarked that in the exact sciences of measurement, such as astronomy, no self-respecting scientist

Principles of the Theory of Probability

will now state his conclusions without their coefficient of probable error. He added that, if this practice is not followed in other disciplines, it is because the probable errors in them are too great to be calculated. The ability of a science to indicate the probable errors of its measurements was thus taken by Peirce as a sign of maturity and not of defect. By his remark Peirce therefore wished to indicate that for the propositions in the most developed empirical sciences, no less than for those in the affairs of everyday life, no finality is obtainable, however well they may be supported by the actual evidence at hand.

The temper of mind which is illustrated by such an appraisal is itself the product of modern science and of a preoccupation with its procedures. It is based on the conviction that the methods of the natural sciences are the most reliable instruments men have thus far devised for ascertaining matters of fact, but that withal the conclusions reached by them are only probable because they rest upon evidence which is formally incomplete. The import of such an insistence upon the fallible character of science can be best appreciated by contrasting it with the classic conception of science, formulated in Greek antiquity and perpetuated in a powerful intellectual tradition. This conception of scientific knowledge was modeled upon the ideal of a completely demonstrative and absolutely indubitable natural science, such as Euclidean geometry was believed to be. It was assumed that the subject matter of genuine science was a realm of precise, unalterable laws, and that scientific knowledge, as distinct from belief, opinion, or mere experience, was to be equated with demonstrated knowledge. For such knowledge facts are not contingent, since they must be apprehended through their "reasons" or "causes," and the propositions which express them must therefore be "necessary." Furthermore, it was maintained that the "basic propositions" required as premisses for demonstrated knowledge could be grasped by the intellect directly and infallibly and could be seen to be true with even greater assurance than any of the conclusions derived from them. The scientific enterprise was accordingly construed as the progressive apprehension of an eternal

order of necessary connections, so that complete certainty was the earmark of genuine knowledge. The changing and the variable could not be subject matter for science; they could at best be the concern of belief and opinion. Variability in the materials studied or in the outcome of measurements was taken to indicate either the obdurateness of subject matter to rational connections or the failure of thought to reach its proper objectives. In a word, experience in the sense of observation and experiment, since it could not yield necessary propositions, could not be the ground for scientific knowledge.

This ideal of science dominated the minds of the great pioneers of modern science and of many of their most illustrious successors; and it is this conception which forms the tacit premiss of many philosophic commentators upon modern science, such as Descartes, Locke, Leibniz, and Kant. It is scarcely possible to exaggerate the significant role which this ideal has played in intellectual history. In proclaiming the ideal of science to be *systematic* knowledge, the rationalist tradition has stimulated research and has led to the development of science as something other than an indigestible miscellany of dubious facts. On the other hand, the great services of classic rationalism cannot hide the fact that its theory of self-evidence rests upon an inadequate analysis of the methods of science, so that it has frequently blocked the progress of inquiry and, though pledged to the ideal of clarity, has not seldom successfully courted obscurantism. Rationalism made complete certitude the theoretical condition for genuine science, but its belief that the latter was obtainable could be maintained only by neglecting or misinterpreting the approximate and contingent character of statements dealing with matters of fact. The long history of science and philosophy is in large measure the history of the progressive emancipation of men's minds from the theory of self-evident truths and from the postulate of complete certainty as the mark of scientific knowledge. Some of the major turning-points in that history consist in radically diminishing the class of statements certifiable simply by a rational insight into their truth. And some of its most dramatic moments have

Principles of the Theory of Probability

occurred when the approximate and incompletely grounded character of allegedly indubitable propositions was recognized.

The forthright admission of the probable or contingent character of even our most soundly based beliefs and the emphasis upon the general reliability of the *methods* of scientific inquiry rather than upon its conclusions are characteristic of contemporary empiricism. For the traditional empiricism of Locke and Mill, which in intent was a revolt against the exaggerated claims of rationalism, accepted in all essentials the standards and preconceptions of the views it nominally opposed. But that admission of the probable character of our beliefs is not the outcome of a capricious decision: it is not a pronouncement made for the sake of wilfully opposing a historically powerful tradition, nor is it a thesis advanced for the sake of a special set of values and an ulterior conception of nature. That admission and that emphasis have been wrung from students as a consequence of their reflection upon the history of science and of a painstaking examination of its methods. Contemporary empiricists who maintain that our knowledge of matters of fact is "probable" do not thereby maintain that such knowledge is inferior to knowledge of some other kind obtainable by methods different from those the natural sciences employ. On the contrary, they maintain that "probable knowledge" is the only kind of knowledge we can find or exhibit, and that the methods and techniques of the sciences are efficacious and dependable precisely because they make available knowledge of that character.

2. Development and Applications of the Theory of Probability

Although the term 'probable' has been employed several times in the preceding section, no precise sense has been attached to it. It is one of the objects of this essay to assign a clear meaning to sentences which contain the term and its derivatives; but it must be admitted at the outset that an analysis of what is meant by 'probable,' which would meet the unanimous approval of competent students of the subject cannot be given at the present stage of research. In the present and preceding

Development and Applications of the Theory of Probability

sections the statement 'Knowledge of matters of fact is probable' is to be understood in the rather loose sense that conclusions of factual inquiry are not in principle incorrigible, because the formal conditions for assuring the logical validity of those conclusions are not completely realized, and because statements having factual content are not logically necessary.

The doctrine that knowledge of matters of fact is only probable is one of the central theses of contemporary analysis of scientific method. The implementation of this doctrine with modern logical and mathematical techniques is relatively recent. But even during the heyday of classic rationalism the status of beliefs which fell short of its ideal of scientific knowledge was frequently and vigorously discussed. Out of the permanent needs which generated such discussions have grown the modern calculi of probability and the diverse interpretations and applications which the term 'probable' has received. The possible equivocality and unquestionable vagueness of the term are therefore in part due to the history of empirical science. The brief survey, to which we now turn, of some of the contexts in which the term is and has been applied aims to achieve three things: to emphasize the intimate connection between the development of empirical science and the growing need for a theory of probability; to indicate the great range of applications of the term 'probable' and so to provide the materials for a discussion of its meaning; and to serve as a convenient introduction to the issues and techniques under contemporary discussion.

a) Aristotle's logical writings formulate the rationalist ideal of science, but his biological works exhibit less exacting standards of scientific adequacy. His evaluation of the then extant theories of sexual reproduction is characteristically judicious; evidence, some of it observational, is presented in opposition to the Hippocratean doctrines and in support of his own views, but there is not even a pretense that the question is settled beyond further debate. His examination of the facts of heredity show him to be familiar with at least the crude elements of a statistical explanation of the similarities and differences between ancestors and descendants. The mechanism which he

Principles of the Theory of Probability

suggested as an explanation for the observed facts was in essentials that of a shuffling and recombination of characters, so that only certain traits would normally recur. Even before Aristotle the principle of natural selection was advanced by Empedocles, though Aristotle had little use for it. No ancient mathematician developed a technique for handling statistical aggregates, and it is possible that the prevalent view of chance as an agent was an insurmountable impediment to a consistent working-out of a statistical view of nature which the theory of natural selection suggests. Nevertheless, passages in Aristotle and in the writings of the Ionians, Democritus, and Hippocrates which could be cited indicate that such a view was not foreign to the ancient mind.

b) Occasions for dealing with evidence which is not conclusive, but which nevertheless carries some weight, presented themselves in the legal and social transactions of both Athens and Rome. For example, there was a rule in Athenian courts excluding hearsay evidence, on the ground of the general untrustworthiness of reported statements as compared with the evidence of eyewitnesses. The courts of Rome took pride in deciding cases before them upon the basis of reason and the evidence of fact rather than caprice, and complicated safeguards were instituted to assure the adequacy of the evidence presented. Curious and distorted survivals of these appear in the formalistic rules of evidence of the Middle Ages. For example, two witnesses were required for a "full proof," the testimony of a single reliable witness counted as "half-proof," while a doubtful witness counted for "less than half." The object apparently aimed at was to convert the process of rendering a decision into a calculation of the "resultant force" of the testimony submitted. There was thus some basis in fact for Rabelais' portrait of Judge Bridlegoose, who made his decisions which were "correct" in the long run by throwing appropriately loaded dice. Years after, Leibniz was again intrigued by the possibility of a calculus of evidence, and the ideal of a quantitative science of proof has frequently hovered before students of probability. When the calculus of probability was finally developed, many

Development and Applications of the Theory of Probability

of its great masters, like Laplace, Poisson, and their followers, attempted to turn it to such a use, though with singularly poor success.

c) Although no individual knows the exact date of his death, he can reasonably expect a definite span of life. His expectations are based on statistical regularities manifesting themselves in large groups of men. The use of such statistical uniformities for predicting individual behavior illustrates a common type of "uncertain inference," which in recent times has become an exacting and important discipline; and it was exemplified in ancient practices as well. Various forms of commercial insurance existed in Babylonia, Greece, and Rome, and the Romans were no strangers to life insurance. Just how and on the basis of what kind of statistical information the various rates were estimated is now unknown, although it is fairly clear that the estimates were not arbitrary. For example, the rates on bottomry and marine insurance depended on the destination of the vessel and the season during which it sailed; and, although the careful gathering of vital statistics is a modern phenomenon, a census of populations was frequently made in antiquity for military and taxation purposes. While therefore the practice of insurance was not placed upon a sound basis until the end of the eighteenth century, it was built on a large body of factual information; and, even though the beliefs which rested on this information fell short of the classic ideal of science, they made possible the planning and execution of important policies.

During the Middle Ages the Italian cities saw the beginning of commercial insurance as a profit-making enterprise; by 1700 the business of insurance was rapidly developing in western Europe, with life insurance in regular demand a century later. These enterprises required to be supported by adequate statistical techniques, and in fairly rapid succession there appeared a number of important statistical studies. For example, in 1662 John Graunt showed how to employ the register of deaths, which began to be kept in London during the Black Death, to make forecasts on population trends; during the same century John De Witt, grand pensioner of Holland, and Halley, the Eng-

Principles of the Theory of Probability

lish astronomer, concerned themselves with annuity problems. Halley laid the basis for a correct theory of the subject, and he showed how to calculate from the mortality tables which he constructed the value of an annuity on the life of a person of given age. While scientific knowledge in accordance with the rationalist ideal was not obtainable for these domains, probable knowledge was, and it became the guide to life.

d) The entire subject of statistical inference now called for a theoretical foundation. The need was supplied from an unexpected quarter—the theory of games of chance. Dice games played with ankle bones were popular in antiquity, and the ancients distinguished between the “likelihoods” of certain combinations of throws. They did not, however, develop any technique for assigning numerical measures to the different “degrees of likelihood.” The quantitative study of games of chance begins with the modern period and was cultivated by a brilliant succession of mathematicians.

Solutions of special problems in the division of stakes and the placing of wagers were first given by Cardan and Galileo (sixteenth century); but the general attack on the theory which was involved in their analyses began with Pascal and Fermat (seventeenth century), who showed that all the special problems under consideration could be reduced to problems in the mathematical theory of permutations and combinations. Upon this basis a convenient calculus was developed, which was subsequently applied to many different fields of inquiry. Huygens, the Bernoullis, Montmort, De Moivre, and Bayes are the most prominent figures in the early history of the subject. Their work was systematized and completed in the great treatise of Laplace (early nineteenth century), and the point of view from which they conducted their analyses remained until quite recently the basis for the interpretation and extension of the mathematical theory. The principle upon which Laplace assigned numerical values to probabilities was that of analyzing the possible outcome of a situation into a set of alternatives which could be judged as “equally possible.” Accordingly, although we might be ignorant of which one of these alternatives

Development and Applications of the Theory of Probability

would occur, a method was provided by the aid of which an appropriate "degree of rational belief" could be assigned to propositions about "chance events." In brief, fortuitous events which had heretofore been denied the status of genuine objects of scientific knowledge could now be handled in an expert manner with the help of probability theory. The intellectual instrument was thus forged for developing what is now known as the statistical view of nature, and for exhibiting important continuities in techniques and methods in different scientific disciplines.

e) The theoretical foundations of the probability calculus as formulated by Laplace still had their roots in traditional rationalism. On the one hand, probability judgments were understood to betoken ignorance: Laplace maintained that all events are regulated by "the great laws of nature" which a sufficiently powerful intelligence could use to foretell the future in the most minute way. On the other hand, judgments of equipossibility were made to rest on a nonexperimental basis. A critique and reformulation of these foundations were not to come for several decades. Nevertheless, these rationalistic preconceptions were conveniently overlooked in the application of probability theory. One of the earliest and most successful of these applications was to the systematization of measurements and observations in the experimental sciences. Astronomy was the first to employ the theory of probability for this purpose. Justly regarded for a long time as the most exact science of measurement, it nevertheless was patent to everyone that the measurements actually performed did not yield identical numerical values for what was presumably the same magnitude, however carefully gross disturbing factors were eliminated. In consequence, the measurable predictions calculated from astronomical theory were not in precise agreement with the numbers obtained by direct measurement. Given the climate of opinion within which astronomical theory was developed, it was congenial to interpret these fluctuations as deviations or "errors" from the "true values" of magnitudes, and to attribute the "inexactitude" of actual measurements to human failing.

Principles of the Theory of Probability

Nevertheless, there was a pressing need for techniques to estimate the "true values" from the actual measurements and to measure the degree to which the latter "approximate" to the former.

This situation is not local to astronomy. As Boyle once explained, "You will meet with several observations and experiments which, though communicated for true by candid authors or undistrusted eyewitnesses, disappoint your expectations, either not at all succeeding constantly or at least varying much from what you expected." Indeed, to test any theory, empirically specified initial conditions must be given, and the consequences logically derived from them with the help of the theory must be compared with the outcome of further observational procedures. Thus, two series of actual measurements or observations must be instituted to test a theory; and, for both series, we find that as a matter of fact there are groups of discordant statements reporting the issue of our measurements. Whether the theory is in accordance with the "facts" cannot therefore be decided without some further hypothesis on the actual measurements we make.

The study of this problem in terms of the theory of probability constitutes what is known as the theory of errors. It was begun in the eighteenth century by Boscovitch, Lambert, Euler, and Thomas Simpson, and was continued by Daniel Bernoulli, Legendre, Gauss, and Laplace. Gauss showed that if we assume that the deviations from the "true magnitude" are produced by a large number of hypothetical "elementary errors" acting independently of one another, the form of the law of distribution of the actual measurements can be deduced, and an approximation to the "true value" can be calculated from the data. The Gaussian "Law of Error" and the Method of Least Squares for systematizing discordant observations have played an important role in subsequent researches in the theory of measurement and statistics. Recent critical work on the foundations of probability shows that Gauss's arguments for the law rest on assumptions which cannot always be made legitimately. In consequence, alternative laws for the distribu-

Development and Applications of the Theory of Probability

tion of errors have been proposed, notably by Poisson, Pearson, Gram, and Charlier, each suitable for different circumstances.

f) The expanding national economies following the breakup of the feudal system required the gathering of extensive factual information in order to guide the formulation of financial, military, and political policies. The earliest attempts to tie up the mathematical theory of probability with the analysis of such descriptive statistics were made by De Moivre, Nicholas and Daniel Bernoulli, Euler, and D'Alembert. Under the influence of the ideas of the French Encyclopedists, who sought a rational basis for monetary undertakings, public-health administration, judicial procedure, and even the conduct of elections, Condorcet tried to apply on a comprehensive scale the new mathematical instrument of probability to all such matters. Like Laplace and Poisson after him, he achieved only a modicum of success. It was characteristic of this group of writers to misunderstand and consequently to overrate the function of the probability calculus; their procedure frequently seemed to rest on the assumption, as one commentator remarked, that valuable results can be obtained from unreliable and insufficiently analyzed data by employing a sufficient number of signs of integration. However, it is to the great credit of these men to have insisted on the fusion of statistical methods with the theory of probability. Interest in this fusion was further stimulated by the Belgian astronomer Quetelet, who saw in the theory of probability the appropriate tool for developing a reliable social science. Poisson had enunciated in a somewhat confused form a "law" which he called "the law of great numbers"; according to this law large aggregates of elements exhibit definite properties with a stable relative frequency, even though these properties occur quite fortuitously within the aggregates. Quetelet popularized this idea in the context of the social disciplines. He regarded the "average man," as computed from the extensive statistics he gathered, as the analogue in social matters of the center of gravity in mechanics; and he saw in the statistical regularities with which certain human actions occur the operation of comprehensive laws of social development. He thus

Principles of the Theory of Probability

found it easy to believe that determinate laws could be formulated to connect the different social averages—determinate laws modeled upon those recorded in Laplace's *Celestial Mechanics*. However, Quetelet was uncritical both in gathering his statistical material and in interpreting it; he was never really clear as to the meaning of statistical averages, and never appreciated the limitations of the probability calculus. His influence, great at first, rapidly waned, and for a time so did the interest in applying theoretical statistics to the social sciences.

When interest in the subject was once more revived, it was supported by research needs in biology, psychology, and theoretical physics. Statistical methods subsequently developed on the basis of probability theory were then applied to matters as remote and different as the calculation of the density of telephone traffic and the maintenance of manufactured products at a certain standard of quality. The determination of the character of an indefinitely large population on the basis of samples drawn from it is a problem common to many disciplines and many daily occupations. The elements of an adequate theory of sampling within the framework of a theory of probability were first laid down by Lexis, and further developed by Bortkiewicz, Tschuprow, Markoff, and others. They showed that the sheer number of instances in a sample is no guaranty of its representative character, criticized statistical practice which relied upon the accumulation of unanalyzed numerical data, and developed a technique for obtaining trustworthy statistical coefficients from data grouped carefully according to the variety, homogeneity, and number of the instances. More recently, R. A. Fisher and his school have approached the problem from a different point of view, and, in addition to devising important criteria for the adequacy of statistical coefficients, he has called needed attention to the serious limitations of many of the Laplacian formulas. Other distinct contributions to the theory of sampling have been made by Fechner, Bruns, Galton, Thiele, Pearson, and Neyman. In consequence of these researches, the theory of errors, the theory of sampling, the theory of curve-fitting, now all fall within a comprehensive theory of probability.

Development and Applications of the Theory of Probability

g) Although the importance of the main ideas of the mathematical theory of probability for systematizing measurements was quickly recognized in the sciences, the theory of probability was for a long time usually regarded as simply ancillary to the theoretical disciplines. Thus, it was commonly assumed in physics that its laws are statable in "deterministic" form, such that the positions and velocities of elementary particles at one time are connected in precise ways with the positions and velocities at any other time. It is today a commonplace, however, that some of the most fruitful applications of the theory of probability occur *within* the theoretical framework of various sciences. The ancient idea that the apparently permanent objects around us as well as the regularities in their behavior could be viewed as aggregate effects of a large number of hypothetical elements undergoing random changes has frequently attracted the creative minds in science. Thus, Kepler played with it to explain the appearance of a new star in 1604; Boyle had a corpuscular theory for the states of aggregation of bodies; Huygens even formulated a corpuscular theory of gravitation; and Daniel Bernoulli's interpretation of Boyle's law for gases in terms of the kinetic theory of matter is well known.

Apparently, the first man to work out such theories with sufficient quantitative detail to make possible an empirical evaluation of the magnitudes associated with the hypothetical elements was Joule. He computed the average velocities of hydrogen molecules on the basis of statistical considerations and showed that, in order to produce the observed effects, the velocities must lie in specified intervals. The statistical explanation of thermal phenomena was carried to much greater lengths by Maxwell: he showed that, if certain assumptions are made concerning the probabilities with which the particles of a gas acquired different positions and velocities, the familiar gas laws could be deduced. But perhaps the greatest triumph of probability theory within the framework of nineteenth-century physics was Boltzmann's interpretation of the irreversibility of thermal processes; this he was able to do in terms of the most probable distribution of the energies of the molecules of a gas.

Principles of the Theory of Probability

In consequence, the second law of thermodynamics can be formulated as a theorem in probability, and irreversible processes turn out to be statistical phenomena.

Thermodynamics is not an isolated instance of the use of statistical concepts within theoretical formulations. Even before Maxwell employed probability theory for the study of gases, G. G. Stokes used it to analyze the effects of polarized light coming from different sources as the average or most probable effects. Again, statistical mechanics, which consistently employs the theorems and the point of view of probability theory, has been fruitfully applied in the study of the history and distribution of the stars. And more recently the entire theory of radiation has been developed to include systematically within itself hitherto unrelated phenomena, on the basis of a profound and radical application of the theory of probability.

But physics is not the only science with has profited from using statistical concepts in its theories. Democritus tried to explain the resemblances and dissimilarities between parents and children in terms of a shuffling of the atoms coming from the ancestors of a child; Aristotle employed related notions in discussing similar problems. And ever since Darwin called attention to the importance of the facts of variation for any adequate biological theory, students of biology have been developing a statistical treatment of the subject. It is obviously essential to distinguish between variations due to heredity and those due to environment; this phase of the subject has been explored by Pearson and his school with the help of the mathematical theory of probability. But attempts such as those of Galton to formulate the laws of heredity in terms of average contributions from the ancestors of a given set of progeny are now known to be unsatisfactory; and Galton's mistakes indicate some of the limitations of statistical methods in general. The theoretical basis for modern experimental genetics was supplied by Mendel. The theory of the mechanism of heredity he proposed, which involved the transmission, segregation, and combinations of unit characters in various proportions, obviously lends itself to be exploited in terms of the fundamental ideas of mathematical

Development and Applications of the Theory of Probability

probability; with its help, the artificial selection of plants and animals has been brought to a high stage of perfection. The mathematical theory of natural selection for those groups in which a Mendelian analysis can be made has been worked out mainly by R. A. Fisher, J. B. S. Haldane, and S. Wright.

In general, therefore, the introduction of probability notions into the theoretical structure of physics and biology has been most fruitful. It has made possible the prediction of the relative frequency with which definite characters occur in groups of individuals, even when it is not feasible to predict the occurrence of such characters for a given individual.

h) The developments which have thus far been surveyed have gradually tended to undermine the authority of classic rationalism in science. For, as points of view borrowed from the theory of probability and statistics assume central roles, both within the theoretical framework of the sciences as well as in the procedures of applying theories to matters of fact, it becomes progressively more difficult to assume that the principles of a science are self-evident or necessary. This change in the climate of opinion has been further supported by a general logical criticism of the assumptions of the classic view which began in antiquity. The Epicureans, as well as Sceptics like Carneades, developed conceptions of the logic of inquiry which made allowances for the formally incomplete character of the evidence for empirical statements. In modern times it was Hume's discussion of causality which put the rationalist notion of necessity on the defensive, and since then every variety of empiricism has had its day in court.

Nevertheless, although the Humean analysis was a powerful dissolvent of ancient preconceptions, it was not so powerful as some of the internal technical developments within the special sciences. A few of these have already been indicated. But perhaps the most significant single technical achievement, from the point of view of its general effect upon the philosophy of science, has been the definitive refutation of the thesis that Euclidean geometry is the apodeictic science of space. For the discovery of the non-Euclidean geometries exhibited logically

Principles of the Theory of Probability

possible alternatives to conceptions previously regarded as indubitable, while the recognition of a distinction between pure and applied mathematics cut the ground from under the claims of traditional rationalism. Whatever doubt still lingered as to the possibility of alternative conceptions of "space" was finally removed when the Newtonian physics and its Euclidean framework for mechanics were displaced by relativity physics and its framework of Riemannian geometry. And perhaps the final coup de grâce to the claim that physical principles are indubitable and necessary was supplied when the familiar physics of continuous action was found to be inadequate for vast ranges of phenomena, and made way for the contemporary physics of quanta.

There are thus both historical and analytic grounds for the view central to empiricism that there is no a priori knowledge of matters of fact, and there are similar grounds for the thesis of contemporary empiricism that no amount of empirical evidence can establish propositions about matters of fact beyond every possibility of doubt or error. On the other hand, the recognition of this state of affairs raises an important problem. Although our beliefs cannot be established with absolute finality, we do, as we must, differentiate between them on the ground of the character of the evidence which supports them. We regard it more probable that Napoleon was a historical character than that he is a solar myth. We believe that the prognoses of a modern physician are more reliable than those made a century ago. A chemist accepts Lavoisier's theory of combustion as better founded than Stahl's phlogiston theory, and a physicist will urge that the quantum theory of radiation is today more securely based than it was twenty years ago. There is clearly an obvious need for canons to evaluate the evidence supporting any proposition, and for the formulation of the principles we employ in deciding that one statement is better grounded than another. Judging by the success of past attempts to supply them, it may be suspected that every proposed list of such canons and formulations will be incomplete and will require emendation with the progress of inquiry. The need, however,

is a permanent one; and the attempts to satisfy it constitute the broader setting and the larger theme in contemporary discussions of probability.

II. The Calculus of Probability and Its Interpretations

3. Preliminary Distinctions

The vast range of material which has just been outlined has been traditionally regarded as constituting the subject matter for a theory of probability. It has been frequently assumed that a precise meaning can be found for the term 'probable' which is common to its use in each of the contexts indicated. Upon this point competent students are not in agreement. Without prejudging the issues involved, it is possible to distinguish between two groups of statements in which the term 'probable' or its derivatives occur. The first group contains statements such as the following:

'The probability that a man of thirty will survive his thirty-first birthday is .955'; 'The probability that a normal coin will present a head after being tossed is $\frac{1}{2}$ '; 'The probability that on the basis of the evidence in 1938 the electronic charge e has a value in the interval $(4.770 \pm .005) \times 10^{-10}$ electrostatic units is .67'; 'The probability that a molecule of hydrogen has a velocity in the interval $v - dv$ and $v + dv$ is p '; 'The probability of a 10° deflection of an α -ray passing through a film is $\frac{1}{4}$ '; 'The intensity of a spectral line is determined by the probability of the corresponding quantum transition'; and 'A snowstorm in New York during January is more probable than during November.'

The second group contains such statements as:

'Relative to our present evidence the theory of light quanta has a probability which is greater than its probability relative to the evidence available in 1920'; 'The evidence makes it highly improbable that Aristotle composed all the works attributed to him'; 'The theory of evolution has a higher probability on the evidence than the theory of special creation'; 'It is probable that, had Cleopatra's nose been a half-inch longer, the course

Principles of the Theory of Probability

of the Roman Empire would have been different'; and 'It is not probable that Christ was a descendant of King David.'

Statements in the first group employ the term 'probable' in a sense which, by practically unanimous consent, is subject to the rules of the calculus of probability; indeed, the calculus has been explicitly devised to handle such "probabilities." Statements in the second group apparently employ the term to indicate the "degree" of the adequacy of the evidence supporting the proposition; students are not agreed whether the mathematical calculus of probability is applicable to such "probabilities." In the present section we shall for the most part confine ourselves to such statements which clearly fall into the first set; the discussion of the second group of statements is reserved for Section III. But nothing said in the present section will exclude the possibility that both classes of statements, in spite of apparent differences between them, are subject to the same interpretation.

Even though we have restricted the scope of the present section in the indicated way, it is still not possible to specify a sense of 'probable' or even a formulation of the calculus of probability, upon which reasonably complete agreement is obtainable. There are, in fact, three major interpretations of the term. According to the first, a degree of probability measures our subjective expectation or strength of belief, and the calculus of probability is a branch of combinatorial analysis; this is the classical view of the subject, which was held by Laplace and is still professed by many mathematicians. It is not always clear whether by 'expectation' proponents of this view understand *actual* expectations or *reasonable* expectations. According to the second, probability is a unique logical relation between propositions, analogous to the relation of deducibility; its most prominent contemporary supporter is the economist Keynes. According to the third, a degree of probability is the measure of the relative frequency with which a property occurs in a specified class of elements; this view already appears in Aristotle, was proposed by Bolzano and Cournot during the last century and further developed by Ellis, Venn, and Peirce, and was finally

Fundamental Ideas of the Frequency Interpretation of Probability

made the basis for a subtle mathematical treatment of the subject by von Mises and other contemporary writers. We shall begin with the exposition of the frequency interpretation of probability and its calculus; subsequently, the other two views will be briefly considered, and it will be argued that the frequency view is the one most suitable for the first of the foregoing two classes of statements; in Section III we shall finally examine some important methodological problems which cluster around the frequency view.¹

4. Fundamental Ideas of the Frequency Interpretation of Probability

The basic ideas of the frequency conception of probability emerge upon an examination of such a statement as 'The probability that a person of thirty residing in the United States survives his thirty-first birthday is .945.' The meaning of such a statement can be ascertained by examining how it is established. That procedure, greatly simplified, is somewhat as follows. Suppose that during a period of years there is no migration to or from the United States, and that during these years exact counts are made of its inhabitants who fall into definite age groups. Thus, suppose that in 1900 there are 2,000,000 persons who have just reached their thirtieth birthday, and that exactly one year later there are 1,890,000 persons who have just reached their thirty-first birthday; that is, of the thirty-year-olds in 1900 a ratio of .9450 survive at least another year. We imagine that similar figures are obtained for the four succeeding years, and that the ratios of thirty-year-olds who survive their thirty-first birthday are .9452, .9456, .9451, and .9454, respectively. We notice that, although these ratios are not constant, the differences do not appear until the fourth decimal is reached. We may say, therefore, that during these five years approximately 945 out of a thousand thirty-year-old residents of the United States live for at least another year; and we may make the further *assumption* that for an *indefinite number of future years* the corresponding ratios of survivals remain in the neighborhood of .945. Accordingly, the statement 'The probability that a thirty-year-old resident of the United States sur-

Principles of the Theory of Probability

vives his thirty-first birthday is .945' means that in the long run the relative frequency with which thirty-year-olds in the United States survive for at least one year is approximately .945.

The following points must be noted in this example. In the first place, the probability statement supplies no information about any *individual* resident of the United States; the information is relevant to the individual Tom Brown only in so far as he belongs to the *class* of thirty-year-old residents. Second, the statement supplies information about no property of this class of residents other than the one explicitly specified, namely, the property of surviving at least one year. Third, the statement supplies a numerical value—the value of a *relative frequency*. Fourth, the statement does not mean that in *every* thousand thirty-year-olds 945 will live for at least another year. And, finally, this numerical value is intended to specify the relative frequency of survivals during an indefinite number of years, or “in the long run,” and not only during the years for which an actual count has been made; that is to say, the statement makes a prognosis.

We now turn to the general definition of probability statements. But at once difficulties arise. A proposed definition must be precise and unambiguous and at the same time should be modeled as closely as possible upon the procedures which the foregoing example illustrates. On the other hand, those procedures have been described with the help of terms which are not precise; in particular, the expressions ‘in the long run’ and ‘approximately’ are highly vague, and it is not easy to develop a mathematical theory in terms of them. Accordingly, the definition to be proposed will replace these expressions by more precise ones, which are appropriate for developing a *calculus* of probability. Hence, although the definition will be modeled upon the illustration, it will employ precise mathematical concepts to which there cannot easily be assigned a simple empirical meaning. The methodological problems which are a consequence of this procedure will have to be considered subsequently.

Fundamental Ideas of the Frequency Interpretation of Probability

Let R be a non-empty class of elements (i.e., it contains at least one member), to be known as the *reference class*; for reasons which will be soon apparent, the elements of R will be supposed to be serially ordered. Let A be some property which the elements of R may exhibit. Suppose R contains n elements, and let $\text{nu}(A \text{ and } R)$ be the number of elements in R which have the property A . We may now define the expression 'the relative frequency with which elements in R have the property A ,' which we abbreviate into ' $\text{fr}_n(A, R)$,' as follows:

$$\text{'fr}_n(A, R)\text{' is short for } \frac{\text{'nu}(A \text{ and } R)\text{'}}{n} .$$

It is evident that a relative frequency is a proper fraction. Suppose now that the number of elements in R increases. In general, the fraction $\text{fr}_n(A, R)$ will be different for different values of n . It may happen, however, that these fractions will crowd around some fixed number p , and will differ from it by a small positive magnitude ϵ which diminishes as n increases: in familiar language, $\text{fr}_n(A, R)$ will tend toward p in the long run. The mathematically precise way of rendering this possibility is to say that $\text{fr}_n(A, R)$ approaches p as a limit with increasing n ; that is,

$$p = \lim_{n \rightarrow \infty} \text{fr}_n(A, R) .$$

What mathematicians understand by 'limit' is illustrated by the following. Consider the infinite series of fractions $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots$; its limit is 1. Suppose we have the infinite series of numbers $x_1, x_2, x_3, \dots, x_n, \dots$, where the subscripts indicate the ordinal position of the numbers in the series. To say that p is the limit of this series means that, however we may select a positive number ϵ , there is a number N such that for every n , if $n > N$, then the absolute difference between x_n and p (i.e., neglecting signs) is less than ϵ . The reason for requiring R to be serially ordered is now clear. If R contains only a finite number of elements, $\text{fr}_n(A, R)$ is unaffected by the order in which the elements are counted; but the limit of $\text{fr}_n(A, R)$ when R is not finite does depend on the order in which the elements of R (and therefore the relative frequencies) are arranged.

It is very convenient in developing the calculus of probability to define 'probability' as 'the limit of relative frequency.' If we abbreviate statements of the form 'The probability that an

Principles of the Theory of Probability

element has the property A if it is a member of R is p ' into 'prob(A, R) = p ,' the definition takes the following form:

$$\text{'prob}(A, R) = p \text{ is short for } 'p = \lim_{n \rightarrow \infty} \text{fr}_n(A, R)'$$

'prob(A, R) = p ' may be read conveniently as 'the probability of A in R is p .'

Expressions like 'prob(A, R)' which describe numbers are called *numerical expressions* and consist of the *functor* 'prob' together with its arguments; 'nu(A and R)' is also a numerical expression, and 'nu' another functor. The expression 'prob(A, R)' is the fundamental numerical expression in the mathematical theory of probability developed on a frequency basis; it describes a real number, which may be irrational, in the interval 0 to 1 inclusive. Within the calculus of probability the statement 'The probability that a thirty-year-old resident of the United States survives his thirty-first birthday is .945' must now be taken as equivalent to 'The limit of the relative frequency with which the property of surviving at least one year occurs in the ordered class of thirty-year-old residents of the United States is .945.'

It has already been pointed out that the foregoing definition of 'probability' has been proposed for the sake of its great convenience in calculations. It employs the notions of infinite ordered classes and of limiting values of relative frequencies in such classes. It is obvious, of course, that in empirical procedures we are occupied with finite classes which may or may not be ordered, and with relative frequencies rather than limits of relative frequencies. Some writers (e.g., Copeland and Popper) have proposed to use as the definition of 'probability' not 'the limit of relative frequencies,' but 'the *condensation point* of relative frequencies.' p is said to be a condensation point of the series $x_1, x_2, x_3, \dots, x_n, \dots$, if for every positive number ϵ and every N there is an n such that $n > N$ and the absolute value of the difference between x_n and p is less than ϵ . Such a definition has the merit that a proof can be given that there is at least one condensation point for relative frequencies in an infinite reference class, even though no limit exists; it suffers from the disadvantage that according to it a property may have more than one probability in a given class, so that the calculus of probability becomes more complicated.²

It is essential to note the following points in connection with probability statements interpreted in terms of relative frequencies:

Fundamental Ideas of the Frequency Interpretation of Probability

a) No meaning can be attached to any expression which, taken literally, assigns a probability to a single individual as having a specified property. Statements of probability predicate something of an individual (e.g., Tom Brown) only in so far as he is an element in a specified reference class. Probability statements which do not do so *explicitly* must be regarded as incomplete if they are to be significant: they must be understood as making an implicit specification of the reference class within which the designated property occurs with a certain relative frequency.

b) Every probability statement of the form thus far considered is a factual statement, into whose determination empirical investigations of some sort must always enter. Probability statements are on par with statements which specify the density of a substance; they are not formulations of the degree of our ignorance or uncertainty. To assert that the probability of a normal coin presenting head after being tossed is $\frac{1}{2}$, is to ascribe a physical property to a coin which is manifested under determinate conditions.

c) Since probability statements require the specification of a reference class with respect to which a given property has some degree of probability, a given property can be associated with different degrees of probability, according to the reference class which is specified. The probability of surviving at least one year may be .945 with respect to the reference class of thirty-year-old residents of the United States; it may be .734 with respect to the reference class of sixty-year-old men; and it may be .345 with respect to the class of domesticated cats.

d) Since the explicit definition of probability statements is in terms of relative frequencies, the *direct* evidence for them is of a *statistical* nature. Thus, waiving difficulties to be mentioned, the direct evidence for the probability of a coin falling head is obtained by counting the frequency with which it falls head. However, probability statements do not always occur singly and are often part of a more or less inclusive *system* of statements or a *theory*. In such cases the estimation of the numerical values of the probabilities and the subsequent testing

Principles of the Theory of Probability

of such values may be made on the basis of *indirect* evidence which in some cases may even be nonstatistical. This point will receive further attention in Section III.

e) Since a probability has been defined as the limit of a relative frequency (or, even more loosely, as the relative frequency in the long run), every probability statement is a hypothesis; such a hypothesis cannot be completely confirmed or finally verified by the (necessarily) finite amount of evidence actually at hand at any given time. It is thus quite possible that the numerical value estimated for a probability on given evidence is not correct, so that revisions of the estimate may have to be made repeatedly. It is partly for this reason that in the history of the subject discussions of probability have run parallel with discussions of the problem of induction. The situation with respect to probability statements is indeed more serious than has been just indicated. For not only cannot probability statements be completely confirmed; they cannot even be completely disconfirmed by any actual evidence. The issues involved will receive further attention below.

f) Finally, it is a mistake to suppose that the successful use of probability statements depends in any way upon the issues of what is popularly known as "determinism." Because current microscopic physics employs theories involving in an essential way probability considerations, many thinkers, including reputable scientists, have been persuaded into supposing that the general breakdown of "mechanistic" explanations has been demonstrated, that processes in nature are "noncausal," and that contemporary physics supplies evidence for the existence of human "freedom" and for a "spiritualistic" world-view. Such suppositions feed upon mistaken or misleading formulations of the actual issues in modern physics, as has been pointed out repeatedly, among others, by Venn, Peirce, Philipp Frank, and Henry Margenau. It is perhaps sufficient to note that the use of probability statements requires no commitment, even by implication, to any wholesale "deterministic" or "indeterministic" world-view; they can be used successfully in such contexts in

Fundamental Ideas of the Frequency Interpretation of Probability

which specified properties occur with stable relative frequencies in specified classes of elements.

One of the main difficulties in most debates on causality is that the term is not explained with sufficient precision to make discussion fruitful. (As a matter of fact, specific contributions to the sciences of nature rarely if ever contain the term.) Without entering into detailed analyses of the issues sometimes raised, the following observations may help clarify some of them.

(i) Questions of causality can be significantly discussed only if they are directed to the *theories* or *formulations* of a science and not to its subject matter. No clear sense can be given to most pronouncements that the world or any segment of it is a causal process. On the other hand, in discussing the causal or noncausal character of a given theory, two factors must be examined: the *state* (or system of properties) in terms of which the physical system under discussion is described and the *laws* (or system of equations) which connect the states at different times and places. The state of a system is sometimes specified with the help of properties belonging to what are taken as "individual elements," sometimes with the help of the properties of a field, and sometimes in statistical terms involving the properties of aggregates of individuals. The laws also can differ markedly in form: they may establish a unique correspondence between states at different times or they may have the form of probability statements; they may be explicit functions of the time variable or they may not, etc. No universally accepted criterion has been formulated for judging whether a theory is "causal." Classical mechanics is frequently considered as the example par excellence of such a theory; the states considered by it are the positions and momenta of material particles, and its laws are certain differential equations of the second order not containing the time variable explicitly. It is often assumed that, in order to be a causal theory, the states employed by the theory must be those of classical mechanics. In that case, however, neither classical electromagnetics nor modern quantum mechanics are causal theories, although the former is usually so regarded. In some cases, on the other hand, the distinction between noncausal and causal theories is made on the basis of whether the states are specified in statistical terms or not, so that classical statistical mechanics and modern quantum mechanics would both be classified as noncausal theories. The main point to be borne in mind is that both factors, specification of state and form of law, are relevant to the discussion. Even theories which employ statistically specified states have been said to be causal because their laws establish a unique correspondence between its states at different times—although with respect to certain properties of individuals in the system the theories have been classified as noncausal, because the equations supply only probability statements concerning the occurrence of properties of individuals.

Principles of the Theory of Probability

(ii) Because probability statements supply no information about any individual member of the reference class, it has been imagined that a physical theory involving probability considerations precludes a "causal" explanation of the phenomena under consideration. Now such a theory will usually specify the state in statistical terms; and, as a consequence, the predictions of the theory may have the form of probability statements concerning the properties of individuals. In some cases, however, it is also possible to describe the situation in terms of nonstatistical states, so that laws of a "causal" type may connect these new states. Whether it is possible or convenient to do so is obviously a matter to be decided for each case by experiment and scientific policy. It so happens that for the phenomena studied by classical statistical mechanics it is possible to do this; and, as a consequence, the "indeterminism" of classical statistical mechanics has been usually regarded as eliminable or inessential. Such an elimination is not possible for modern quantum mechanics within the framework of its procedures, and marks an important difference between classical and recent physics. In any case, nothing more than a very technical scientific difference is involved; and at least some physicists are of the opinion that future research may remove this difference. It should also be noted, moreover, that if the Ψ -function in modern quantum mechanics is taken to specify the state of the system, without seeking to interpret this function statistically, quantum theory may also be regarded as a "causal" theory, for its laws have the form of equations usually regarded as of the causal type: they establish a unique correspondence between states at different spatio-temporal regions.³

5. Fundamental Theorems in the Calculus of Probability

1. *The function of the calculus.*—It should now be clear that probability statements cannot in general be certified on purely formal grounds, so that pure mathematics and logic are not in the position to assert probability statements of the form considered thus far. What then, it may be asked, are the function and nature of the mathematical calculus of probability? To readers of the preceding monographs in this *Encyclopedia* the answer will be familiar. The calculus of probability has the same general function as a demonstrative geometry or a demonstrative arithmetic: given certain initial probabilities, the calculus of probability makes it possible to calculate the probabilities of certain properties which are related to the initial ones in various ways. Thus, arithmetic cannot tell us how many people live in either China or Japan; but, if the population of China and the population of Japan are given, we can compute the

Fundamental Theorems in the Calculus of Probability

combined population of these countries. The calculus of probability functions in the same way. It is important to recognize that the propositions asserted *in the calculus* are not factual or empirical statements: they are all certifiable on formal grounds alone, and are analytic of the definitions and rules initially laid down. The proposal to establish the theorems of the calculus of probability by experimentation, which has sometimes been made, is as ill-considered as would be the proposal to prove experimentally that $3^2 + 4^2 = 5^2$. The function of the probability calculus, like that of other calculi, is to make possible the *transformation* of probability statements in order that their theoretical content be made evident. The calculus thus has an *instrumental function* in the context of empirical investigations. It permits us to derive the relative frequencies with which certain properties occur from initial probability statements which do not explicitly mention those frequencies; in this way the calculus makes possible a more adequate testing of the probability statements which we entertain by making explicit the predictions they involve.

The detailed discussion of the calculus of probability can be undertaken only with the help of the technical apparatus of mathematical analysis. Some familiarity with at least the elementary theorems of the calculus is, however, essential for a just appraisal of its function and limitations. In the present section we shall accordingly state a few standard theorems of the calculus and, incidentally, obtain important material for evaluating the claims of standpoints in the philosophy of science which do not subscribe to an empirical outlook.

2. *Elementary theorems of the calculus.*—Suppose we wished to obtain the probability that children of white parents are both blue-eyed and blond. The reference class R consists of children born to white parents; the problem requires for its answer the (limit of the) relative frequency with which the properties A (being blue-eyed) and B (being blond) jointly occur in R . This number, $\text{prob}(A \text{ and } B, R)$, could be estimated directly. It may, however, be *calculated* from the following two numbers: the probability of A in R ; and the probability of B in the refer-

Principles of the Theory of Probability

ence class consisting of blue-eyed children of white parents; i.e., from $\text{prob}(A, R)$ and $\text{prob}(B, A \text{ and } R)$. The following theorem, known as the General Product Theorem, can be easily demonstrated: The probability of A and B in R is equal to the probability of A in R , multiplied by the probability of B in A and R . Using familiar mathematical symbolism this can be stated as follows:

$$\text{prob}(A \text{ and } B, R) = \text{prob}(A, R) \times \text{prob}(B, A \text{ and } R). \quad (1.1)$$

We happen to know that the relative frequency of blond hair in the class of blue-eyed children of white parents is not equal to the relative frequency of blond hair among children of white parents in general. In some cases, however, the probability of B in R does equal the probability of B in the narrower reference class A and R . The properties A and B are then said to be "independent" of each other with respect to R . In such cases we obtain the Special Product Theorem:

$$\text{prob}(A \text{ and } B, R) = \text{prob}(A, R) \times \text{prob}(B, R). \quad (1.2)$$

Theorem 1.1 may itself be generalized for the joint occurrence of n properties A_1, A_2, \dots, A_n .

Suppose now that we required the probability that children of white parents are either blond or black-haired. The properties A (being blond) and B (being black-haired) cannot as a matter of fact jointly occur in the class R ; they are said to be "exclusive" with respect to R . For such exclusive properties the following Special Addition Theorem can be easily proved: The probability of A or B in R is equal to the probability of A in R plus the probability of B in R . Again employing mathematical symbolism we obtain:

$$\text{prob}(A \text{ or } B, R) = \text{prob}(A, R) + \text{prob}(B, R). \quad (2.1)$$

Let us next obtain the probability that children of white parents are either male or female. Properties such as male and female are called "contradictory properties" in the class of human births because they are both exclusive and exhaustive. It is obvious that the probability of being male or female in

Fundamental Theorems in the Calculus of Probability

the class of children of white parents must be equal to 1. In particular, we can demonstrate the following theorem:

$$\text{prob}(A \text{ or not-}A, R) = 1, \quad (2.2)$$

and with the help of theorem 2.1 we also obtain

$$\text{prob}(A, R) + \text{prob}(\text{not-}A, R) = 1. \quad (2.3)$$

Thus if the probability of a male birth among humans is .51, the probability of a female birth in that class must be .49.

Theorems 1.1 and 2.1 are fundamental in the elementary calculus of probability. From them a large number of important consequences can be derived by applying the ordinary rules of logic and arithmetic. A few of them will be mentioned because of their practical and methodological importance.

There is clearly no difference between the probability of A and B in R and the probability of B and A in R . Accordingly,

$$\begin{aligned} \text{prob}(A \text{ and } B, R) &= \text{prob}(A, R) \times \text{prob}(B, A \text{ and } R) \\ &= \text{prob}(B, R) \times \text{prob}(A, B \text{ and } R), \end{aligned}$$

from which we obtain the Division Theorem:

$$\text{prob}(B, A \text{ and } R) = \frac{\text{prob}(B, R) \times \text{prob}(A, B \text{ and } R)}{\text{prob}(A, R)}, \quad (3.1)$$

which can be given the following more convenient form:

$$\text{prob}(B, A \text{ and } R) = \frac{\text{prob}(B, R) \times \text{prob}(A, B \text{ and } R)}{\text{prob}(B, R) \times \text{prob}(A, B \text{ and } R) + \text{prob}(\text{not-}B, R) \times \text{prob}(A, \text{not-}B \text{ and } R)}. \quad (3.2)$$

Theorem 3.2 is one form of what is known as Bayes's theorem. A more general form is the following: Let B_1, B_2, \dots, B_n be a set of mutually exclusive and exhaustive properties with respect to R , and let B_i be any one of them. Then

$$\text{prob}(B_i, A \text{ and } R) = \frac{\text{prob}(B_i, R) \times \text{prob}(A, B_i \text{ and } R)}{\sum_1^n \text{prob}(B_i, R) \times \text{prob}(A, B_i \text{ and } R)}, \quad (3.3)$$

where, as usual, ' Σ ' is the sign of summation.

Principles of the Theory of Probability

Bayes's theorem and the consequences which have been drawn from it have played important roles in discussions of the foundations of probability, induction, and scientific method. It is therefore important to illustrate how it may be employed, especially since the limitations of its use have not always been clearly understood or remembered. Let R be the very numerous class of shots fired at a certain target; let A be the property of a shot hitting the bull's eye; and, finally, let B_1 be the property of a shot that it is fired from Rifle 1, B_2 from Rifle 2, and B_3 from Rifle 3. All the shots are supposed to be fired from these rifles. The (limiting) relative frequency of shots from Rifle 1 is $\frac{3}{8}$, from Rifle 2 is $\frac{1}{8}$, and from Rifle 3 is $\frac{4}{8}$; furthermore, the probability that a shot fired from Rifle 1 hits the bull's eye is $\frac{1}{8}$, while from Rifle 2 it is $\frac{2}{8}$, and from Rifle 3 it is $\frac{1}{8}$. What is the probability that a shot which hits the bull's eye is fired from Rifle 2? The question asks for the value of $\text{prob}(B_2, A \text{ and } R)$; it is obtainable from theorem 3.3 if we remember that $n = 8$, $\text{prob}(B_1, R) = \frac{3}{8}$, $\text{prob}(B_2, R) = \frac{1}{8}$, $\text{prob}(B_3, R) = \frac{4}{8}$, $\text{prob}(A, B_1 \text{ and } R) = \frac{1}{8}$, $\text{prob}(A, B_2 \text{ and } R) = \frac{2}{8}$, and $\text{prob}(A, B_3 \text{ and } R) = \frac{1}{8}$. A simple calculation shows that the required probability is $\frac{2}{3}$.

Bayes's theorem is frequently referred to as a theorem in "inverse probabilities," and it has been traditionally regarded as the instrument for discovering the probability of "causes" or "hypotheses" from known "effects" or "consequences." The reason for this terminology is perhaps evident from the illustration: the probability which is sought is that of the "cause" (namely, of a shot being fired from Rifle 2), on the assumption that certain "effects" have set in (namely, of the shot hitting the bull's eye). But although Bayes's theorem can be demonstrated in the calculus of probability, it can be employed to determine the probability of "causes" only if *all* the probability coefficients in the right-hand side of the formula are given. Of special importance are the probabilities of the form ' $\text{prob}(B_i, R)$ ' which are sometimes designated as the "antecedent probabilities of the causes." Now it has been often assumed that, if we possess no information to the contrary, these

Fundamental Theorems in the Calculus of Probability

antecedent probabilities are equal to one another. This assumption has been supported by what is known as the Principle of Indifference. With the help of this principle it has been supposed that probabilities could be determined a priori—that is, without recourse to empirical, and more particularly to statistical, investigations. Consequently, this assumption proceeds from a different conception of probability than the one developed in §4; and for a relative frequency conception of probability the equating of probabilities to one another simply on the ground that we know no reason why they should be unequal is a major error. Proceeding within this different conception of probability, Laplace deduced from Bayes's theorem the so-called Rule of Succession, which for a long time was accepted by eminent thinkers as the basis for reliable scientific predictions. According to this rule, if n events of a certain kind have been observed in succession, then the probability of its recurrence is $(n + 1)/(n + 2)$. Following Laplace, Quetelet declared that, "after having seen the sea rise periodically ten successive times at an interval of about twelve hours and a half, the probability that it will rise again for the eleventh time would be $\frac{11}{12}$." But it also follows from the rule that, if the tide has not been observed to rise at all, the probability of its rising is $\frac{1}{2}$; and such a consequence is a *reductio ad absurdum* of the rule and of its premisses for any view of probability which defines it in terms of relative frequencies.

In most problems it is not practically or theoretically possible to assign values to the antecedent probabilities in Bayes's theorem which could have any empirical significance. For this reason Bayes's theorem has only a limited use, and few writers today take it seriously as a means for determining the probability of a given hypothesis on the basis of given evidence.

3. *Theorems depending on irregularity in the reference class.*—The theorems which have been mentioned thus far can be demonstrated on the sole assumption that the relative frequency of a property in its reference class has a limit. But many theorems in the calculus which are of greatest importance in practice require that other conditions are satisfied as well.

Principles of the Theory of Probability

Suppose that in the class R (e.g., tosses with a coin, where the tosses may be imagined as temporally ordered) the property H (head falling uppermost) occurs as follows, where T (tail falling uppermost) and H are exclusive and exhaustive properties with respect to R :

$$H T H T H T H T H T H T \dots \quad (i)$$

That is to say, we suppose that every other toss yields a head, so that the probability of a coin falling head is $\frac{1}{2}$. Here the property H occurs with an obvious regularity; and, if such were indeed the case for actual throws with a coin, we would very likely not employ probability considerations with respect to it. In fact, however, in actual cases heads and tails occur in no such regular order, but with an irregularity somewhat as follows:

$$H H T T T H H T H T T T H H T T H H H H T H T H T T T \dots \quad (ii)$$

In this finite segment of a hypothetically infinite series the relative frequency of H is .51; but we may imagine that the limiting value of this ratio is also $\frac{1}{2}$. The second series is like the first in having $\frac{1}{2}$ as the limit for the relative frequency of H ; it is unlike the first in that H occurs in it irregularly or at random. Various theorems in the calculus of probability depend upon the assumption that the reference classes involved possess such a random character.

It is, however, not easy to give a precise sense to what we mean by 'at random,' and an extensive technical literature now exists which deals with the problems of defining 'irregularity' in a manner suitable for mathematical purposes. The first one to have called attention to the importance of conditions of irregularity and to have worked out systematically a mathematical theory of probability with them in mind is von Mises. His procedure takes its point of departure from the following observation: If in the first of the foregoing series we select the (nonfinite) subseries R' by including in it only the odd terms of R , the probability of H in R' is no longer $\frac{1}{2}$ but is 1. On the other hand, if we select the subseries R' from a random series R (such as the second series above is supposed to be) in the same way as before, the probability of H in R' is still $\frac{1}{2}$; that is, $\text{prob}(H, R) = \text{prob}(H, R')$. Now let S be any nonfinite subseries of R , subject to the sole condition that the elements of S

Fundamental Theorems in the Calculus of Probability

are not selected on the basis of their possessing or not possessing H . If for every selection of such a subseries S from R , $\text{prob}(H, R) = \text{prob}(H, S)$, the reference class R is said by von Mises to be irregular. He believes that this definition makes precise our intuitive notion of irregularity and that it formulates the conditions found in games of chance and other fortuitous events. Moreover, he maintains that in order to demonstrate many of the standard theorems in the calculus of probability his condition for irregularity must be assumed. (It is well to bear in mind, however, that considerations such as these which involve infinite classes or classes having certain types of order are pertinent primarily to the *calculus* of probability. They are introduced for the sake of constructing a consistent and powerful instrument of symbolic transformations.)

However, many students have found von Mises' definition unsatisfactory. It can be shown that, if a reference class satisfies von Mises' condition of irregularity, the order in which the specified property occurs in it cannot be formulated by any mathematical function; and doubts have therefore been raised as to the *logical* possibility of a reference class which is to satisfy so stringent a condition of irregularity. Indeed, if the phrase 'every selection of such a subseries S ' in the definition is taken seriously, a contradiction can be exhibited in the notion of an irregular reference class. Various attempts have accordingly been made by a number of writers to overcome such difficulties (e.g., by Doerge, Kamke, Tornier, Reichenbach, Popper, Copeland) by distinguishing between different types of irregularity and by proposing conditions of irregularity whose consistency can be established. None of these substitutes, however, is sufficiently strong logically for demonstrating the standard theorems in question in their full generality. But more recently it has been shown by A. Wald that by suitably relativizing the selection of subseries in von Mises' definition to certain very general classes of selections, the logical difficulties can be obviated, while at the same time the consequent restrictions upon those theorems do not seriously impair their general validity.⁴

We shall assume that a mathematically satisfactory definition of irregularity can be given, and proceed to mention a few important theorems which may be demonstrated for reference classes satisfying it. Let R be such a reference class (e.g., throws with a coin) in which the property H (head uppermost) has the probability p while the property T (tail uppermost) has the probability $1 - p$. We now suppose the elements of R to be grouped into sets of n successive elements each, and ask for the probability that exactly r elements in a set (where $r \leq n$) have the property H while the remaining $n - r$ elements have

Principles of the Theory of Probability

T. The numerical value of this probability can be shown to be equal to

$$\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \quad (4.1)$$

where $r! = 1 \times 2 \times 3 \times \dots \times (r-1) \times r$, with $0! = 1$.

It is of some importance to understand clearly what this number signifies. Suppose that *H* and *T* occur in *R* as in series (ii) on page 32 above, and suppose that *R* is broken up into sets of four successive elements each. The following sequence of sets then results:

$$\begin{aligned} (HHTT) \quad (HTTT) \quad (TTTH) \quad (TTHH) \quad (THHT) \\ (HHTH) \quad (HTHT) \quad (THTT) \dots \quad (iii) \end{aligned}$$

Some of these sets, such as the first and second, are overlapping, in the sense that they contain common terms from *R*; others, such as the first and fifth, are nonoverlapping. If we let $r = 1$, the number given by theorem 4.1 is the limit of the relative frequency with which these sets contain one *H* and three *T*'s; that is, $4p(1-p)^3$.

Suppose now that $p = \frac{1}{2}$. The probability that in sets of four successive elements from *R* there is just one element with the property *H* and three with *T*, is then $\frac{1}{4}$. But the probability that in such sets there are just two heads and two tails (here $r = 2$) is $\frac{3}{8}$. Hence, when n and p are fixed, the number determined by theorem 4.1 will vary with r . What value of r will yield a maximum value for this number? It can be shown that r must satisfy the condition

$$pn + p \geq r \geq pn + p - 1. \quad (4.2)$$

When n is very large, the value for which r yields a maximum may be taken to be pn . This means that the probability of sets with n successive elements containing just r elements with the property *H* is a *maximum*, when r is approximately equal to pn ; that is to say, the most probable value occurs for the case when the relative frequency of *H* in a set of n elements is approximately equal to the *limit* of the relative frequency of *H* in *R*.

Fundamental Theorems in the Calculus of Probability

A very important consequence, known as Bernoulli's theorem, can now be derived, which plays a central role in the practical use of the probability calculus. It can be stated as follows:

Let R be a reference class which is irregular with respect to a property H , and let $\text{prob}(H, R) = p$. Let R be broken up into sets of n successive elements each, and let ϵ be any positive number no matter how small. The probability that H will occur in these sets with *frequencies* lying in the interval $pn \pm \epsilon n$ (or with *relative frequencies* lying in the interval $p \pm \epsilon$) approaches 1 as a limit as n increases (4.3).

The following will illustrate the theorem: R is the irregular class of throws with a coin, and the probability of getting a head is taken to be $\frac{1}{2}$. We ask for the probability that in sets of n successive throws each, the frequency of heads will differ from $n/2$ by not more than $n/10$ (or that the relative frequency of heads will differ from $\frac{1}{2}$ by not more than $\frac{1}{10}$). According to Bernoulli's theorem, this probability tends to 1 as the value of n is increased. Thus, the probability that in ten successive throws there will be anywhere from four to six heads (i.e., that the relative frequency of heads will lie in the interval $\frac{1}{2} \pm \frac{1}{10}$ or $\frac{4}{10}$ to $\frac{6}{10}$) is .47; the probability that in thirty successive throws there will be anywhere from twelve to eighteen heads is .73; the probability that in fifty throws there will be anywhere from twenty to thirty heads is .84; the probability that in one hundred successive throws there will be anywhere from forty to sixty heads is .95; the probability that in five hundred throws there will be anywhere from two hundred to three hundred heads is .99, etc.

These numerical values are calculated with the help of mathematical techniques explained in treatises on probability. Of particular importance in the application of the probability calculus is the analytic formula

$$\phi(t) = \frac{e^{-t^2}}{\sqrt{\pi}} \int_0^t e^{-x^2} dx,$$

which is obtained from theorem 4.1 by a series of approximations. $\phi(t)$ is the probability that in sets of n successive elements of R which is irregular with respect to H , H occurs with a frequency lying in the interval

$$pn \pm t \sqrt{2p(1-p)n},$$

Principles of the Theory of Probability

where $\text{prob}(H, R) = p$; and tables of values of ' $\phi(t)$ ' for different values of ' t ' have been constructed. In many problems the elements of R are assumed to take on any one of an infinite set of properties. Thus, in measuring the length of objects we may suppose that the measurements are carried out with great precision, and we may accordingly find it convenient to assume that the possible values of the length are real numbers (in the strict mathematical sense). The problems arising in such cases lead to the theory of continuous or geometrical probability.

The theorem of Bernoulli has been generalized by Poisson, and more recently Cantelli and Polya have given an important extension for it. A more general theorem than that of Bernoulli has been established by Tchebycheff, which was further elaborated by Markoff.

There is also an inverse of Bernoulli's theorem, which is sometimes referred to as Bayes's theorem and is obtained with the help of theorem 3.3; it has played an important part in the theory of statistics. But this theorem in inverse probability, with whose help the probability of a statistical hypothesis is to be established on the basis of the samples that have been drawn, suffers from the serious limitations and difficulties already pointed out in connection with theorem 3.3. Critical statisticians no longer make use of it. Statisticians have now developed more suitable procedures for handling the sort of problems Bayes's theorem was intended to solve; the method of maximum likelihood, recently proposed by R. A. Fisher, is a valuable and interesting contribution to this phase of theoretical statistics.

As already indicated, many theorems of the probability calculus are demonstrable only on the assumption that the reference classes are irregular—or in easily understood intuitive terms, that there is a general independence between the occurrence of a property on one occasion and its occurrence on another. However, in many fields of research (e.g., the behavior of gases) such independence cannot, on physical grounds, be assumed to exist. Nonetheless, it has been shown that the calculus of probability may be applied even to such domains with consistency and success.⁵

It is worth while mentioning a seemingly fatal criticism of the definition of 'probability' as 'the limit of relative frequencies.' Let R be irregular with respect to H , and let $f_1, f_2, \dots, f_n, \dots$, be the series of relative frequencies of H in R after the first, second, n th terms. (Thus, in the series [ii] of p. 32, the relative frequencies are: $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \dots$.) Suppose that p is the limit of these frequencies. Then, once a number ϵ has been selected, there must be an N such that for every n greater than N the difference between f_n and p is less than ϵ ; and this means that after the N th term in R , the relative frequency of H in R will have to remain close to p . But according to theorem 4.1 there is a probability, which though small is not zero, that a very long run of successive H 's will occur; and, according to the criticism being considered, there is this probability that even after the N th term in R such a long run of

Fundamental Theorems in the Calculus of Probability

H 's will set in. However, the criticism continues, a sufficiently long run of H 's will make some of the f_n 's (with $n > N$) differ from p by more than ϵ . A contradiction is thus alleged in the calculus of probability developed on a limit basis. (Thus, suppose $p = .50$, $\epsilon = .01$, and that N is taken to be 100; and suppose that beginning with the one hundred and first throw a run of two hundred heads sets in. If x is the number of heads which have appeared in the first one hundred throws, $f_{300} = (x + 200)/300$ which differs from $.50$ by more than $.01$.)

However, the allegation of contradiction itself rests on a blunder and proceeds from a conception of probability according to which it is significant to ascribe a probability to a single occurrence. The probability specified by theorem 4.1 does *not* permit us to infer a long run of H 's starting with an *assigned* term in R , for example, with the $N + 1$ th. That probability has for its reference class R^* , the class having as *its* elements *sets* of n successive elements from R ; while p has R for its reference class. (It is possible and significant to ask for the probability that a definite run of H 's begins at some assigned term; but the answer to it is *not* given by theorem 4.1. Such a question involves the consideration of a *series* of reference classes such as R . An examination of this more complicated problem shows that the objection being considered confuses convergence in R with *uniform convergence* in a series of R 's.) There is thus no incompatibility between the statement that there is a nonvanishing probability of H occurring with a relative frequency *different* from p , in sets of n elements each (here the reference class is R^*); and the statement that the probability of H occurring in R is *equal* to p . It is true, of course, that in assigning a certain value to N we may be committing an error, because for a time the relative frequencies of H may diverge from p . But this does not establish a *contradiction* in the limit definition of probability; it simply testifies to the difficulty in fixing a value for N . That definition does not supply us with an effective method for obtaining a value for N either by calculation or in some other way; it merely asserts the existence of such a number N . For this reason it has been subjected to various criticisms by finitists, some of which will be considered below.⁶

4. *Formalization of the calculus of probability.*—Two points should be noted in the foregoing presentation of theorems in the calculus. In the first place, the theorems were formulated and explained in terms of an explicit definition of 'probability' as 'the limit of relative frequencies.' And, second, no primitive propositions were specified from which the theorems of the calculus may be derived with the help of the rules of logic. From the standpoint of a formal mathematical discussion, as well as from the point of view of modern methodology, these are defects, and they require a brief discussion.

Principles of the Theory of Probability

If the functor 'prob' is introduced as the defined equivalent of 'the limit of relative frequencies,' every proposition of the calculus is simply a transcription of a theorem in the theory of limits; and every proposition is an analytic statement which can be certified on formal grounds alone. When the calculus is developed in this way, there is no need to supply a special set of primitive sentences: the primitive sentences sufficient for the theory of real numbers are also sufficient to establish every theorem in the calculus.

However, while it may be an advantage to have every theorem of the calculus an analytic sentence of arithmetic, the frequency interpretation of the functor 'prob' is not the only one that is possible. The state of affairs here is strictly analogous to what obtains in geometry. As geometry is employed in physics, the terms 'point,' 'line,' 'plane,' etc., which occur in Euclid, designate certain physical configurations; consequently, the propositions of geometry (such as that the angle-sum of a triangle equals two right angles) formulate measurable relations between physical configurations in exactly the same way as do the propositions of mechanics. But the derivation of geometric theorems from the primitive propositions of Euclid does not depend upon the correlations which happen to be established between terms like 'point' and 'line' and determinate physical configurations. Indeed, formal or demonstrative geometry is not a branch of physics: its theorems cannot be significantly characterized as empirically true or false, because the nonlogical terms in them (e.g., 'point') are uninterpreted. Only after *semantical rules* have been introduced (sometimes also called *co-ordinating definitions*), which correlate such uninterpreted terms with terms employed to designate empirical subject matter, is a formal geometry transformed into a part of natural science. By distinguishing between pure and physical geometry, not only do we avoid confusing questions of formal validity with questions of empirical fact but we also increase the applicability of pure geometry. Alternative co-ordinating definitions may be introduced, so that qualitatively different subject matters may be explored in terms of the same formal system. On the other

Fundamental Theorems in the Calculus of Probability

hand, we may also find that, of the many distinct pure geometries which are logically possible, one system is a more effective means than another for organizing the materials of an empirical subject matter.

Similarly, it is not necessary to interpret 'probability' in terms of frequencies in order to develop a formal calculus of probability. The formalization of the probability calculus is of special importance because of the conflicting interpretations which have been given to the term 'probable,' as well as because of the wide range of opinion concerning the conditions under which probability statements are to be regarded as significant. As in the case of geometry, the probability calculus can be formalized in different ways, depending on what terms are selected as primitive, on the mathematical apparatus which is to be employed in developing it, and also upon the use to which it is to be put subsequently. Only one condition is usually observed in formalizing the calculus of probability: it is required that theorems which have been traditionally regarded as standard ones in the subject (such as the addition theorem or Bernoulli's theorem) be derivable from the primitives of the system.

Only a brief mention is here possible of some of the points of view from which the calculus may be formalized. To understand some of them, the distinctions (made in Vol. I, No. 3) will have to be recalled between the language of a science itself, the syntax language whose object-language is the language of science, and the semantic language of the language of science. Statements in the first language refer to what is commonly called the subject matter of the science, statements in the second refer to the order and possible arrangements of the expressions in the object-language, while statements in the third refer to the relations between an expression in the object-language and its subject matter. One difference between probability calculi arises from the fact that probability statements have been formulated in each of these three languages; another difference is due to the fact that some probability statements are metricized while others are not; and a third difference is due to several attempts to incorporate the probability calculus into

Principles of the Theory of Probability

a general logic which would include both necessary and probable inference.

Two broad classes of calculi of probability may be distinguished: those which provide a metric for the fundamental functor 'prob' and those which do not. Nonmetrical probability calculi may be further distinguished according as they introduce a definite serial order for probabilities or not. The motivation for the construction of nonmetrical calculi has usually been the desire to interpret probability statements in a nonfrequency sense. Such interpretations are often used by writers who have their eyes on the problems of induction and the estimation of evidence in history and legal procedures. A nonmetrical calculus has been developed by Keynes, but the subject is still in a very unsatisfactory and primitive state.

It is possible to formulate a frequency theory of probability both in the object-language and in the semantic language of a science, the choice between these alternatives being largely a matter of convenience. The probability statements of physics occur in its object-language, and most writers who approach the problems of probability from the natural sciences prefer an object-language formulation. Calculi in the object-language usually associate a number p with a probability, such that $0 \leq p \leq 1$. Some writers restrict the values of ' p ' to rational numbers; others permit it to vary in the field of real numbers. A formalized calculus may be developed by taking 'prob(A, R)' as an uninterpreted two-place numerical expression; the logical properties of the expression are then determined by a set of postulates from which, with the help of the usual rules of logic, the standard theorems may be derived. These postulates are abstract in the sense that no restrictions are imposed on the possible interpretations of the functor other than the trivial one that every such interpretation satisfy these postulates. Abstract sets of postulates for probability have been given by Borel, Cantelli, Kolmogoroff, Popper, Reichenbach, and several other writers. (It is also possible to formalize the calculus by taking a one-place numerical functor as primitive, and subse-

Fundamental Theorems in the Calculus of Probability

quently defining a two-place functor in its terms; and there are other possibilities as well.) From an abstract mathematical point of view, the probability calculus is a chapter in the general theory of measurable functions, so that the mathematical theory of probability is intimately allied with abstract point-set theory. This aspect of the subject is under active investigation and has been especially cultivated by Borel, Fréchet, and a large number of French, Italian, Polish, and Russian mathematicians. In object-language calculi, the arguments ' A ' and ' R ' to the numerical expression ' $\text{prob}(A, R)$ ' are usually predicates or predicate variables; in semantic and syntactical calculi the arguments are usually names of sentences or variable designations of sentences. Postulates for metricized semantic calculi are similar to metricized object-language calculi except for the difference in the kind of arguments the functors take. Such semantic postulates have been given by Mazurkiewicz, Popper, and others.⁷

The possibility of interpreting a formal calculus of probability in different ways can be illustrated by the following list: (i) The functor ' prob ' may be interpreted as the limit of relative frequencies in an infinite reference class; the postulates are then transformed into analytic propositions in the theory of real numbers. (ii) The functor may be defined as a relative frequency in a finite reference class; some of the postulates then become analytic propositions in the elementary arithmetic of rational numbers, while others must be suppressed. (iii) The functor may be interpreted, as in the classical Laplacian formulation, as the ratio of the cardinality of two sets of alternatives; the postulates are again converted into analytic propositions in elementary arithmetic. (iv) The functor may be interpreted, as by F. P. Ramsey, as a measure of "partial beliefs," where a degree of probability is a measure of the extent to which a man is prepared to act on a belief. (v) The functor may be interpreted as the ratio of two areas; the postulates become statements in some system of geometry. (vi) A proposal has been made by C. G. Hempel to introduce co-ordinating definitions

Principles of the Theory of Probability

for the functor in such a way that, while it will refer to relative frequencies in a class, the postulates are converted into synthetic statements of physics. (vii) A semantical interpretation has been given to the functor (Reichenbach), according to which it designates the truth-frequency of sentences in certain ordered classes of sentences. (viii) According to another interpretation, which also appears to be semantical, the functor denotes the "degree of falsifiability" of a theory (K. Popper). (ix) The functor has been interpreted as referring to the degree of a unique relation between a "proposition" and a set of premisses. (It is not clear, however, how this view is to be understood. The language in which it is proposed sometimes suggests that the relation is a syntactical one holding between sentences, sometimes that it holds between the "possible facts" which the sentences designate, and sometimes that it is a semantical relation.) Some writers who take this interpretation do not regard the functor as a numerical one (Keynes), while others explicitly do so (H. Jeffreys).⁸

There is another standpoint from which the formalization of the calculus has been undertaken. Leibniz was one of the earliest writers to broach the possibility of a general formal logic in which the calculus of probability would occupy a central place. According to such a project, the standard relations of deducibility between propositions are to be regarded as limiting cases of a more inclusive relation of "probability implication." Many writers after Leibniz, including Boole and Peirce, kept the ideal of such a general logic a live one; and Clerk-Maxwell went to the extent of declaring that "the true logic for this world is the calculus of probability." However, little was done to actualize this possibility until the very recent development of polyvalent logical calculi. The fusion of familiar formal logic and the calculus of probability into one compendent formal system is now actively investigated. But, although much important work has been already done, there is at present still no satisfactory system of such a general logic.

The calculi of n -valued logics of sentences (with n a finite integer) were first developed by J. Lukasiewicz and E. Post. These calculi reduce to the

Fundamental Theorems in the Calculus of Probability

standard sentential calculus (e.g., of *Principia mathematica*) when $n = 2$; that is, when a sentence is permitted to take just two "truth-values," namely, truth and falsity. There are certain partial analogies between the theorems of polyvalent logics and theorems of the probability calculus when the latter is suitably formulated; and a number of writers, including Mazurkiewicz, Reichenbach, and Zawirski, have been exploiting these analogies, with the intent of formalizing the calculus of probability as a polyvalent sentential calculus. Reichenbach's method, stated in outline, consists in interpreting an infinite-valued sentential calculus (with values lying in the interval 0 to 1 inclusive) so that each truth-value is the limit of the relative frequency with which the members of definite sequences of propositions are true. He has urged, moreover, that such an infinite-valued "probability logic" is the one most appropriate for science—on the ground that no empirical statement can be completely verified and can therefore be associated with a "truth-value" which in general is different from 0 (falsity) and 1 (truth). Reichenbach's proposal is not free from technical difficulties, and most students are not convinced that he has achieved a fusion of the probability calculus and a general logic of propositions. For example, Reichenbach's polyvalent "probability logic" contains expressions which apparently are subject to the rules of the ordinary two-valued logic; and it therefore seems that his probability logic is constructed upon a basic two-valued schema. Again, his probability logic is nonextensional, in the strict sense of this term in standard use, while the general system of logic commonly employed in mathematics and physics is extensional; it is therefore not easy to see how the latter can be a specialization of the former.⁹

Interest in the fusion of formal logic and the calculus of probability into one comprehensive system has also been exhibited by physicists impressed by the part which probability statements play in modern quantum theory. In that theory certain noncommutative operators occur, as a consequence of some of the fundamental physical assumptions of the theory; and it is possible to regard such operators as a species of logical multiplication upon propositions dealing with subatomic phenomena. However, instead of superimposing such noncommutative multiplications upon the general framework of a logic of propositions in which multiplication (i.e., and-connection) is commutative, proposals have been made to revise the general logic of propositions. According to some of these proposals, a multiplication which is noncommutative will be governed by the formal rules of the logic of propositions and will not be introduced simply as a consequence of a special physical theory. Attempts to re-write quantum mechanics upon the basis of an altered sentential calculus have been made by J. von Neumann, G. Birkhoff, M. Strauss, and others. But researches in this field have not yet gone far enough to permit a judgment on the feasibility and convenience of the proposed emendations.¹⁰

6. Nonfrequency Interpretations of Probability Statements

We must now briefly consider other interpretations of probability statements than the frequency view proposed earlier in the present section.

1. *The classic conception of probability.*—As already noted, the mathematical theory of probability was first developed in connection with games of chance, and the point of view from which it was cultivated received its classic formulation in the treatise of Laplace. According to the Laplacian view, all our knowledge has a probable character, simply because we lack the requisite skill and information to forecast the future and know the past accurately. A degree of probability is therefore a measure of the amount of certainty associated with a belief: "I consider the word *probability*," De Morgan explained, "as meaning the state of the mind with respect to an assertion, a coming event, or any other matter on which absolute knowledge does not exist." What is required for a mathematical treatment of probability, however, is an exact statement of how this measure is defined; and the classical account is as follows.

Judgments of probability are a function of our partial ignorance and our partial knowledge. We may know that in a given situation the process studied will have an issue which will exhibit one out of a definite number of alternative properties; thus, in tossing a die any one of the six faces may turn up. (These alternative properties have been called the "possible events.") On the other hand, we may have no reason to suppose that one of these events will be realized rather than another, so that, as Laplace remarked, "in this state of indecision it is impossible for us to announce their occurrence with certainty." But a measure of the appropriate degree of belief in a specific outcome of the process can be obtained. We need simply analyze the possible outcome into a set of "equipossible alternatives," and then count the number of alternatives which are favorable to the event whose probability is sought. This measure, the probability of the event, is a fraction whose numerator is the number of favorable alternatives, and whose de-

Nonfrequency Interpretations of Probability Statements

nominator is the total number of possible alternatives, provided that all the alternatives in question are equipossible. Thus, the probability of obtaining six points with a pair of dice is $\frac{5}{36}$, because the dice can fall in any one of thirty-six equally possible ways, five of which are favorable to the occurrence of six points in all. On the basis of this definition, the probability calculus was developed as an application of the theory of permutations and combinations.

Almost all writers on probability in the nineteenth century (e.g., Poisson, Quetelet, De Morgan, Boole, Stumpf), and many contemporary mathematicians (e.g., Borel, De Finetti, Cantelli, Castelnovo), follow Laplace with only relatively minor variations. Because of its historical role, as well as because of its contemporary influence, we shall briefly examine this view.¹¹

a) According to the Laplacian definition, a probability statement can be made only in such cases as are analyzable into a set of equipossible alternatives. But, while in some cases it seems possible to do this, in most cases where probability statements are made this is not possible. Thus suppose that a biased coin is assigned the probability of .63 that it presents a head when tossed; there is no clear way in which this number can be interpreted as the ratio of equipossible alternatives. This is perhaps even more evident for statements like 'The probability that a thirty-year-old man will live at least another year is .945.' It is absurd to interpret such a statement as meaning that there are a thousand possible eventuations to a man's career, 945 of which are favorable to his surviving at least another year. Moreover, the Laplacian definition requires a probability coefficient to be a rational number. But irrational numbers frequently occur as values for such coefficients, and there is no way of interpreting them as ratios of a number of alternatives. Thus, on the basis of certain assumptions, it can be calculated that the probability that two integers picked at random are relatively prime is $6/\pi^2$. This number cannot be made to mean that there are π^2 equally possible ways in which pairs of integers can be picked, six of which are favorable to getting relative primes.

Principles of the Theory of Probability

b) Writers on the subject have not always been clear as to whether they regarded a probability as the measure of a (psychological) belief, or whether they regarded it as a measure of the degree of belief one *ought* to entertain as reasonable. If a probability coefficient is the measure of a degree of actual certainty or the strength of a belief, the addition and multiplication of probabilities require that we determine procedures for combining certainties or beliefs in some corresponding manner. There are, however, no known methods for adding beliefs to one another, and indeed it is difficult to know what could be meant by saying that beliefs are additive. The proposals of Ramsey and De Finetti, to measure strength of beliefs by the relative size of the bets a man is willing to place, are based on a dubious psychological theory; and at least Ramsey's proposal leads directly to a definition of probability in terms of relative frequencies of actions. On the other hand, if probability is a measure of the amount of confidence one ought to have in a given situation, the Laplacian view offers no explanation of the source of the imperative. It is possible, finally, that a probability coefficient is simply a conventional measure of a degree of belief; in that case, however, probability statements turn out to be bare tautologies.

c) According to the Laplacian definition, the alternatives counted must be equally possible. But if 'equipossible' is synonymous with 'equiprobable,' the definition is circular, unless 'equiprobable' can be defined independently of 'probable.' To meet this difficulty, a rule known as the Principle of Indifference (also as the Principle of Insufficient Reason and as the Principle of the Equal Distribution of Ignorance) has been invoked for deciding when alternatives are to be regarded as equiprobable. According to one standard formulation of the rule, two events are equiprobable if there is no known reason for supposing that one of them will occur rather than the other.

It can be shown, however, that, when this form of the rule is applied, incompatible numerical values can be strictly deduced for the probability of an event. An emended form of the rule has been therefore proposed, according to which our *relevant*

Nonfrequency Interpretations of Probability Statements

evidence must be *symmetrical* with respect to the alternatives, which must not, moreover, be divisible into further alternatives on the given evidence. This formulation seriously restricts the application of the Principle of Indifference. Apart from this, however, two points should be noted: A coin which is known to be symmetrically constructed (so that according to the principle its two faces are to be judged as equiprobable) may nevertheless present the head more frequently than the tail on being tossed; for the relative frequency of heads is a function not only of the physical construction of the coin, but also of the conditions under which it is tossed. Second, no evidence is perfectly symmetrical with respect to a set of alternatives. Thus, the two faces of a coin are differently marked, they do not lie symmetrically with respect to the earth's center at the instant before the coin rises into the air, etc. The emended rule therefore provides that it is only the relevant evidence which is to be considered. But if 'relevance' is defined in terms of 'probable,' the circle in the Laplacian definition is once more patent; while, if judgments of relevance are based on definite empirical knowledge, the ground is cut from under the basic assumption of the Laplacian point of view.

d) It is usually assumed that the ratio of the number of favorable alternatives to the number of possible ones (all being equipossible) is also a clue to the relative frequency with which an event occurs. There is, however, no obvious connection between the 'probability of obtaining a head on tossing a coin' as defined on the classical view, and 'the relative frequency with which heads turn up.' For there is in fact no *logical* relation between *the number of alternative ways* in which a coin can fall and *the frequency* with which these alternatives in fact occur. It has, however, often been supposed that Bernoulli's theorem demonstrates such a connection. For as already explained, according to that theorem if the probability of head is $\frac{1}{2}$, then the probability approaches 1 that in n tosses there are approximately $n/2$ heads as n increases. But the supposition that Bernoulli's theorem establishes a relation between a priori (i.e., determined in accordance with the classical definition) and a

Principles of the Theory of Probability

posteriori probabilities (i.e., determined on the basis of relative frequencies of occurrence) is a serious error. It commits those who make it to a form of a priori rationalism. For within the framework of the classical interpretation of the calculus, Bernoulli's theorem simply specifies the relative number of certain types of equiprobable alternatives, each consisting of n tosses; it is no more than a theorem in arithmetic and does *not* permit us to conclude that these alternatives will *occur equally often*. That is to say, only if the expression 'The probability of heads is $\frac{1}{2}$ ' designates a relative frequency of occurrence, can the phrase 'The probability approaches 1' be legitimately interpreted as designating relative frequencies of occurrences.

2. *Probability as a unique logical relation.*—A number of modern writers, conscious of the difficulties in the classical view of probability as a measure of strength of belief, have advanced the view that probability is an objective logical relation between propositions analogous to the relation of deducibility or entailment. According to this version, a degree of probability measures what is often called "the logical distance" between a conclusion and its premisses. The evaluation of a degree of probability therefore depends upon recognizing the inclusion, exclusion, or overlapping of logical ranges of possible facts. Though varying considerably among themselves, something like this view (which has had its forerunners in Leibniz and Bolzano) is central to von Kries, Keynes, J. Nicod, F. Waismann, and several other writers. Only the standpoint of Keynes will be examined here.¹²

For Keynes, probability is a unique, unanalyzable relation between two propositions. No proposition as such is probable; it has a degree of probability only with respect to specified evidence. This relation of probability is not a degree of subjective expectation; on the contrary, it is only when we have perceived this relation between evidence and conclusion that we can attach some degree of "rational belief" to the latter. (As already noted, Keynes's formulation of his view is not unambiguous. His occasional language to the contrary notwith-

Nonfrequency Interpretations of Probability Statements

standing, it does not seem likely that he regards his probability relation as a syntactical one. The present writer is inclined to the opinion that it is a semantical relation.) It is characteristic of Keynes's standpoint that the *secondary* proposition, which asserts that a proposition p has the probability relation of degree a to the proposition h , can and must be known to be true "with the highest degree of rational certainty." Such a highest degree of rational certainty is obtainable, according to Keynes, when we see that the conclusion of a syllogism follows from its premisses, as well as when we see that a conclusion "nearly follows" from its premisses with degree a of probability. However, degrees of probability are not quantitative and are not in general capable of measurement; indeed, according to Keynes, probabilities cannot in general be even ordered serially, although in some cases they are comparable. The comparison of probabilities, whenever this is possible, is effected with the help of the modified Principle of Indifference mentioned above; and the judgments of relevance which the principle presupposes are themselves direct judgments of degrees of probability. In terms of such an apparatus of concepts, Keynes develops a calculus which formulates the relations between comparable probabilities, and finally explains how and under what limited circumstances numerical values may be assigned to degrees of probability.

Although Keynes avoids some of the difficulties of the classical view of probability, his general standpoint has difficulties of its own. Omitting all discussion of the technical difficulties in his calculus, we shall confine ourselves to a brief mention of three central issues.

a) On Keynes's view we must have a "logical intuition" of the probable relations between propositions. However, few if any students can be found who claim for themselves such an intuitive power; and no way has been proposed to check and control the alleged deliverances of such direct perceptions in cases where students claim it. Moreover, the possession or lack of this power is wholly irrelevant in the actual estimation of probabilities by the various sciences. No physicist will seriously

Principles of the Theory of Probability

propose to decide whether two quantum transitions are equiprobable by appealing to a direct perception of probability relations; and, as N. R. Campbell remarked, "anyone who proposed to attribute to the chances of a given deflection of an α -ray in passing through a given film any sense other than that determined by frequency could convince us of nothing but his ignorance of physics."

b) Since on Keynes's view numerical probabilities can be introduced only when equiprobable alternatives are present, he cannot account for the use of numerical probabilities when such an analysis is not possible. Moreover, like the classic interpretation, Keynes cannot establish any connection between numerical probabilities and relative frequencies of occurrences. His theory, when strictly interpreted, is incapable of application to the problems discussed in physics and statistics, and at least from this point of view remains a vestal virgin.

c) On Keynes's view it is significant to assign a probability, with respect to given evidence, to a proposition dealing with a single occasion. For example, it is permissible to declare that on given evidence the probability of a given coin falling head uppermost *on the next toss* is $\frac{1}{2}$. However, the coin, after it is thrown and comes to rest, will show a head or it will show a tail; and no matter what the issue of the given throw is, the probability of obtaining a head on the initial evidence is and remains $\frac{1}{2}$. No empirical evidence is therefore relevant either for the confirmation or for the disconfirmation of that probability judgment, unless we invoke indirectly a relative frequency in a group of statements—which would be contrary to Keynes's intent. But this is to fly in the face of every rule of sound scientific procedure. A conception of probability according to which we cannot in principle control by experiment and observation the probability statements we make is not a conception which recommends itself as germane to scientific inquiry.

Except for matters to be discussed in Section III, the difficulties which have been pointed out for the classical and the logical interpretations of probability do not embarrass the fre-

Problems of Frequency Interpretation of Probability Calculus

quency view. For this negative reason, but especially because it is in accord with scientific practice, the frequency interpretation of probability is the one most suitable for the first class of statements which was specified at the beginning of this section.

III. Unsettled Problems of General Methodology

7. Logical Problems of the Frequency Interpretation of the Probability Calculus

It was shown in Section II that the definition of 'probability' as 'the limit of relative frequency' is *suggested* by common practice in assigning probability coefficients. It has been argued that a probability of $\frac{1}{2}$ for head turning up when a coin is tossed means, roughly, that in half the cases of flipping a coin the head is presented. However, such a statement does not mean that in every two tosses a head turns up just once, for in that case it would be absurd to apply it to an odd number of throws; and we would not regard the statement as erroneous if, after getting a tail, we did not get a head on the next succeeding throw. Accordingly, a less misleading explanation of what a probability of $\frac{1}{2}$ signifies is that in a long run of throws the relative frequency of heads is approximately $\frac{1}{2}$. But it has also been pointed out that a definition of 'probability' as 'the approximate ratio of frequencies in the long run' is not precise and is not suitable for mathematical purposes. A definition in terms of limits, on the other hand, has the requisite precision, and a logically consistent calculus can be developed on such a basis. The convenience and fruitfulness of such a definition for the purposes of a calculus of probability are indeed beyond question.

However, from the point of view of the application of the calculus to empirical matters, it would be of little profit to have a precise mathematical definition of 'probability' if as a consequence every probability statement would acquire a theoretical content which cannot be controlled by acknowledged empirical methods. But an examination of the form of probability statements, when these are interpreted in terms of limits of relative frequencies, seems to indicate that such is indeed the

Principles of the Theory of Probability

case. This may be seen concretely in the following way. Suppose we test the hypothesis that the probability of heads is $\frac{1}{2}$ by flipping the coin a thousand times, and suppose we get a run of a thousand heads. We might be inclined to conclude that the hypothesis has been definitely proved erroneous. However, on that very hypothesis such a run of heads is not excluded, since that hypothesis asserts something about the limiting ratio of heads in an *infinite* class and not in a *finite* one. In general, that hypothesis is compatible with *any* results obtained in any finite number of throws; and, conversely, a given result within a finite class of throws is compatible with any hypothesis about the numerical values of the probability. In short, it seems that no direct statistical evidence obtainable from actual trials (which must obviously be finite in number) can establish or refute a probability statement.

(It should be observed, moreover, that this difficulty is not obviated, as some writers have thought, by employing a less precise definition for 'probability.' For example, if we define it in terms of approximate ratios in long runs, a finite number of observations on the direct evidence for a probability statement will still not suffice to establish or refute it completely and unambiguously.)

The formal argument is as follows. If $f_1, f_2, \dots, f_n, \dots$, is the series of relative frequencies of heads and ϵ a positive number, to say that the probability of getting a head is $\frac{1}{2}$ is to say that $\frac{1}{2}$ is the limit of these ratios. And this means that for every ϵ there is an N , such that for every n , if $n > N$, then the absolute difference of f_n and $\frac{1}{2}$ is less than ϵ . Or, in the notation of modern logic,

$$(\epsilon) (\exists N) (n) [(n > N) \supset (|f_n - \frac{1}{2}| < \epsilon)].$$

This statement contains three quantifiers, the two universal quantifiers 'for every ϵ ' and 'for every n ,' and the existential quantifier 'there is an N .' Because of the presence of the universal quantifiers, this statement cannot be established by examining a finite number of ϵ 's and n 's; or, in the language proposed by Carnap, the statement is not completely confirmable.

This situation is familiar throughout science. For example, the statement 'All bodies attract each other inversely as the square of their mutual distances' is not completely confirmable either. But, according to strict logic and textbook scientific method, this latter statement is capable of *complete disproof* by *one negative instance*; and it is usually said, therefore, that, although we

Problems of Frequency Interpretation of Probability Calculus

can never be in the position to assert the truth of universal statements, we may be in the position to assert their falsity.

However, probability statements do not fall under this dictum. For, in order to completely disprove such a statement, its formal contradictory would have to be completely confirmed. But the formal contradictory of the specimen probability statement is: There is an ϵ , such that for every N , there is an n , such that $n > N$, and the difference between f_n and $\frac{1}{2}$ is not less than ϵ . In symbolic notation

$$(\exists \epsilon) (N) (\exists n) [(n > N) \cdot (|f_n - \frac{1}{2}| \geq \epsilon)].$$

However, this statement also contains a universal quantifier, namely, 'for every N ,' so that it cannot be completely confirmed.

In sum, therefore, a probability statement can be neither completely confirmed nor completely disconfirmed.

Many writers have therefore concluded that probability statements interpreted in terms of relative frequencies are devoid of empirical meaning because what they assert cannot be controlled by determinate empirical procedures. Such a conclusion, if it were warranted by the facts, would be fatal to a frequency interpretation of probability. For it is a cardinal requirement of modern science that its statements be subject to the criticism of empirical findings. This simply means that not every state of affairs can be confirmatory evidence for a given statement and that observable states of affairs must be specifiable which would be acknowledged as incompatible with its truth. On the other hand, such a conclusion is paradoxical because in actual practice probability statements interpreted in terms of frequencies *are* accepted or rejected on the basis of empirical evidence; and no one seriously doubts that we order affairs of everyday living, of industry, and of science with their help.

What is required, therefore, is a specification of the semantical and pragmatic *rules* in accordance with which probability statements are accepted and rejected on the basis of empirical findings. Although a complete set of rules cannot be given at present, so that the problem is in a very unsettled condition, it is believed that the following observations will be found relevant to the issue raised.

Principles of the Theory of Probability

a) An objection often made to the limit definition of probability is that limits, in the strict sense of the term, do not exist for empirically determined relative frequencies and that in actual statistical material the ratios of frequencies fluctuate more or less widely. Such an objection, however, should in all consistency be made also to the use of general mathematical analysis in the natural sciences. For the limit concept is employed not only in probability but elsewhere also. For example, the masses or centers of gravity of bodies are frequently calculated with the help of the integral calculus, and the integrations are performed on the assumption that the mathematical functions which specify the density of the bodies are continuous; the calculation of these quantities thus involves limits at several places. Moreover, the assumption of a continuous density distribution is not warranted by our present theories of matter as discontinuous. We do not, however, reject the powerful tools of analysis for these reasons. An even simpler illustration of the use of limits occurs in measurement, against which no one seems to raise difficulties of the sort indicated. Every actual measurement, for example, of the length of the diagonal of a square, yields a *rational* number; nonetheless, in theoretical work we frequently employ irrational numbers, such as $\sqrt{2}$, for specifying lengths; and irrational numbers involve limit notions. The reason for employing terms involving limits in probability theory, as elsewhere, is the same: we thereby obtain powerful and economical methods in making *mathematical transformations*. And the reason why the use of such "calculus terms" and the procedures requiring them is countenanced in the natural sciences (even when direct empirical evidence and theoretical considerations indicate that the conditions for their use are not fully satisfied) is that we know how to *correlate* with them *groups* of directly measured magnitudes lying in certain intervals.

b) It is indeed a naïve conception of scientific method according to which the statements of science (whether singular or general) are to be rejected on the ground of a single negative instance. It was pointed out in Section I that, even in the exact

Problems of Frequency Interpretation of Probability Calculus

sciences of measurement, the numerical values of magnitudes as predicted by a theory are not in precise agreement with the numerical values obtained by actual measurement and observation; the theory of errors had its genesis in the study of just such situations. A theory is not in general dismissed as false or worthless because the confirmation of its predictions by observation is only approximate—even though *formally* every deviation from a predicted value of a magnitude is a negative instance for a theory. The amount of allowable deviation between predicted and observed values is not specified by the theory itself, and even a “large” deviation may not be decisive against the theory. The reasons for this are twofold: An empirically testable consequence of a theory does not follow from the theory *alone*, but from it conjoined with statements reporting matters of observation and possibly other theories. Consequently, an apparent negative instance for a theory may be argued to be incompatible not with *it* but only with some of the other premisses of the argument; and by a suitable alteration in the assumptions from which the testable consequences are drawn, the theory itself may be retained as in accordance with the “facts.” Second, the amount of allowable deviation between predicted and observed values of a magnitude may be a function of a number of variable factors, such as the number of observations made, the purposes for which the inquiry is conducted, the kind of activity which the theory is intended to coordinate and foretell, or the character of the instruments by means of which the testing is carried on. These factors cannot in general be completely enumerated or specified in detail, although those who conduct researches have been trained to make allowance for them in the concrete cases before them.

A crude illustration of this second point, for the case when *direct statistical evidence* for a probability statement is evaluated, can be constructed as follows. Suppose the hypothesis that $\frac{1}{2}$ is the probability of obtaining a head with a coin is to be tested, by tossing it one hundred times. According to the hypothesis, we may expect approximately fifty heads. If heads turned up forty-nine times, we would regard this as confirming

Principles of the Theory of Probability

the hypothesis; if heads turned up forty-five times, this may still be regarded as confirmatory; but if heads turned up only twenty times, we might suspect that the coin is loaded and doubtless propose a different value for the probability of getting a head. That is to say, somewhere between getting twenty and getting fifty heads in one hundred throws, we might fix a value such that a frequency less than it is to be taken as disconfirming the hypothesis of $\frac{1}{2}$. In other words, the actual hypothesis which would be tested under these circumstances is that the relative frequency of heads lies in an *interval* $\frac{1}{2} \pm \delta$, where the positive number δ is not fixed once for all but varies with circumstances. Now the probability of obtaining deviations of specified magnitudes from $\frac{1}{2}$ (on the assumption that *sets* of such trials are repeated indefinitely) can be calculated with the help of Bernoulli's theorem; and *this* probability depends upon the initial hypothesis that p is the probability of getting a head as well as upon the number n of throws which are made. Hence δ will often be a function of p and n . But it may be a function of other factors as well: e.g., of our knowledge of the physical construction of the coin and of the circumstances under which it is thrown, of the size of our fortune if we are gambling, etc. The definition of 'probability' in terms of 'limit' is therefore important for the purpose of constructing a consistent and powerful *calculus*. The calculus itself is instrumental in effecting transitions from one set of empirically controllable statements to other such sets. Provided that appropriate semantical and pragmatic rules are instituted for applying the calculus, it is not a serious objection to it that some of its terms cannot be taken as *descriptive* of the subject matter of science.

In modern theoretical statistics various methods have been devised for evaluating the goodness of an estimate of parameters, such as p , which characterize a hypothetical infinite population. According to the older methods of Lexis, the *aggregate sample* on the basis of which the estimate is made requires to be analyzed into sets of elements which are similar in certain relevant respects; the stability or fluctuation of the estimate in these various groups is then studied. In the more recent methods of R. A. Fisher, J. Neyman, and others, "measures of credibility" are introduced, some of which are

Problems of Frequency Interpretation of Probability Calculus

carefully distinguished from probabilities. According to these methods, the values assigned to the hypothetical probabilities must meet explicitly stipulated conditions of stability under repeated samplings, and must also make these measures of credibility a maximum. It is not possible at this place to enter into this subject in greater detail.

In many cases, no determinate numerical value can be assigned to a probability, not because a frequency interpretation of probability statements is not relevant but because relevant statistical information is lacking. For example, the proposition is often asserted that when the barometer falls it is highly probable that it will rain, although no numerical value is usually specified for this "high probability." Such a statement clearly means that the relative frequency of rain within a few hours, in the class of cases where the barometer falls, is greater than $\frac{1}{2}$ and possibly close to 1. But lacking precise statistical information, the high probability is assigned and confirmed on the basis of general impressions as to the behavior of the weather. In still other cases, such as that involved in estimating the probability of a witness speaking the truth, the statistical data may be even more meager, and the general impressions upon which we base our estimates may be highly unreliable and even worthless.

We have no final assurance that a hypothesis as to the numerical value of a probability is a correct one. However, the method of inquiry we employ is a self-corrective one, and in general we place greater reliance upon our rules of procedure and their *net* results than upon particular conclusions obtained. We are not in a position to assert with finality that the empirical frequencies we obtain do converge to a limiting value. But as Peirce and more recently Reichenbach have pointed out, if these ratios do tend to remain within certain narrow intervals, we can discover what those intervals are by a repeated and systematic correction of the estimates which are suggested by the samples we continue to draw.

c) Thus far, only the *direct statistical* evidence for a probability statement has been considered. But it was explained in Section II that, whenever such a statement is part of an in-

Principles of the Theory of Probability

clusive system of statements, the evidence may be indirect and even of a nonstatistical character. There are, indeed, the following possibilities: Let S be a probability statement of the form ‘ $\text{prob}(A, R) = p$ ’; and let Σ be a class of statements which in general will contain *singular* statements reporting matters of observation (e.g., statements which ascribe a property to a definite space-time region), as well as general or theoretical statements some of which may have the form of probability statements.

(i) The value of p in S may be estimated directly from statistical evidence concerning the frequency of A in R ; this case has already been considered.

(ii) From S and Σ another probability statement S_1 may be derived which may be tested by direct statistical evidence for S_1 . Thus, if S ascribes the probability of $\frac{1}{2}$ to a coin falling head uppermost, S_1 may ascribe the probability of $\frac{1}{4}$ to the coin falling heads up twice in succession.

(iii) From S and Σ a statement S_2 may be derived which is nonstatistical. Thus, let S ascribe the probability of $\frac{1}{2}$ to an atom in a state with a magnetic moment of one suffering a transition into a state with a magnetic moment of two when a deflecting field is introduced; then S_2 may assert that the intensity of the ionic current across the path of the molecular beam is of a specified magnitude. In this case, no problems arise in connection with the empirical control of probability statements which do not arise in connection with other statements of science.

(iv) The value of p in S may be deduced from Σ . Thus, if Σ contains the Schrödinger equation together with a number of boundary conditions, we can calculate the numerical value of the probability that an atom in a given space-time region will be in a specified state.

Although some of the formal logical problems in connection with cases (iii) and (iv) have not been thoroughly worked out, such cases do occur. And it is evident from them that the correctness of a given hypothesis as to the numerical value of a probability may be controlled in much the same way as the

Problems of Frequency Interpretation of Probability Calculus

more familiar nonstatistical hypotheses of science are controlled.

In recent years, following a suggestion of Poincaré, what is sometimes known as "a causal theory of probability" has been developed by G. D. Birkhoff, E. Hopf, and others. The main idea of these researches is the deduction of a probability value (e.g., the probability of the ball in roulette coming to rest in a red sector) from underlying dynamical assumptions governing the average values of certain quantities with increase of time. It is incorrect to maintain, as some have done, that a probability value can be deduced from a dynamical theory which contains no material assumptions about the distribution of frequencies or average values. Nevertheless, these researches, apart from their technical interest, emphasize one very important point: An estimate of a probability which is made simply on the basis of unanalyzed samples or trials is not likely to be a safe basis for prediction. If nothing is known concerning the mechanism of a situation under investigation, the relative frequencies obtained from samples may be poor guides to the character of the indefinitely large population from which they are drawn. Thus, because we know very little about the mechanism of historical changes in human societies, it would be unsafe to use the life-probabilities computed in the first quarter of the present century as a basis for conducting a life insurance business in America two centuries hence. On the other hand, because we know something about the mechanism of biological heredity, a relatively few observations on the number and types of descendants of a plant may suffice to confirm hypotheses about the probability of certain types recurring. Again, we assign a value to the probability of getting heads on a freshly minted coin with great assurance, even before making any actual trials with it, because the homogeneity of the products of national mints, as well as of the conditions under which the coin would be thrown, are fairly well established. In general, therefore, the amount and kind of evidence required for probability statements depend on their interconnections with the body of our knowledge and theories at a given time.¹³

Principles of the Theory of Probability

(v) Some writers, notably Reichenbach, have maintained that, while probability statements are incapable of complete confirmation or disconfirmation, nevertheless a degree of probability (in the frequency sense) can be attached to them. Such a proposal, it turns out, involves a hierarchy of probabilities, in which every probability statement on one level is subject matter for probability statements on a higher level; it is a conception which has stimulated the development of a "probability logic" referred to in Section II. If such a proposal could be implemented with an unambiguous and convenient method for assigning probabilities to probability statements, it would go a long way to solving definitively the logical problem to which the present section has been devoted. Reichenbach's writings make important contributions toward formulating such a method. However, a probability statement is a *general* statement, as was explained on page 52; and we reserve the discussion of the probability of general statements, hypotheses, or theories for § 8.

8. Probability and Degree of Confirmation or Weight of Evidence

At the outset of Section II two classes of statements containing the term 'probable' were distinguished. The members of the first class have now been shown to require a frequency interpretation, and the statements in it are subject to the rules of the calculus of probability. We shall now inquire whether the second class is similar to the first in these respects.

A common objection to the frequency theory of probability is that, although probability statements concerning single occasions or single propositions are often asserted and debated, it is meaningless to assert such statements in terms of the frequency theory. For example, writers like Keynes have urged that such statements as 'It is probable on the evidence that Caesar visited Britain' and 'The evidence makes it improbable that all crows are black' cannot be analyzed in terms of relative frequencies; and they have concluded that a conception of probability is involved in them which is different from, and "wider" than, the frequency view. Frequentists have retorted, quite

Probability and Degree of Confirmation or Weight of Evidence

rightly, that such statements *are* without meaning, if they *literally* attribute a probability in the frequency sense to a single proposition; but frequentists have also urged that such statements do have significance if they are understood as *elliptic* formulations.

There is little doubt that many probability statements which are apparently about single propositions are incomplete formulations and that, when they are suitably expanded, they conform to the conditions required by the frequency theory. On the question, however, whether *all* probability statements about single propositions are to be analyzed in this way there is considerable difference of opinion. This disagreement not only divides frequentists from nonfrequentists like Keynes but it also represents a division among those who subscribe to a frequency interpretation for the first class of statements previously mentioned.

This difference of opinion concerning the range of applicability of the calculus of probability has a long history. Earlier writers on the subject believed that the calculus was the long-sought-for instrument for solving all problems connected with estimating the adequacy of evidence. In particular, it was maintained that the problems associated with establishing general laws on the basis of examined instances and with obtaining some measure for the reliability of predictions (the traditional problems of induction) were part of the subject matter of the mathematical theory of probability. Bayes's theorem and the Rule of Succession were commonly employed for these purposes, and Jevons explicitly regarded induction as a problem in inverse probabilities. On the other hand, writers like Cournot and Venn, two of the earliest writers to propose a frequency interpretation of the probability calculus, were equally convinced, though for different reasons, that the calculus was not relevant to the problems of induction. More recently, Keynes and Reichenbach, arguing from diametrically opposite standpoints, agree on the point that the term 'probable' can be given a consistently univocal meaning; and Reichenbach has given the most complete account at present available of how to extend the fre-

Principles of the Theory of Probability

quency view to the consideration of the probability of scientific theories. But other contemporary frequentists, such as Carnap, von Mises, Neurath, and Popper, though supporting the frequency interpretation for a very large class of probability statements, do not believe such an interpretation is appropriate for every statement which contains the word 'probable.' This latter group of writers rejects the notion of a "logical probability" as developed by Keynes and others; but it distinguishes between 'probable' employed in the sense of 'relative frequency' and 'probable' employed in the sense of 'degree of confirmation' or 'weight of evidence.'

It is possible, therefore, to distinguish writers on probability according to the following schema: (1) Writers who interpret 'probable' in a *univocal* sense; such writers differ among themselves according as they accept the classical view, the view of probability as a unique logical relation, or the frequency view. (2) Writers who do not believe that the term 'probable' can be interpreted in precisely the same manner in every one of the contexts in which it occurs.

The present state of research, therefore, leaves the issue unsettled as to the scope of the frequency theory of probability. We shall examine the points at issue, but our conclusion will of necessity have to be highly tentative. We shall concern ourselves explicitly with statements ascribing a probability to a theory, because of lack of space; but the discussion will apply without essential qualifications to probability statements about singular statements like 'Caesar visited Britain,' whenever such probability statements are not analyzable as elliptic formulations involving relative frequencies. By 'theory' will be understood any statement of whatever degree of complexity which contains one or more universal quantifiers, or a set of such statements.

1. *The probability of theories.*—We begin with examining the proposal to interpret probability statements about theories in terms of relative frequencies; and, since Reichenbach has expounded this proposal more fully than anyone else, we shall examine his views. Reichenbach has given two distinct but

Probability and Degree of Confirmation or Weight of Evidence

allied methods for defining "the probability of a theory." The first of these methods has received an improved formulation by C. G. Hempel, which avoids serious difficulties present in Reichenbach's own version. It should be noted that the definitions given by both methods are semantic ones.

a) Let T be some theory, for example, the Newtonian theory of gravitation. Let C_n be a class of n singular statements, each of which specifies an initial state of a system. (For from T alone, without the specification of initial conditions, no empirically controllable consequences can be obtained; thus, the mass, initial position, and velocity of a planet must be assigned before a future state of the planet can be predicted.) From every such statement with the help of T , other statements may be derived, some of which are empirically controllable by an appropriate observation. Therefore, let E_n be the class of n such singular statements derived from C_n with the help of T . We suppose that a one-to-one correspondence is established between the elements of C_n and E_n ; and without loss of generality we shall suppose that every statement in C_n is true. (From a single statement in C an indefinite number of statements belonging to E may be derived; but we can simply *repeat* a statement in C for every one of the distinct consequences drawn from it.) Let $\text{nu}(E_n)$ be the number of statements in E_n which are true. The relative frequency with which a statement in E_n is true when its corresponding statement in C_n is true is given by $\text{nu}(E_n)/n$. Suppose now that n increases indefinitely, so that C_n will include all possible true initial conditions for T , while E_n will include all the possible predictions which are made from them with the help of T . The numerical expression

$$\text{prob}(E, C) = \lim_{n \rightarrow \infty} \frac{\text{nu}(E_n)}{n}$$

will then be the probability that the consequences, obtained with the help of T from appropriate initial conditions, are true. This, in essence, is Reichenbach's first method of assigning a probability to a theory T .

Although the foregoing exposition requires supplementation

Principles of the Theory of Probability

in several ways, there seems to be little question that a precise definition for 'the probability of a theory' can be given on a relative frequency basis. It is, however, by no means evident that such a definition formulates the concept people seem to be employing when they discuss the probability of theories.

(i) On the foregoing definition the probability of a theory is the limiting value of relative frequencies in an infinite ordered class *E*. This value is therefore independent of the *absolute number* of true instances in *E*, and is also independent of the absolute or relative number of instances in *E* which we know to be true *at a given time*. However, we often do say that on the basis of *definite evidence* a theory has some "degree of probability." Thus, a familiar use of this phrase permits us to say that, because of the accumulated evidence obtained since 1900, the quantum theory of energy is more probable today than it was thirty years ago. The foregoing definition is not suitable for this use of the phrase.

(ii) Because the probability of a theory is defined as the limit of relative frequencies, the probability of a theory may be 1, although the class *E* of its empirically confirmable consequences contains an infinite number of statements which are in fact false. This conclusion could follow even if some of these exceptions to the theory are ruled out as not being genuine negative instances (see the discussion of this point in §7). But, according to the familiar usage of 'probability of a theory' already referred to, if a theory did have an infinite number of exceptions, not only would not a "high degree of probability" be assigned to it: it would be simply rejected.

(iii) It is difficult to know how even the approximate value of the probability of a theory, in Reichenbach's first sense, is to be determined. The situation here is not quite the same as for the probability statements which occur *within* a natural science and which have been already discussed in §7. In the present case it does not seem possible to obtain other than direct statistical evidence for an assigned numerical value; for it is not apparent how a statement about the probability of theories can be part of an inclusive system, so that the statement might

Probability and Degree of Confirmation or Weight of Evidence

possibly be confirmed indirectly, perhaps even by nonstatistical evidence. Reichenbach's proposal of a hierarchy of probabilities, according to which the probability of a probability statement may be estimated, postpones this problem by referring it to a higher level of probabilities; but postponing a problem does not solve it.

b) The second method proposed by Reichenbach for assigning a probability to a theory in a frequency sense depends upon the first method. The theory T under consideration will now be regarded as an element in an infinite class K of theories. These theories are supposed to be alike in some respects and unlike in others; and the theory T will share with a number of others in K a certain definite property P . (The following crude illustration may help fix our ideas: Suppose T is the Newtonian theory, and K the class of possible theories dealing with the physical behavior of macroscopic bodies. P may then be the property that the force functions in the theory are functions of the coordinates alone.) The probability of the theory T is then defined as the limit of the relative frequency with which theories in K , possessing the property P , have a probability in Reichenbach's first sense which is not less than a specified number q .

We can comment only briefly on this proposal.

(i) Although it is easy to introduce the reference class K and the property P in the formal definition, in practice it is by no means easy to specify them. The class K must not be selected too widely or arbitrarily, but no way is known for unambiguously grouping together a set of allegedly "relevant" theories. The difficulty is even greater in specifying the property P for a concrete case. We might wish to say, for example, that the theory of relativity is more probable than the Newtonian theory. But just what is the property P in this case on the basis of which they are to be distinguished?

(ii) We do not at present possess a sufficiently extensive collection of theories, so that appropriate statistical inquiries cannot be made with respect to them in accordance with this proposal. This proposal therefore completely lacks practical relevance. Indeed, there is some ground for suspicion that the pro-

Principles of the Theory of Probability

posal would be feasible only if, as Peirce suggested, "universes were as plentiful as blackberries"; only in such a case could we determine the relative frequency with which these different universes exhibit the traits formulated by a theory under consideration.

(iii) If we could assign a probability value to a theory according to the first of Reichenbach's two proposals, there would be little need for estimating its probability by the second method. It is consistent with these proposals that a theory which has a probability of 1 on the first method, has the probability of only 0 on the second method. But since we are, by hypothesis, interested in that *one* theory, of what particular significance is it to know that theories of such a type have almost all their instances in conformity with the facts with only a vanishingly small relative frequency? This second proposal, like the first, does not therefore formulate the sense of those statements which assign a "degree of probability" to a theory on the basis of *given* finite evidence. For this second proposal does not permit us to talk *literally* about the degree of probability which *one definite theory* has on the evidence at hand; and it is just this which is intended when the evidence for a theory at one time is compared with the evidence at another time.

2. *Degree of confirmation or weight of evidence.*—These difficulties with the two proposals for assigning a probability to a theory, in the relative frequency sense of the term, are serious enough to have led competent students to seek a different interpretation for such statements. Guided by the actual procedure of the sciences, a long line of writers have urged that a different concept is involved in such statements from the one specified by the frequency theory of probability. This concept has been designated as "degree of confirmation" or "weight of evidence," in order to distinguish it from the various interpretations given to the term 'probable.' We shall briefly explain what is meant by 'degree of confirmation' and discuss some of the problems which center around its use. •

The initial task which must be performed before a satisfactory account of 'degree of confirmation' can be given is a careful

Probability and Degree of Confirmation or Weight of Evidence

analysis of the logical structure of a theory in order to make precise the conditions under which a theory may be confirmed by suitable experiments. This has been partially done by Carnap with considerable detail and refinement. We shall, however, not reproduce the results of his analyses, and shall employ distinctions inexactly formulated but which are familiar in the literature of scientific method. In particular, we shall take for granted the following, of which use has already been made: No theory (or for that matter no singular statement) can be established completely and finally by any finite class of observations. But a theory can be tested by examining its instances, that is, the singular sentences *E* derived with the help of the theory from the sentences *C* stating the initial conditions for the application of the theory. Both *C* and *E* may increase in number; but, while theoretically there are an infinite number of instances of a theory, no more than a finite number will have been tested at any given time. Indeed, a theory is said to be capable of being confirmed or verified only incompletely, just because no more than a finite number of its instances can be actually tested. The instances may be confirmed by observation, in which case they are called the *positive instances* for the theory; or they may be in disaccord with the outcome of observations, in which case they are called the *negative instances*.

We shall assume for the sake of simplicity that there are no negative instances for a given theory *T*. Then as we continue the process of testing *T*, the number of positive instances will usually increase. Now it is generally admitted that, by increasing the positive instances, the theory becomes more securely established. What is known as 'the weight of evidence' for the theory is thus taken to be a function of the number of positive instances. And we may accordingly state as a preliminary explanation of what is meant by 'the degree of confirmation' for a theory that the degree of confirmation increases with the number of the positive instances for *T*.¹⁴

This explanation is, of course, far from precise; but at present no precise definition for the term is available. As matters stand, the term is used in a more or less intuitive fashion in the actual

Principles of the Theory of Probability

procedures of testing theories. It would obviously be highly desirable to have carefully formulated semantical rules for employing the term; but there is no early prospect that the rules for weighing the evidence for a theory will be reduced to a formal schema. The following observations, however, indicate some of the conditions under which the weighing of evidence is carried on, and will contribute something to making more precise the meaning of 'degree of confirmation.'

a) It does not seem possible to assign a quantitative value to the degree of confirmation of a theory. Thus, at one stage of investigation a theory *T* may have twenty positive instances in its favor, while at a later stage it may have forty such instances. While the degree of confirmation of *T* at the second stage would in general be acknowledged as greater than at the first stage, it is nevertheless not appropriate to say one degree of confirmation is twice the other. The reason for this inappropriateness is that, if degrees of confirmation could be quantized, all degrees of confirmation would be comparable and be capable therefore of a linear ordering. That this does not seem to be the case is suggested by the following hypothetical situation.

Suppose that the positive instances for *T* can be analyzed into two nonoverlapping classes K_1 and K_2 , such that the instances in K_1 come from one field of inquiry and those in K_2 from another field. For example, if *T* is the Newtonian theory, K_1 may be the confirmatory instances for it from the study of planetary motions, while K_2 may be those coming from the study of capillarity phenomena; each set of instances is in an obvious sense qualitatively dissimilar from the other. Now imagine the following possibilities as to the number of instances in K_1 and K_2 :

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
K_1	50	50	100	101	99	100	200	100	198
K_2	0	50	0	49	52	90	0	100	2
<i>E</i>	50	100	100	150	151	190	200	200	200

The last row of figures gives the total number of positive instances for *T*. These nine possibilities are arranged in order of

Probability and Degree of Confirmation or Weight of Evidence

increasing number of positive instances. Would we say, however, that this order also represents the order of increasing degrees of confirmation?

It would generally be granted that for both P_2 and P_3 the degree of confirmation is greater than for P_1 , simply because of the total number of positive instances. On the other hand, many scientists would be inclined to assign a greater degree of confirmation to P_2 than to P_3 , even though the total number of positive instances is the same in these cases. And the reason they would give is that in P_2 there are *different kinds* of instances, while in P_3 there is only one kind. For this reason also P_4 would be assigned a higher degree of confirmation than P_1 , even though the total number of positive instances in the former case is less than in the latter case. Again, P_4 and P_5 would often be assigned the *same* degree of confirmation, even though the total number of instances is different in these cases, because the relative number of instances of each kind is approximately the same. Finally, P_6 and P_7 would often be regarded as *incomparable* with respect to their degrees of confirmation, because of the disparity in the relative number of different kinds of instances.

Variety in the kinds of positive instances for a theory is a generally acknowledged factor in estimating the weight of the evidence. The reason for this is that experiments which are conducted in qualitatively different domains make it easier to control features of the theory whose relevance in *any* of the domains may be in question. Hence, by increasing the possibility of eliminating what may be simply accidental successes of a theory under special or unanalyzed circumstances, the possibility of finding negative instances for the theory is increased. In this way of conducting experiments, the theory is subjected to a more searching examination than if all the positive instances were drawn from just one domain. A large increase in the number of positive instances of one kind may therefore count for less, in the judgment of skilled experimenters, than a small increase in the number of positive instances of another kind. It follows, however, that the degree of confirmation for

Principles of the Theory of Probability

a theory seems to be a function not only of the absolute number of positive instances but also of the kinds of instances and of the relative number in each kind. It is not in general possible, therefore, to order degrees of confirmation in a linear order, because the evidence for theories may not be comparable in accordance with a simple linear schema; and a fortiori degrees of confirmation cannot, in general, be quantized.

Indeed, the foregoing hypothetical situation is only a highly simplified outline of the considerations which are usually taken to be relevant in estimating the weight of the evidence for a theory. Among other factors usually considered is the precision with which the confirmable consequences of a theory are in agreement with experimental findings. Although, as has been repeatedly explained, a theory is not rejected simply because perfect agreement between predicted and experimentally determined magnitudes does not occur, the more closely the observed values center around the theoretically expected magnitudes, the greater weight is usually attached to the supporting observations for a theory. Furthermore, evidence for a theory often consists not only of its own positive instances but also of the positive instances for *another* theory, related to the first within a more inclusive theoretical system. The number of direct positive instances may in such cases be regarded as of small importance, in comparison with the fact that support is given to the theory by the accumulated positive instances for the inclusive system.

b) How large must the number and kinds of positive instances be in order that a theory can be taken as adequately established? No general answer can be given to such a question, since the answer involves practical decisions on the part of those who conduct a scientific inquiry. There is an ineradicable conventional element among the factors which lead to the acceptance of a theory on the basis of actual evidence at hand. It is always theoretically possible to demand further evidence before agreement is reached that a theory has been sufficiently well tested. However, the practical decision is in part a function of the contemporary scientific situation. The estimation

Probability and Degree of Confirmation or Weight of Evidence

of the evidence for one theory is usually conducted in terms of the bearing of that evidence upon alternative theories for the same subject matter. When there are several competing theories, a decision between them may be postponed indefinitely, if the evidence supports them all with approximately the same precision. Furthermore, the general line of research pursued at a given time may also determine how the decision for a theory will turn out. For example, at a time when a conception of discontinuous matter is the common background for physical research, a theory for a special domain of research formulated in accordance with the dominant leading idea may require little direct evidence for it; on the other hand, a theory based on a continuous notion of matter for that domain may receive little consideration even if direct empirical evidence supports it as well as, or even better than, it does the alternative theory.

In particular, the acceptance of definite numerical values for probabilities also involves practical decision, for which no general rules can be given. As already explained, such numerical values are often computed on the basis of more or less comprehensive theoretical systems, and the confidence which we have in the correctness of those values depends on the confidence we have in those systems. It may happen that we can determine the value of a probability with only small accuracy by a theory which has a relatively high degree of confirmation, while a different value may be computed with great precision by an alternative theory with an inferior degree of confirmation. The supposition that in such a case the dilemma can be resolved by a clear-cut method neglects the human and accidental factors which determine the history of science. Certainly no mathematical or logical formula can be given which would mechanically supply a coefficient of weight for the correctness of the decisions which are made in many analogous cases.

c) Assuming that these desultory observations are based on the study of actual scientific procedure, it may be asked why it is that we seem to feel that theories with a greater degree of confirmation deserve our confidence on logical grounds more than those with less—whenever such comparisons can be made.

Principles of the Theory of Probability

Why, in other words, should a theory be regarded as "better established" if we increase the number and kinds of its positive instances?

Perhaps a simple example will help suggest an answer. Suppose a cargo of coffee is to be examined for the quality of the beans. We cannot practically examine every coffee bean, and so we obtain some sample beans. We do not, however, sample the cargo by taking a very large number of beans from just one part of the hold; we take many relatively small samples from very many different parts of the ship. Why do we proceed in this way? The answer seems to be that our general experience is such that, when we conduct our samplings in this manner, we approximate to the distribution of qualities in the entire hold; and, in general, the larger our individual samples and the more diversified our choice of the parts of the ship from which they are taken, the more reliable (as judged by subsequent experience) are the estimates we form. It is at least a plausible view that in testing a theory we are making a series of samplings from the class of its possible instances. A theory is "better established" when we increase the number and kinds of its positive instances, because the *method* we thereby employ is one which our general experience confirms as leading to conclusions which are stable or which provide satisfactory solutions to the specific problems of inquiry. At any rate, this was the answer which Charles Peirce proposed to the so-called "problem of induction," and which has been independently advanced in various forms by many contemporary students of scientific method (e.g., M. R. Cohen, J. Dewey, H. Feigl, O. Neurath, and many others). As Peirce succinctly put the matter, "Synthetic inferences are founded upon the classification of facts, not according to their characters, but *according to the manner of obtaining them*. Its rule is that a number of facts obtained in a given way will in general more or less resemble other facts obtained in the same way; or, *experiences whose conditions are the same will have the same general characters*." A degree of confirmation is thus a rough indication of the extent to which our general *method of procedure* has been put into operation. While no prob-

Probability and Degree of Confirmation or Weight of Evidence

ability in a frequency sense can be significantly assigned to any formulation of our method (because it is that very method which is involved in estimating and testing such probabilities), scientific inquiry is based upon the assumption, which is supported by our general experience, that the method of science leads to a proportionately greater number of successful terminations of inquiry than any alternative method yet proposed.¹⁶

Attempts to find a systematic answer to "the problem of induction" within the framework of a theory of probability, though often made, have not in general been regarded as successful. The *process* of induction has been usually conceived as the search for more or less stable and pervasive relations between properties of objects; and the *problem* of induction has been taken to be the discovery of a principle (the principle of induction) which would "justify" the various conclusions of that process. Stated in this way, it is rather difficult to know just how the "problem" is to be conceived in empirical terms. On the face of it, the "problem" seems to involve a futile infinite regress; and indeed the Achilles heel of attempted solutions of it has usually been the status of the proposed principle of induction: how is the principle itself to be "justified"? The number of different types of answers which have been given to this last question is relatively small; among them are the following: the inductive principle is a synthetic a priori proposition concerning the nature of things in general, it is an a priori proposition concerning the fundamental constitution of the human mind, it is a generalization from experience, and it is a "presupposition" or "postulate" of scientific procedure. It would take too long to examine these answers in detail. It is perhaps sufficient to note that the first two involve positions incompatible with the conclusions of modern logical research; that the third commits a *petitio principii*; and that the fourth, assuming it to have a clear meaning, cannot make of the proposed inductive principle a "justification" of the procedure of science or of its conclusions, since according to this answer the principle is simply an *instrument* of scientific procedure. The position taken in the present monograph is that no antecedent principle is required to justify the procedure of science, that the sole justification of that procedure lies in the specific solutions it offers to the problems which set it into motion, and that a *general* problem of induction in its usual formulation does not exist. Since the notion of the probability of theories (in the specific senses discussed above) has been found to involve serious difficulties, and since the degree of confirmation for a theory has been argued to indicate the extent to which the theory has been tested by the procedure of science, the problem of induction which the present writer recognizes as genuine is the formulation of the general features of scientific method—of the method which, in short, leads to a proportionately greater number of successful terminations of inquiry than the number which other methods may have to their credit.

Principles of the Theory of Probability

One brief final remark: It has been customary in the traditional discussions of scientific theories to seek grounds for our knowledge of their *truth* or at least of their *probability* (in some one of the many senses previously discussed). Omitting more than mention of those students (e.g., Wittgenstein and Schlick) who have dismissed such discussions as meaningless because, according to them, theories are not "genuine" propositions since they are not completely verifiable, reference must be made to another group of writers. According to this group, the traditional discussions have not fruitfully illuminated the character of scientific inquiry because those who take part in them neglect the *function* which theories have in inquiry. When this function is examined, it has been urged, it turns out that questions of the *truth* of theories (in the sense in which theories of truth have been traditionally discussed) are of little concern to those who actually use theories. Reflective inquiry is instituted for the sake of settling a *specific* problem, whether it be practical or theoretical, and inquiry terminates when a resolution of the problem is obtained. The various procedures distinguishable in inquiry (such as observation, operation upon subject matter including the manipulation of instruments, symbolic representation of properties of subject matter, symbolic transformation and calculation, etc.) are to be viewed as instrumental to its end product. The use of theories is one patent factor in reflective inquiry. They function primarily as means for effecting transitions from one set of statements to other sets, with the intent of controlling natural changes and of supplying predictions capable of being checked through manipulating directly experienceable subject matter. Accordingly, in their actual use in science, theories serve as *instruments* in specific contexts, and in this capacity are to be characterized as good or bad, effective or ineffective, rather than as true or false or probable. Those who stress the instrumental function of theories are not necessarily committed to identifying truth with effectiveness and falsity with uselessness. Their major insight does not consist in denying the meaningfulness of certain types of inquiries into the truth of theories but in calling attention to the way theories

Concluding Remarks

function and to the safeguards and conditions of their effectiveness. A theory is confirmed to the degree that it performs its specific instrumental function. From this point of view, which has been developed with much detail by Dewey, the degree of confirmation for a theory may be interpreted as a mark of its proved effectiveness as an intellectual tool for the purposes for which it has been instituted.

9. Concluding Remarks

In consonance with the discussion and terminology of the theory of signs in Volume I, Number 2, it is convenient to classify the problems connected with probability into three distinct though connected groups. *Syntactical problems*: these are concerned primarily with perfecting the calculus of probability, making more precise its assumptions, simplifying its procedure, establishing its consistency, developing alternative formal techniques, and indicating its relation to other branches of formal mathematics. Some of these matters were considered in Section II. *Semantical problems*: these are concerned with establishing and formulating appropriate rules for applying the calculus to various existential affairs, by indicating under what conditions certain complexes of signs in the calculus are to be co-ordinated with experimentally controllable situations. *Pragmatic problems*: these are concerned with formulating the procedures and conditions involved in the acceptance of probability statements, and with evaluating the efficacy of the calculus in solving the problems set for it in scientific inquiry. Semantical and pragmatic problems were outlined in the present section.

In recent years a growing number of mathematicians and logicians have devoted themselves to the solution of the syntactical problems of probability. Although there are still a number of outstanding difficulties, these are being attacked with the most subtle instruments of modern mathematics. In any case, the calculus has been refined and generalized to an extent undreamed of a century ago. The discussion of the semantical problems of probability is perhaps still in its infancy, though important spade work has already been done. The very

Principles of the Theory of Probability

recognition of the existence of such problems bodes well for the future, since the classical discussions of probability have been shown to be inadequate largely because semantical problems were not clearly distinguished from syntactical ones. The discussion of pragmatic problems has been carried on in the United States for many years. The most obvious fruits of this activity are the number of substantial contributions to an objective psychology dealing with scientific inquiry. And the present co-operative attack upon this group of problems by biologically oriented thinkers and those trained in the mathematical sciences gives a bright promise that, perhaps for the first time in the modern period, an adequate account of human behavior in the context of getting knowledge will soon be available.

The present section has stressed problems associated with the discussion of probability which are still largely unsettled. An unsettled situation in an intellectual discipline has often been seized upon by those hostile to free inquiry as an opportunity to cry out the "bankruptcy of science," to charge it with "confusion," to preach a wholesale skepticism with respect to its findings, and to invoke dogmatically "perennial truths" in the interest of private and institutionalized vested interests. However, unsettled situations in science usually mark important departures from traditional modes of analysis and are concomitants of active research; and the present state of probability discussions is typical of such situations. Disagreement among competent students certainly indicates that the last word upon the topic under discussion has not been said; but it may also indicate that a community of workers is co-operatively engaged in contributing to the solution of complicated issues. Such in fact is patently the case in current discussions of probability. Even where sharp disagreements occur, those engaged in the discussion have been drawing upon one another's insights, have been influencing one another to state their proposed solutions with greater precision, have been led to recognize alternative possibilities in solutions, and have consequently guarded themselves against a premature commitment to theses which may block the course of further inquiry. What is essen-

tial for the future development of probability considerations, as for the development of science in general, is that trained minds play upon its problems freely and that those engaged in discussing them illustrate in their own procedure the characteristic temper of scientific inquiry—to claim no infallibility and to exempt no proposed solution of a problem from intense criticism. Such a policy has borne precious fruit in the past, and it is reasonable to expect that it will continue to do so. In the history of the study of probability it has brought into existence a perfected calculus of probability; it has led to an extension of its range of application to many diverse domains; and it has contributed to showing that the various sciences, however distinct their specific subject matters may be, employ a common logic and common procedures, are faced with common logical problems, and are mutually indebted to one another for indispensable tools of inquiry.

NOTES

These very limited bibliographical notes aim to do no more than suggest further reading on some of the topics mentioned in the text.

1. For the classical view of probability consult Laplace, *Essai philosophique sur les probabilités* (Paris, 1814); A. De Morgan, *An Essay on Probability* (London, 1838). For the logical view see J. M. Keynes, *Treatise on Probability* (London, 1921); J. von Kries, *Die Principien der Wahrscheinlichkeitsrechnung* (Tübingen, 1886); F. Waismann, "Analyse des Wahrscheinlichkeitsbegriffs," *Erkenntnis*, Vol. I. For the frequency view see J. Venn, *Logio of Chances* (London, 1886); Charles S. Peirce, *Collected Papers*, Vol. II (Cambridge, Mass., 1932); R. von Mises, *Wahrscheinlichkeit, Statistik, und Wahrheit* (Vienna, 1936) and *Wahrscheinlichkeitsrechnung* (Leipzig, 1931); H. Reichenbach, *Wahrscheinlichkeitslehre* (Leiden, 1935).

2. For these alternative definitions consult K. Popper, *Logik der Forschung* (Vienna, 1935), and A. H. Copeland, "Admissible Numbers in the Theory of Probability," *American Journal of Mathematics*, Vol. L, as well as his "Predictions and Probabilities," *Erkenntnis*, Vol. VI.

3. Peirce's comments on these matters are scattered throughout his writings, especially in Vols. II and VI of his *Collected Papers*. Philipp Frank has written many monographs on this subject, but the fullest account will be found in *Das Kausalgesetz und seine Grenzen* (Vienna, 1932). Henry Margenau develops his point of view in several articles in *Philosophy of Science* and in R. B. Lindsay and H. Margenau, *Foundations of Physics* (New York, 1936). See also E. Cassirer, *Determinismus und Indeterminismus in der modernen Physik* (Göteborg, 1937).

4. Reichenbach's book cited in n. 1 contains fairly full references to these discussions. A. Wald's paper, "Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung," appeared in K. Menger's *Ergebnisse eines mathematischen Kolloquiums*, Heft 8.

Principles of the Theory of Probability

5. The technical details referred to will be found in any book on the mathematical theory of probability; for example, the books of von Mises cited in n. 1. Fisher's writings are scattered in many periodicals but are summarized in his *Statistical Methods for Research Workers* (Edinburgh and London, 1925), which also contains a list of his papers, and also in his *The Design of Experiments* (Edinburgh and London, 1935).

6. This criticism has been made by a number of writers, e.g., F. Cantelli, "Considération sur la convergence dans le calcul des probabilités," *Annales de l'Institut Henri Poincaré*, Vol. V; T. C. Fry, *Probability and Its Engineering Uses* (New York, 1928); and R. B. Lindsay and H. Margenau in their *Foundations of Physics*.

7. Consult A. Kolmogoroff, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik*, Vol. II, and the important series of works edited by E. Borel entitled *Traité du calcul des probabilités et de ses applications*.

8. These interpretations will be found in F. P. Ramsey, *Foundations of Mathematics* (London and New York, 1931); C. G. Hempel, "Über den Gehalt von Wahrscheinlichkeitsaussagen," *Erkenntnis*, Vol. V; H. Jeffreys, *Scientific Inference* (Cambridge, 1931); S. Mazurkiewicz, "Über die Grundlagen der Wahrscheinlichkeitsrechnung," *Monatshefte für Mathem. u. Physik*, Vol. XLI; and other works by authors already cited in previous notes.

9. Reichenbach's views have been stated by him in their most complete form in the work cited in n. 1 and in his *Experience and Prediction* (Chicago, 1938).

10. Consult G. Birkhoff and J. von Neumann, "The Logic of Quantum Mechanics," *Annals of Mathematics*, Vol. XXXVII; M. Strauss, "Zur Begründung der statistischen Transformationstheorie der Quantenmechanik," *Sitzungsber. der preuss. Akad. d. Wiss.* (1936); Paulette Février, "Les Relations d'incertitude de Heisenberg et la logique," *Comptes rendus des sciences*, Vol. CCIV.

11. For a recent exposition of the subjective view consult B. De Finetti, "La Prevision: ses lois logiques, ses sources subjectives," in *Annales de l'Institut Henri Poincaré*, Vol. VII. Criticisms of the classic view will be found in the writings of Peirce, Venn, von Kries, and Keynes already referred to.

12. In addition to the works cited in n. 1, consult J. Nicod, *Le Problème logique de l'induction* (Paris, 1923).

13. Poincaré's method is explained in his *Calcul des probabilités* (Paris, 1896) and also by Reichenbach in his *Wahrscheinlichkeitslehre*. For the work of E. Hopf see "On Causality, Statistics and Probability," *Journal of Mathematics and Physics*, Vol. XIII; see also G. D. Birkhoff and D. C. Lewis, "Stability in Causal Systems," *Philosophy of Science*, Vol. II.

14. Carnap's discussion is contained in his "Testability and Meaning," *Philosophy of Science*, Vols. III and IV.

15. For further discussion of these matters consult M. R. Cohen, *Reason and Nature* (New York, 1931); John Dewey, *Essays in Experimental Logic* (Chicago, 1916), *Quest for Certainty* (New York, 1929), and *Logic: The Theory of Inquiry* (New York, 1938); H. Feigl, "The Logical Character of the Principle of Induction," *Philosophy of Science*, Vol. I; Otto Neurath, "Pseudorationalismus der Falsifikation," *Erkenntnis*, Vol. V.

Selected Bibliography

- BIRKHOFF, G. D., and LEWIS, D. C. "Stability in Causal Systems," *Philosophy of Science*, Vol. II (1935).
- BIRKHOFF, G., and NEUMANN, J. VON. "The Logic of Quantum Mechanics," *Annals of Mathematics*, Vol. XXXVII (1936).
- BOREL, E. *Traité du calcul des probabilités et de ses applications*. Paris, 1925.
- CARNAP, R. "Testability and Meaning," *Philosophy of Science*, Vols. III (1936) and IV (1937).
- CASSIRER, E. *Determinismus und Indeterminismus in der moderne Physik*. Göteborg, 1937.
- COHEN, M. R. *Reason and Nature*. New York, 1931.
- COPELAND, A. H. "Admissible Numbers in the Theory of Probability," *American Journal of Mathematics*, Vol. L (1900).
- . "Predictions and Probabilities," *Erkenntnis*, Vol. VI (1936).
- DE FINETTI, B. "La Prevision: ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, Vol. VII (1937).
- DE MORGAN, A. *An Essay on Probability*. London, 1838.
- DEWEY, JOHN. *Essays in Experimental Logic*. Chicago, 1916.
- . *Logic: The Theory of Inquiry*. New York, 1938.
- FEIGL, H. "The Logical Character of the Principle of Induction," *Philosophy of Science*, Vol. I (1934).
- FISHER, R. A. *Statistical Methods for Research Workers*. Edinburgh and London, 1925.
- . *The Design of Experiments*. Edinburgh and London, 1935.
- FRANK, PHILIPP. *Das Kausalgesetz und seine Grenzen*. Vienna, 1932.
- FRY, T. C. *Probability and Its Engineering Uses*. New York, 1928.
- HEMPEL, C. G. "Über den Gehalt von Wahrscheinlichkeitsaussagen," *Erkenntnis*, Vol. V (1935).
- HOPF, E. "On Causality, Statistics and Probability," *Journal of Mathematics and Physics*, Vol. XIII (1934).
- JEFFREYS, H. *Scientific Inference*. Cambridge, 1931.
- KEYNES, J. M. *Treatise on Probability*. London, 1921.
- KOLMOGOROFF, A. "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik*, Vol. II (1933).
- KRIES, J. VON. *Die Principien der Wahrscheinlichkeitsrechnung*. Tübingen, 1886.
- LAPLACE, P. *Essai philosophique sur les probabilités*. Paris, 1814.
- LINDSAY, R. B., and MARGENAU, H. *Foundations of Physics*. New York, 1936.
- MISES, R. VON. *Wahrscheinlichkeit, Statistik und Wahrheit*. Vienna, 1928.
- . *Wahrscheinlichkeitsrechnung*. Leipzig, 1931.
- NEURATH, OTTO. "Pseudorationalismus der Falsifikation," *Erkenntnis*, Vol. V (1935).

Principles of the Theory of Probability

- NEYMAN, J. *Lectures and Conferences on Mathematical Statistics*. Washington, D.C., 1938.
- NICOD, J. *Le Probleme logique de l'induction*. Paris, 1918.
- PEIRCE, CHARLES S. *Collected Papers*, Vols. II and VI. Cambridge, Mass., 1932 and 1935.
- POPPER, K. *Logik der Forschung*. Vienna, 1935.
- RAMSEY, F. P. *Foundations of Mathematics*. London and New York, 1931.
- REICHENBACH, HANS. *Wahrscheinlichkeitslehre*. Leiden, 1935.
- . *Experience and Prediction*. Chicago, 1938.
- STRAUSS, M. "Zur Begründung der statistischen Transformationstheorie der Quantenmechanik," *Sitzungsber. der preuss. Akad. d. Wiss.* (1936).
- VENN, J. *Logic of Chance*. London, 1886.
- WAIMANN, F. "Analyse des Wahrscheinlichkeitsbegriffs," *Erkenntnis*, Vol. I. (1930).
- WALD, A. "Die Widerspruchsfreiheit des Kollektivsbegriffes der Wahrscheinlichkeitsrechnung," *Ergebnisse eines mathem. Kolloquiums*, Heft 8.

Foundations of the Unity of Science

Foundations of the Unity of Science, though a self-contained work, serves as the first two introductory volumes of the proposed *International Encyclopedia of Unified Science*. It is concerned with the scientific enterprise as a whole.

The wish to insure impartiality has led to a selection of collaborators with somewhat different points of view, but who agree in considering the unity of science as the ideal aim of their efforts, in eliminating any form of speculation other than that recognized in science, in stressing the importance of logical analysis in various fields, and in taking into account the historical development of scientific concepts and regulative principles. Such collaborators include, for instance, persons stemming from the Vienna Circle, from the Berlin group of scientific philosophers, from the Polish school of logicians, from the group centering around *Scientia* and the Centre de *synthèse*, as well as representatives of American pragmatism, the English analytical school, French conventionalism, various groups of scientific philosophers in Belgium, Holland, Switzerland, Scandinavia, and other countries, and a large number of scientists from the various special branches of science.

For these and other reasons there will be a certain divergence of opinion within the wider set of agreements which give unity to the work; tendencies which are often called scientific empiricism and logical empiricism will find a place by the side of other tendencies which prefer to be called scientific or experimental rationalism. Collaborators of various nationalities have been invited; only their personal competence has been considered or the benefits to be obtained from a variety of cultural viewpoints—their political views or the political ideologies of the countries they come from have not entered into consideration, since the *Encyclopedia* is a scientific and not a political enterprise. Each collaborator will, of course, be responsible only for the ideas which he himself expresses.

Foundations of the Unity of Science will contain the following twenty monographs (some variations of author and title may be made):

VOLUME I

1. *Encyclopedia and Unified Science*—NEURATH, BOHR, DEWEY, RUSSELL, CARNAP, MORRIS
2. *Foundations of the Theory of Signs*—CHARLES W. MORRIS (Chicago)
3. *Foundations of Logic and Mathematics*—RUDOLF CARNAP (Chicago)
4. *The Language of Science*—LEONARD BLOOMFIELD (Chicago)
5. *Procedures of Empirical Science*—VICTOR F. LENZEN (Berkeley)
6. *Principles of the Theory of Probability*—ERNEST NAGEL (New York City)
7. *Foundations of Physics*—PHILIPP FRANK (Prague)
9. *Cosmology*—E. FINLAY FREUNDLICH (Istanbul)
9. *Foundations of Biology*—FELIX MAINK (Prague)
10. *Formal Procedures in Biology*—JOSEPH H. WOODGER (London)

VOLUME II

1. *Theory of Behavior*—EGON BRUNSWIK (Vienna, Berkeley) and ARNE NESS (Oslo)
2. *Foundations of Social Science*—OTTO NEURATH (The Hague)
3. *Empirical Theory of Value*—JOHN DEWEY (New York City)
4. *Sociology of Science*—LOUIS WIRTH (Chicago)
5. *History of Science*—FEDERIGO ENRIQUES (Rome)
6. *History of Logic*—JAN LUKASIEWICZ (Warsaw)
7. *From Rationalism a Priori to Empiricism*—LOUIS ROUGIER (Besançon, Cairo)
8. *Problems of Empiricism and Rationalism*—EDGAR ZILSEL (Vienna) and GEORGE SANTILLANA (New York City)
9. *Logical Empiricism*—JOERGEN JOERGENSEN (Copenhagen)
10. *Bibliography and Index*—NEURATH, JOERGENSEN, CARL G. HEMPEL (Brussels)

It is planned to publish the monographs between June, 1938, and August, 1939. The price for the series is \$18.00. The monographs will be available singly at \$1.00.

The University of Chicago Press. Chicago, Illinois