U.S. AIR FORCE Project RAND

ON THE EPISTEMOLOGY OF THE INEXACT SCIENCES

Olaf Helmer and Nicholas Rescher

R-353

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

Copyright 1959 by
THE INSTITUTE OF MANAGEMENT SCIENCES

Copyright © 1960
THE RAND CORPORATION

PROJECT RAND

ON THE EPISTEMOLOGY OF THE INEXACT SCIENCES

Olaf Helmer and Nicholas Rescher

February, 1960

R-353

This research is sponsored by the United States Air Force under contract No. AF 49(368)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Development, Hq USAF.

The views, conclusions, and recommendations expressed herein do not necessarily reflect the official views or policies of the United States Air Force.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

SUMMARY

This report considers a new epistemological approach to the inexact sciences, which include applied physical sciences such as engineering or medicine as well as most of the social sciences. The purpose of all science is to explain past events and predict future ones, and to do so in an objective manner. While explanation and prediction have the same logical structure in the exact sciences, this is not so in the inexact sciences; and this fact opens the door to the development of specifically predictive instrumentalities in these fields, and thus to various methodological innovations. Among these are the systematic employment of expert judgment and the use of pseudo-experimentation, involving simulation processes and, in particular, operational gaming.

CONTENTS

Summary iii

ON THE EPISTEMOLOGY OF THE INEXACT SCIENCES

The Mythology of Exactness 1
Plan of Procedure 4
Historical Laws 5
Quasi-laws in the Physical Sciences 9
Explanation and Prediction 10
Probability 13
Some Examples of the Use of Evidence in Prediction 18
The Role of Expertise in Prediction 20
The Problem of the Predictive Use of Evidence in an Inexact Context 21
The Intrinsic Use of Experts for Prediction 23
The Role of Prediction as an Aid to Decisionmaking 24
Justification of the Intrinsic Use of Expertise 26
Criteria for the Selection of Predictive Experts 27
The Dependence of Predictive Performance on Subject Matter 29
Predictive Consensus Techniques 31
Simulation and Pseudo-experimentation 33
Operational Gaming 36
Review of the Main Theses 37
Some Tasks for Methodological Research 39

"The laws of economics are to be compared with the laws of the tides, rather than with the simple and exact law of gravitation. For the actions of men are so various and uncertain, that the best statement of tendencies, which we can make in a science of human conduct, must needs be inexact and faulty. This might be urged as a reason against making any statements at all on the subject; but that would be almost to abandon life. Life is human conduct, and the thoughts and emotions that grow up around it. By the fundamental impulses of our nature we all—high and low, learned and unlearned—are in our several degrees constantly striving to understand the courses of human action, and to shape them for our purposes, whether selfish or unselfish, whether noble or ignoble. And since we *must* form to ourselves some notions of the tendencies of human action, our choice is between forming those notions carelessly and forming them carefully. The harder the task, the greater the need for steady patient inquiry; for turning to account the experience, that has been reaped by the more advanced physical sciences; and for framing as best we can well thought-out estimates, or provisional laws, of the tendencies of human action."

—Alfred Marshall, *Principles of Economics*, 1892

ON THE EPISTEMOLOGY OF THE INEXACT SCIENCES

THE MYTHOLOGY OF EXACTNESS

It is a fiction of long standing that there are two classes of sciences, the exact and the inexact, and that the social sciences by and large are members of the second class—unless and until, like experimental psychology or some parts of economics, they mature to the point where admission to the first class may be granted.

This widely prevalent attitude seems to us fundamentally mistaken; for it finds a difference in principle where there is only one of degree, and it imputes to the so-called exact sciences a procedural rigor that is rarely present in fact. For the sake of a fuller discussion of these points, let us clarify at the very outset the terms "science," "exact science," and "inexact science," as they are intended here.

For an enterprise to be characterized as *scientific* it must have as its purpose the explanation and prediction of phenomena within its subject-matter domain and it must provide such explanation and prediction in a reasoned, and therefore intersubjective, fashion. We speak of an "exact science" if this reasoning process is formalized in the sense that the terms used are exactly defined and reasoning takes place by formal logico-mathematical derivation of the hypothesis (the statement of the fact to be explained or predicted) from the evidence (the body of knowledge accepted by virtue of being highly confirmed by observation). That an exact science frequently uses mathematical notation and concerns itself with attributes that lend themselves to exact measurement we regard as incidental rather than defining characteristics. The same point applies to the precision, or exactness, of the predictions of which the science may be capable. While precise predictions are indeed to be preferred to vague ones, a discipline that provides predictions of a less precise character, but makes them correctly and in a systematic and reasoned way, must be classified as a science.

In an inexact science, conversely, reasoning is informal; in particular, some of the terminology may, without actually impeding communication, exhibit some inherent vagueness, and reasoning may at least in part rely on reference to intuitively perceived facts or implications. Again, an inexact science rarely uses mathematical notation or employs attributes capable of exact measurement, and as a rule does not make its predictions with great precision and exactitude.

Using the terms as elucidated here—and we believe that this corresponds closely to accepted usage—purely descriptive surveys or summaries, such as the part of history that is mere chronology or, say, purely descriptive botany or geography, are not called "sciences." History proper, on the other hand, which seeks to explain historical transactions and to establish historical judgments having some degree of generality, is a science; it is in fact largely coincident with political science, except that its practitioners focus their interest on the past, while the political scientists' main concern is the present and the future.

As for *exactness*, this qualification, far from being attributable to all of the so-called natural sciences, applies only to a small section of them, in particular to certain subfields of physics, in some of which exactness has even been put to the ultimate test of formal axiomatization. In other branches of physics, such as parts of aerodynamics and of the physics of extreme temperatures, exact procedures are still intermingled with unformalized expertise. Indeed the latter becomes more dominant as we move away from the precise and usually highly abstract core of an exact discipline and towards its applications to the complexities of the real world. Both architecture and medicine are cases in point. Aside from the respective activities of building structures and healing people, both have a theoretical content—that is, they are explanatory and predictive ("This bridge will not collapse, or has not collapsed, because . . . . " "This patient will exhibit, or has exhibited, such and such symptoms because . . . ."). They must therefore properly be called "sciences," but they are largely inexact since they rely heavily on informal reasoning processes.

If in addition to these examples we remember the essentially in-between status of such fields as economics and psychology, both of which show abundant evidence of exact derivations as well as reliance on intuitive judgment (exhibiting intermittent use of mathematical symbolism and of

measurable attributes and an occasional ability to predict with precision), it should be obvious that there is at present no clear-cut dichotomy between exact and inexact sciences, and, in particular, that inexactness is not an attribute of only the social sciences.

However, leaving aside their present comparative status, it still might be possible to hold the view that there exists an epistemological difference *in principle* between the social sciences on the one hand and the natural or physical sciences on the other, in the sense that the latter, though not necessarily quite exact as yet, will gradually achieve ultimate exactness, while the former, due to the intangible nature of their subject-matter and the imperfection in principle of their observational data, must of necessity remain inexact. Such a view would be based on a false premise, namely, a wholly misguided application of the exactness-versus-inexactness distinction. Indeed, the artificial discrimination between the physical sciences with their (at least in principle) precise terms, exact derivations, and reliable predictions, as opposed to the social sciences with their vague terms, intuitive insights, and virtual unpredictability has retarded the development of the latter immeasurably.

The reason for this defeatist point of view regarding the social sciences may be traceable to a basic misunderstanding of the nature of scientific endeavor. What matters is not whether or to what extent inexactitudes in procedures and predictive capability can eventually be removed (with regard to predictive precision, the social sciences may perhaps always fall short of the physical sciences); rather it is *objectivity*—that is, the intersubjectivity of findings independent of any one person's intuitive judgment—that distinguishes science from intuitive guesswork, however brilliant. This has nothing to do with the intuitive spark that may be the origin of a new discovery; pure mathematics, whose formal exactness is beyond question, needs that as much as any science. But once a new fact or a new idea has been conjectured, no matter on how intuitive a foundation, it must be capable of objective test and confirmation by anyone. And it is this crucial standard of scientific objectivity, rather than any purported criterion of exactitude, to which the social sciences must conform.

In rejecting precision of form or method as well as degree of predictability as basic discriminants between the social and the physical sciences,

it thus remains to be seen whether there might not in fact be a fundamental epistemological difference between them with regard to their ability to live up to the same rigorous standard of objectivity. Our belief is that there is essentially no such difference, in other words, that the social sciences cannot be separated from the physical sciences on methodological grounds. We hope to convince the reader of the validity of our position by offering, in what follows, at least some indications as to how the foundations for a uniform epistemology of all of the inexact sciences might be laid—be they social sciences or "as yet" inexact physical sciences.

PLAN OF PROCEDURE

Our goal is more modest than that of presenting a comprehensive epistemology of the inexact sciences. We merely wish to outline an epistemological attitude toward them that we would like to see adopted more widely. Since epistemology is concerned with the role of evidence in the attainment of scientific laws and with the scientific procedures implied by that role, we need to re-examine the status of such things as laws, evidence, confirmation, prediction, and explanation, with special reference to the case of inexact sciences.

We shall begin with a brief look at historical laws, as typical examples of the explanatory generalizations encountered in the social sciences. They will typically be found to involve escape clauses of the "other things being equal" kind, which makes them different from the neat laws of the exact physical sciences; and the possibility of their application despite this defect will have to be discussed. In this context we shall examine the role of laws in prediction and in explanation and try to clarify the similarities and differences in the structure of prediction and explanation.

An analysis of this kind necessarily involves the prominent use of the notion of probability. In order to avoid misunderstandings, we shall sketch briefly our own views regarding this somewhat controversial concept. We prefer to distinguish—in addition to the familiar statistical concept of probability based on observed frequencies—an *objective* and a *subjective* notion of probability, the former called "degree of confirmation," the latter, "personal probability"; the two will be seen to be linked via the concepts of "rational person" and "fair bets."

We shall then use these considerations regarding probability to clarify some difficulties arising in the use of evidence for predictive purposes, and consider ways in which these difficulties may be surmounted by supplementing the objective evidence by judicious reliance on apparently subjective means (personal probabilities). This deviation from the standard procedures of the exact sciences will necessitate examining the possible justification of reliance on expert judgment and a search for improved and systematic methods for its use. Proposals for obtaining a consensus among a group of experts will be examined, as well as controlled opinion-feedback techniques that can aid their performance as a group. We next take up explicitly the question of whether, and if so to what extent, the standard procedures of the exact sciences have to be further augmented to meet the needs of the inexact sciences. Here we shall, in particular, discuss the concept of simulation models and their perhaps most important subcase, operational gaming models, together with the use of experts in conducting pseudo-experiments within such models. Finally, some suggestions will be made as to how the application of relatively unorthodox methods (such as the use of experts and gaming), tried successfully in other inexact sciences, might aid the applied social sciences, especially in exercising their predictive function with regard to decisionmaking processes.

HISTORICAL LAWS

Let us, then, first take a brief look at historical science in order to obtain some illustrative examples of the form of laws in the social sciences and of the function they perform. A *historical law* may be regarded as a well-confirmed statement concerning the actions of an organized group of men under certain restrictive conditions (such group actions being intended to include those of systems composed conjointly of men and nonhuman instrumentalities under their physical control). Examples of such laws are "A census takes place in the United States in every decade year"; "Heretics were persecuted in 17th century Spain"; and "In the sea fights of sailing vessels in the period 1653–1803, large formations were too cumbersome for effectual control as single units." Such statements share three features of epistemological importance and interest: they are

law-like, *spatio-temporally restricted*, and *loose*. These features require elaboration.

To consider law-likeness, let us take for example the statement about the cumbersomeness of large sailing fleets in sea fights. On first view, this statement might seem to be a mere descriptive list of characteristics of certain particular engagements: a shorthand version of a long conjunction of statements about large-scale engagements during the century and a half from Texel (1653) to Trafalgar (1803). This view is incorrect, however, because the statement in question is more than an assertion regarding characteristics of certain actual engagements. Unlike a mere description, it can serve to explain developments in cases to which it makes no reference. Furthermore, the statement has counterfactual force. It asserts that in literally any large-scale fleet action fought under the conditions in question (with sailing vessels of certain types, having particular modes of armament and using contemporaneous communications methods), effectual control of a great battle line is hopeless. It is claimed, for example, that had Villeneuve issued from Cadiz some days earlier or later he would all the same have encountered difficulty in the management of the great allied battle fleet of over thirty sail of the line, and Nelson's strategem of dividing his force into two virtually independent units under prearranged plans would have facilitated effective management equally well as at Trafalgar.

The statement in question is thus no mere descriptive summary of particular events; it functions on the more general plane of law-like statements, specifically in that it can serve as a basis for explanation and that it can exert counterfactual force. To be sure, the individual descriptive statements that are known and relevant do provide a part of the appropriate evidence for the historical generalization. But the content of the statement itself lies beyond the sphere of mere description and, in taking this wider role, historical laws become marked as genuinely law-like statements.

Nevertheless, such historical generalizations are not unrestricted or universal in the manner in which the laws of the physical sciences are: they are not valid for all times and places. A historical law is limited, either tacitly or expressly, to applicability within specific geographic and temporal bounds. Usually historical laws are formulated by explicit use

of proper names: names of places, of groups of persons, of periods of time, of customs or institutions, of systems of technology, of culture, or the like. The restriction of application in such cases is overt in the formulation of the law. Sometimes, however, historical laws are formulated as unconditionally universal statements. But in such cases the statement properly interpreted takes on a conditional form of such a kind that its applicability is *de facto* tightly restricted. If sailing ships and contemporary naval technology and ordnance *were* reinstated, the tactical principles developed from the time of Tromp and de Ruyter to that of Rodney and Nelson would prove valid guidance. But the applicability of these tactical principles depends on the fulfillment of conditions that cannot reasonably be expected to recur. (This is not true of certain experimental situations in physics—such as the Michelson-Morely experiment—which merely are rarely repeated.) Thus although explicit use of proper names might be avoided in the formulation of a historical law, the statement is restricted nonetheless by conditions effecting in an oblique manner the same delimitation of its applicability to specific geographic and temporal bounds.

Finally, an important characteristic of historical laws lies in their being "loose." It has been said already that historical laws are (explicitly or obliquely) conditional in their logical form. However, the nature of these conditions is such that they can often not be spelled out fully and completely. For instance, the statement about sailing-fleet tactics has (among others) an implicit or tacit condition relating to the state of naval ordnance in the 18th century. In elaborating such conditions, the historian delineates what is typical of the place and period. The full implications of such reference may be vast and inexhaustible; in our example, ordnance ramifies *via* metalworking technology into metallurgy, mining, etc. Thus the conditions that are operative in the formulation of a historical law may only be indicated in a general way and are not necessarily (indeed in most cases they cannot be expected to be) exhaustively articulated. This characteristic of such laws is here designated as *looseness*.

It is this looseness of its laws that typifies history as an inexact science in the sense in which we have used the term: in a domain where laws are not fully and precisely articulated there exists a limit to exactitude in terminology and reasoning. In such a sphere, mathematical precision must not be expected. To say this implies no pejorative intent whatever, for the

looseness of historical laws is clearly recognized as being due, not to slipshod formulation of otherwise precise facts, but to the fundamental complexities inherent in the conceptual apparatus of the domain.

A consequence of the looseness of historical laws is that they are not universal, but merely quasi-general, in that they admit exceptions. Since the conditions delimiting the area of application of the law are often not exhaustively articulated, a supposed violation of the law may be explicable by showing that a legitimate (but as yet unformulated) precondition of the law's applicability is not fulfilled in the case under consideration. The laws may be taken to contain a tacit caveat of the "usually" or "other things being equal" type. A historical law is thus not strictly universal in that it must be taken as applicable to all cases falling within the scope of its explicitly formulated conditions; rather it may be thought to formulate relationships that obtain generally, or better, *as a rule*.[1]

Such a "law" we will term a *quasi-law*. In order for a quasi-law to be valid, it is not necessary that there be no apparent exceptions; it is only necessary that if an apparent exception should occur, an adequate explanation be forthcoming, an explanation demonstrating the exceptional characteristic of the case at hand by establishing the violation of an appropriate (if hitherto unformulated) condition of the law's applicability.[2]

For example the historical law that in the pre-revolutionary French navy only persons of noble birth were commissioned is not without apparent exceptions, since in particular the regulation was waived in the case of the great Jean Bart, son of a humble fisherman, who attained great distinction in the naval service. We may legitimately speak here of an *apparent* exception; for instead of abandoning this universal law in view of the cited counterexample, it is more expedient to maintain the law but to interpret

[1] This point has been made by various writers on historical method. Charles Frankel, for example, puts it as follows in his lucid article on "Explanation and Interpretation in History": "It is frequently misleading to take statements such as 'Power corrupts, and absolute power corrupts absolutely,' when historians use them, as attempts to give an exact statement of a universal law. . . . But such remarks may be taken as statements of strategy, rules to which it is best to conform in the absence of very strong countervailing considerations." (*Philosophy of Science*, Vol. 24, 1957, p. 142.)

[2] In his book *The Analysis of Matter*, London, 1957, Bertrand Russell writes: "Our pre-scientific general beliefs are hardly ever without exceptions; in science, a law with exceptions can only be tolerated as a makeshift" (p. 191). But this is true only in the physical sciences, if at all. A far juster view is that enunciated in Alfred Marshall's classic *Principles of Economics* (1892), cited in our motto.

it as being endowed with certain tacit amendments which, fully spelled out, would cause the law to read somewhat as follows: "In the pre-revolutionary French navy, as a rule, only persons of noble birth were commissioned—that is, unless the regulation was explicitly waived or an oversight or fraud occurred or some other similarly exceptional condition obtained." While it may be objected that such a formulation is vague—and indeed it is—it cannot be said that the law is now so loose as to be vacuous; for the intuitive intent is clear, and its looseness is far from permitting the law's retention in the face of just *any* counterexample.[3] Specifically, if a reliable source brings to light one counterinstance for which there is no tenable explanation whatsoever to give it exempt status, a historian may still wish to retain the law in the definite expectation that some such explanation eventually will be forthcoming; but should he be confronted with a succession or series of unexplained exceptions to the law, he would no doubt soon feel compelled to abandon the law itself.

We thus have the indisputable fact that in a generally loose context, that of history being typical of the inexact sciences, it would be hopeless to try to erect a theoretical structure that is logically, perhaps even esthetically, on a plane with our idealistic image of an exact theory. Yet, if we consider the situation not from the standpoint of the wishful dreamer of neat and tidy theory construction, but from that of the pragmatist in pursuit of a better understanding of the world through reasoned methods of explanation and prediction, then we have good reason to take heart at the sight of even quasi-laws; and we should realize that the seemingly thin line between vagueness and vacuity is solid enough to permit the distinction of fact from fiction reasonably well in practical applications.

QUASI-LAWS IN THE PHYSICAL SCIENCES

We have chosen to illustrate the nature of limited generalizations (quasi-laws) by means of the graphic example of historical laws. Use of this example from a social-science context must not, however, be construed as implying that quasi-laws do not occur in the natural, indeed even in the physical, sciences. In many parts of modern physics, formalized

[3] Michael Scriven, in a paper shortly to be published, speaks of historical generalizations as having a "selective immunity to counterexamples."

theories based wholly on universal principles are (at least presently) unavailable, and use of limited generalizations is commonplace, particularly in applied physics and engineering.

Writers on the methodology of the physical sciences often bear in mind a somewhat antiquated and much idealized image of physics as a very complete and thoroughly exact discipline in which it is never necessary to rely on limited generalizations or expert opinion. But physical science today is very far from meeting this ideal. Indeed the laws of some branches of the social sciences are no less general than those of various branches of physics, such as the theory of turbulence phenomena, high-velocity aerodynamics, or extreme temperatures. Throughout applied physics in particular, when we move (say, in engineering applications) from the realm of idealized abstraction ("perfect" gases, "homogeneous" media) to the complexities of the real world, reliance on generalizations that are, in effect, quasi-laws becomes pronounced. (Engineering practice in general is based on "rules of thumb" to an extent undreamed of in current theories of scientific method.)

Thus no warrant whatever exists for using the presence of quasi-laws in the social sciences as validating a methodological separation between them and the physical sciences. A realistic assessment of physical-science methods shows that quasi-laws are operative here too, and importantly so.

With this in mind, let us now turn to a closer examination of the role played by laws—or quasi-laws—in explanation and prediction.

EXPLANATION AND PREDICTION

A somewhat simplified characterization of scientific explanation—but one that nonetheless has a wide range of applicability, particularly in the physical sciences—is that explanation consists in the *logical derivation* of the statement to be explained from a complex of factual statements and well-established general laws. One would, for example, explain the freezing of a lake by adducing (1) the fact that the temperature fell below 32°F and (2) the law that water freezes at 32°F. These statements, taken together, yield the statement to be explained deductively.[4]

[4] For a full discussion of this matter, see C. G. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science*, Vol. 15, 1948, pp. 135–175.

This deductive model of explanation, while adequate for many important types of explanations encountered in the sciences, cannot without at least some emendation be accepted as applying to all explanations. For one thing there are probabilistic explanations, which can be based on statistical (rather than strictly universal) laws. ("I did not win the Irish Sweepstakes because the chances were overwhelmingly against my doing so.") And then there are the quasi-laws occurring in the inexact sciences that because of their escape clauses cannot serve as the basis of strict *derivation* and yet can carry explanatory force. For example, the quasi-law quoted earlier surely explains—in the accepted sense of the word—why the French fleet that supported Washington's Yorktown campaign was commanded by a nobleman (namely, the Comte de Grasse).

The uncertainty of conclusions based on quasi-laws is not due to the same reason as that of conclusions based on statistical laws. A statistical law asserts the presence of a characteristic in a certain (presumably high) percentage of cases; a quasi-law asserts it in all cases for which an exceptional status (in an ill-defined but clearly understood sense) cannot be claimed. We note for the moment, however, that the schema of explanation when either type of nonuniversal law is involved is the same, and in fact identical with what it would be were the law universal; and an explanation is regarded as satisfactory if, while short of logically *entailing* the hypothesis, it succeeds in making the statement to be explained highly *credible* in the sense of providing convincing evidence for it. We shall return below to a discussion of the concept of evidence.

With regard to prediction as opposed to explanation, analyses of scientific reasoning often emphasize the similarities between the two, holding that they are identical from a logical standpoint, inasmuch as each is an instance of the use of evidence to establish a hypothesis, and the major point of difference between them is held to be that the hypothesis of a prediction or of an explanation concerns respectively the future or the past. This view, however, does not do justice to several differences between prediction and explanation that are of particular importance.[5]

---

[5] On the contrast between prediction and explanation see further I. Scheffler, "Explanation, Prediction, and Abstraction," *British Journal for the Philosophy of Science*, Vol. 7, 1957, pp. 293–309, and N. Rescher, "On Prediction and Explanation," *British Journal for the Philosophy of Science*, Vol. 8, 1958, pp. 281–290.

First of all, there are such things as *unreasoned* predictions—predictions made without any articulation of justifying argument. The validation of such predictions lies not in their being supported by plausible arguments, but may, for example, reside in proving sound *ex post facto* through a record of successes on the part of the predictor or predicting mechanism.

It is clear that such predictions have no analogue in explanations; only reasoned predictions, based on the application of established theoretical principles, are akin to explanations. However, even here there is an important point of difference.

By the very meaning of the term, an explanation must *establish* its conclusion, showing that there is a strong warrant why the fact to be explained—rather than some possible alternative—obtains. On the other hand, the conclusion of a (reasoned) prediction need not be well established in this sense; it suffices that it be rendered *more tenable than comparable alternatives*. Here then is an important distinction in logical strength between explanations and predictions: An explanation, though it need not logically rule out alternatives altogether, must beyond reasonable doubt establish its hypothesis as *more credible than its negation*. Of a prediction, on the other hand, we need to require only that it establish its hypothesis as *more credible than any comparable alternative*. Of course predictions may, as in astronomy, be as firmly based in fact and as tightly articulated in reasoning as any explanation. But this is not a general requirement to which predictions *must* conform. A doctor's prognosis, for example, does not have astronomical certitude, yet practical considerations render it immensely useful as a guide in our conduct because it is far superior to reliance on guesswork or on pure chance alone as a basis for decisionmaking.

Generally speaking, in any field in which our ability to forecast with precision is very limited, our actions of necessity are guided by only slight differences in the probability that we attach to possible future alternative states of the world; consequently we must permit predictions to be based on far weaker evidence than explanations. This is especially true of a science such as history, or rather its predictive counterpart — political science. Here, in the absence of powerful theoretic delimitations that narrow down the immense variety of future possibilities to some manage-

able handful, the *a priori* likelihood of any particular state of affairs is insignificant, and we can thus tolerate considerable weakness in our predictive tools without rendering them useless. Consider, for example, the quasi-law that in a U.S. off-year election the opposition party is likely to gain. This is certainly not a general law, nor is it intended to be a summary of statistics. It has implicit qualifications of the "other things being equal" type, but it does claim to characterize the course of events "as a rule" and it generates an expectation of the explainability of deviations. On this basis, a historical (or political) law of this kind can provide a valid, though limited, foundation for sound predictions.

The epistemological asymmetry between explanation and prediction has not, it seems to us, been adequately recognized and taken into account in discussions of scientific method. For one thing, such recognition would lead to a better understanding of the promise of possibly unorthodox items of methodological equipment, such as quasi-laws, for the purposes of prediction in the inexact sciences. But more generally it would open the way to explicit consideration of a *specific methodology of prediction* —a matter that seems to have been neglected to date by the philosophers of science. As long as one believes that explanation and prediction are strict methodological counterparts, it is reasonable to press further with solely the explanatory problems of a discipline, in the expectation that only the tools thus forged will then be usable for predictive purposes. But once this belief is rejected, the problem of a specifically predictive method arises, and it becomes pertinent to investigate the possibilities of predictive procedures autonomous of those used for explanation.

Before discussing such possibilities in greater detail, it is imperative, to avoid various misunderstandings, that we clarify briefly the meaning of probability and of some associated concepts.

PROBABILITY

From the viewpoint of the philosophy of science, the theory of probability occupies a peculiar position. To the extent that it deals with relations among propositions it is part of semantics and thus of pure logic. To the extent that it deals with credibility, rational beliefs, and personal expectations, it is part of empirical pragmatics and thus a social science.

(The view, not held by us, that probability theory properly belongs entirely in the second field rather than the first is sometimes referred to as *psychologism*.)[6] Even for the logical part of the theory, the foundations are not yet established very firmly, and only in applications to the simplest forms of one-place predicate languages has real progress been made to date.[7] For this reason, some vagueness must still be accepted even in discussing the purely logical aspects of probability—that is, unless we are content to confine ourselves to the simplest case just mentioned, which we are not because the linguistic demands of the inexact sciences transcend these limits of simplicity even more frequently than do those of the exact sciences.

It is convenient to distinguish three probability concepts, namely, *relative frequency*, *degree of confirmation*, and *personal* (or *subjective*) *probability*. Of these, the first is an objective, empirically ascertainable property of classes of physical objects or physical events; the second is also purely objective, namely, a logical relation between sentences; the third is a measure of a person's confidence that some given statement is true and is thus an essentially subjective matter. Let us briefly consider each of these three probability concepts.

Relative Frequency

Relative frequency requires the statement of a reference class (of objects or events), also called the "population." If the class is finite, the relative frequency is simply the ratio of the number of elements having some property or trait, divided by the total number of elements in the class. Thus we speak of the relative frequency of males in the present U.S. population, or of rainy days in Los Angeles in the first half of this century.[8] Sometimes the notion of relative frequency is extended to classes of either indefinite or infinite size. For example, we may speak of the relative frequency of male births in the United States over an extended

[6] An incisive critique of psychologism is given in Chap. 2 of R. Carnap's book *Logical Foundations of Probability*, Chicago, 1950.

[7] See R. Carnap's massive study of the *Logical Foundations of Probability*, Chicago, 1950, and various studies cited by him in the extensive Bibliography, in particular those of Helmer, Hempel, and Oppenheim.

[8] It is a technical refinement, into which we need not enter here, that in applications it is common to use, instead of the relative frequency proper, some statistical *estimate* thereof.

period, without precisely specifying that period; or we may speak of the relative frequency "in the long run" of "heads" in tosses with a particular coin, where the sequence of tosses is of indefinite length and may even be idealized into an infinite sequence (in which case the "relative frequency" is the limit of the relative frequencies of the finite sub-sequences). In a situation like this, it is even customary to ascribe this probability, that is, the relative frequency of "heads" in the long run, as a property of the coin itself (in particular, a "fair coin" is one for which this probability is one-half). But it is best to interpret such a statement merely as a paraphrase for the longer statement that in a long sequence of possible tosses with this coin (but not so long as to alter the physical characteristics of the coin) the relative frequency of "heads" will be one-half.

Degree of Confirmation

The degree of confirmation is a logical relation between two sentences, the hypothesis H and the evidence E. The degree of confirmation of H on the basis of E is intended to be a measure of the credibility rationally imparted to the truth of H by the assumed truth of E. Precise definitions have thus far been suggested only for the one-place predicate calculus. In the simplest case, where E has the form of a statistical record of n observations, to the effect that exactly m out of n objects examined had a property P, and where the hypothesis H ascribes this property P to an as-yet unexamined object, the degree of confirmation of H on the basis of E, or $dc(H,E)$, is defined to be either the observed relative frequency m/n, or else a quantity very close to it (and having the same limit as n becomes large), which may differ somewhat from m/n because of technical requirements of elegance of the formalism. It is irrelevant for our present purposes which particular definition we adopt, but to fix the idea let us assume simply that in the above case $dc(H,E) = m/n$.

If E does not have the simple form of a statistic or H does not just affirm another like instance, then some plausible extension of the definition of "dc" is required; this may lead to cases where no single number can reasonably be specified but where the evidence merely warrants a narrowing down of the probability of H to several possible numbers or an interval of numbers. For instance, if H is the hypothesis that a certain

Irish plumber will vote Democratic in the next presidential election, and the evidence E amounts solely to saying that 70 per cent of the Irish vote Democratic and 20 per cent of the plumbers do, then all that can reasonably be asserted is that the required probability lies somewhere between 0.2 and 0.7.

Ambiguities of this kind can, of course, be removed by fiat (and in fact this has been the path followed in the formalisms proposed to date by Carnap). That is to say, one can transfer the ambiguity from the object language to the meta-language, by stating the matter as follows: There are several ways in which "dc" can be defined, but under each particular definition the degree of confirmation is a single-valued function.

No matter which of these alternatives is chosen, the situation can still be resolved, as long as we are dealing with one-place predicates only. When we move into a subject matter where adequate discourse requires multiplace predicates of several logical levels, no formal proposals for an extended definition of "dc" are as yet available, and we have to rely largely on trained intuition as to how a numerical measure of the "credibility rationally imparted to H by E" should be estimated in specific cases.

Since it is not the purpose of this report to deal at length with the foundations of probability theory, while on the other hand the use of some notion of degree of confirmation in the vague sense introduced here seems to us unavoidable, we shall largely have to ignore the technical problems pointed out above. For practical purposes this does not mean that we shall maintain the fiction of there being a well-defined formula available that permits computation of $dc(H, E)$ for all H and E. Rather we shall assume, in specific cases arising in situations of interest, that reasonable and knowledgeable persons, when confronted with the question of ascertaining a value of $dc(H, E)$, will find this definitely, if vaguely, meaningful and will arrive at estimates of the value that will not be too widely disparate. This leads us to the next probabilistic concept that must be discussed.

Personal Probability

Personal, or subjective, probability is a measure of a person's confidence in, or subjective conviction of, the truth of some hypothesis. According to

Savage,[9] it is measured behavioristically in terms of the person's betting behavior. If a person thinks that H is just about as likely as its negation $\sim H$, then, if he were placed in a situation where he had to make an even bet on either H or $\sim H$, he would presumably be indifferent to this choice. Similarly, if he thought H to be twice as likely as $\sim H$, he would have no preference as to which side to take in a 1:2 bet on $H:\sim H$. Generalizing this idea, we shall say that the person attaches the personal probability p to the hypothesis H if he is found to be indifferent between the choice of receiving, say, one dollar if H turns out to be true or receiving $p/(1-p)$ dollars if H turns out to be false (his "personal expectation" in either case being p dollars).[10]

We shall call a person "rational" if (1) his preferences (especially with regard to betting options) are mutually consistent or at least, when inconsistencies are brought to his attention, he is willing to correct them; (2) his personal probabilities are reasonably stable over time, provided he receives no new relevant evidence; (3) his personal probabilities are affected (in the right direction) by new relevant evidence; and (4) in simple cases where the evidence E at his disposal is known, and E and H are such that $dc(H, E)$ is defined, his personal probability regarding H is in reasonable agreement with the latter; in particular, he is indifferent as to which side to take in a bet that to his knowledge is a "fair" bet.

A (predictive) "expert" in some subject matter is a person who is *rational* in the sense discussed, who has a large background knowledge E in that field, and whose predictions (actual or implicit in his personal probabilities) with regard to hypotheses H in that field show a record of comparative successes in the long run. This is very much a relative concept, as it depends on the predictive performance of which the average nonexpert in the field would be capable. (In a temperate climate, a lay predictor can establish an excellent record by always forecasting good weather, but this would not support a claim to meteorological expertise.) Below we will give more detailed attention to predictive expertise.

---

[9] L. J. Savage, *The Foundations of Statistics*, New York, 1952.

[10] There are certain well-known difficulties connected with this behavioristic approach, which we will ignore here. We will merely mention that in experimental situations designed to elicit personal probabilities, care must be taken that the stakes involved are in a range where the utility of money is effectively linear and the utility (or disutility) of gambling is negligible.

With regard to the relationship between degree of confirmation and personal probability, it may be said that $dc(H, E)$ is intended to be a conceptual reconstruction of the personal probability that an entirely rational person would assign to H, given that his entire relevant information is E. In practice this relation can be applied in both directions: In simple cases where we have a generally acceptable definition of "dc" we may judge a person's rationality by the conformity of his personal probabilities—or of his betting behavior—with computable (or, if his information E is uncertain, estimable) dc values. Conversely, once a person has been established as rational and possibly even as an expert in a field, we may use his personal probabilities as estimates, on our part, of the degrees of confirmation that should be assigned to given hypotheses.

We shall make use of these probability concepts below, primarily in connection with the use of expert judgment for predictive purposes. But we must first consider the use of evidence in prediction, beginning with some examples to illustrate the problems arising in the predictive use of probabilistic evidence.

SOME EXAMPLES OF THE USE OF EVIDENCE IN PREDICTION

The simplest use of evidence occurs when there is a direct reference to prior instances. Will my car start on this cold morning? Its record of successful starts on previous cold mornings is about 50 per cent. I would be unduly hopeful or pessimistic in assigning as personal probability of its starting today a number significantly different from one-half. This use of a record of past instances as a basis for probability assignments with regard to future events is a common, and generally justified, inductive procedure (and of course is the basis on which a definition of degree of confirmation is constructed). However, under some circumstances it is a very poor way indeed of marshaling evidence.

Consider the case of Smith, who has been riding the bus to work for a year for a fare of 10 cents. One morning he is required to pay 15 cents. Smith may wonder if his return fare that evening will be 10 cents; but it is highly unlikely—despite the great preponderance of 10-cent rides in Smith's sample. For Smith well knows that public transportation fares

do change, and not by whim but by adoption of a new fare structure. In the light of this item of *background information*, it is unreasonable for Smith to base his personal probability directly on the cumulative record of past instances.

This illustrates the need for the use of background knowledge as indirect evidence, in the sense of furnishing other than direct-instance confirmation. This need is encountered constantly in the use of evidence, and it constitutes one of the prime obstacles to a more sophisticated definition of degree of confirmation than has hitherto been achieved. Consider another example. Will my new neighbor move away again within 5 years? He is a carpenter (the average carpenter moves once every 10 years) and a bachelor (the average bachelor moves once every 3 years). I can assess the likelihood of my neighbor's moving within the next 5 years relative to either the reference class of carpenters or that of bachelors. Which one I should choose, or what weight I should give to each, must depend strongly on my background information as to the relative relevance of occupation versus marital status as a determining factor in changes of domicile.

Such reference-class problems arise even with statistical information of the simplest kind. Consider a sample of 100 objects drawn at random from a population, with the following outcome as regards possession of the properties P and Q:

	has Q	has not -Q
has P	1	9
has not -P	89	1

Given this information, what is the probability that another object drawn from the population, which is known to have the property P, will also have the property Q? Should we use a value around 0.1 (since only 1 of 10 observed P's is a Q) or a value around 0.9 (since altogether 90 per cent of the observed sample has the property Q)? Here again, an expedient use of the statistical evidence before us must rely on background information, if any, regarding the relevance of P-ness to Q-ness. If we know that most Texans are rich and most barbers poor, and are given as the only specific item of information about a man by the name of

Jones that he is a Texan barber, we would do well to assign a low probability to the statement that Jones is rich, precisely because occupation is known to us to be more relevant to financial status than is location.

THE ROLE OF EXPERTISE IN PREDICTION

The implication of the examples we have been discussing is that a knowledge about past instances or about statistical samples—while indeed providing valuable information—is not the sole, and sometimes not even the main, form of evidence in support of rational assignments of probability values. In fact the evidential use of such *prima facie* evidence must be tempered by reference to background information, which frequently may be intuitive in character and have the form of a vague recognition of underlying regularities, such as analogies, correlations, or other conformities, the formal rendering of which would require the use of predicates of a logical level higher than the first.

The consideration of such underlying regularities is of special importance for the inexact sciences, particularly (but not exclusively) the social sciences,[11] because in this sphere we are constantly faced with situations in which statistical information matters less than knowledge of regularities in the behavior of people or in the character of institutions, such as traditions and customary practices, fashions and mores, national attitudes and climates of opinion, institutional rules and regulations, group aspirations, and so on. For instance, in assessing the chances of a Republican presidential victory in 1964, a knowledge of the record of past election successes matters less than an insight into current trends and tendencies; or in answering a question as to the likelihood, say, of U.S. recognition of Communist China by 1964, it is hard to point to any relevant statistical evidence, yet there exists a mass of relatively undigested but highly relevant background information.

This nonexplicitness of background knowledge, which nonetheless may be significant or even predominantly important, is typical of the inexact sciences, as is the uncertainty as to the evidential weight to be

[11] Use of background information to temper the application of statistical information is just as operative in the physical sciences—in engineering, for example. Therefore no difference in principle is involved here.

accorded various pieces of *prima facie* information in view of indirect evidence provided by underlying regularities. Hence the great importance that must be attached to experts and to expertise in these fields. For the expert has at his ready disposal a large store of (mostly inarticulated) background knowledge and a refined sensitivity to its relevance, through the intuitive application of which he is often able to produce trustworthy personal probabilities regarding hypotheses in his area of expertness.

The important place of expert judgment for predictions in the inexact sciences is further indicated by the prominence of quasi-laws among the explanatory instrumentalities of this domain. Since the conditions of applicability of such generalizations are neither fully nor even explicitly formulable, their use in specific circumstances presupposes the exercise of sound judgment as to their applicability to the case at hand. The informed expert, with his resources of background knowledge and his cultivated sense of the relevance and bearing of generalities in particular cases, is best able to carry out the application of quasi-laws necessary for reasoned prediction in this field.

THE PROBLEM OF THE PREDICTIVE USE OF EVIDENCE IN AN INEXACT CONTEXT

In summary, the foregoing illustrations of the predictive use of evidence may be said to indicate that we are frequently confronted with what must be considered as a problematical, and far from ideal, epistemological situation. The examples we have been considering show that in assessing the probability of a hypothesis H—typically a description of some future event—we are in many instances required to rely not merely on some specific and explicit evidence E, but also on a vast body of potentially relevant background knowledge K, which is in general not only vague in its extent (and therefore indefinite in content) but also deficient in explicit articulation. In many practical applications, particularly in the inexact sciences, not even that part of K that is suitably relevant to H can be assumed to be explicitly articulated, or even articulable. One is unable to set down in sentential form everything that would have to be included in a full characterization of one's knowledge about a familiar room; and the same applies equally, if not more so, to a

political expert's attempt to state all he knows that might be relevant to a question such as that of U.S. recognition of Communist China.

These considerations point up a deficiency for present purposes in the usual degree-of-confirmation concept quite apart from those already mentioned. For such an indefinite K, we cannot expect $dc(H, E \,\&\, K)$ to be determinable or even defined. This suggests, as a first step, the desirability of introducing a concept $dc_K(H, E)$—the "degree of confirmation of H on E in view of K"—which is defined to be equal to $dc(H, E \,\&\, K)$ whenever it is possible to articulate K fully within the same language in which H and E are stated. But how is such a quantity to be determined when K is not fully formulated? Furthermore, in addition to the difficulty involved in formulating it completely, K almost invariably contains probability statements (both of an objective, or dc, type, and of the indirect form "So-and-so attaches to H the personal probability p"). To date, there is no hint of any suggestion as to how "$dc(H, X)$" might be formally defined when X contains incompletely formulated matter.

Faced with this situation—which is surely not likely to be resolved in the near future—we must either for the present renounce all claims to systematized prediction in the inexact sciences, or, as indicated earlier, turn to unorthodox methods that are based on judicious and systematic reliance on expert judgment. One such course, to which we previously alluded, may possibly help us out of the present perplexity. Let A be an expert and $K(A)$ his relevant background knowledge. Then A's personal probability, $pp_A(H, E)$, may be taken as an estimate on *our* part of $dc_{K(A)}(H, E)$. Thus the device of using the personal probabilities of experts, extracted by appropriately devised techniques of interrogation, can serve as a means for measuring quantities of the dc type even in cases where there is no hope of applying the formal degree-of-confirmation concepts.

It might seem that in resorting to this device we conjure up a host of new problems, because—to all appearances—we are throwing objectivity to the winds. Of course, since we insist on remaining within our own definition of scientific activity, we do not propose to forego objectivity. However, before attempting to analyze the possibility of salvaging objectivity in this situation, it may be well to look at a few examples illustrating the application of expertise in the sense just described.

THE INTRINSIC USE OF EXPERTS FOR PREDICTION

A source of characteristic examples of the predictive use of expert judgment is provided by the field of diagnostics, especially medical diagnostics.[12] A patient, let us assume, exhibits a pattern of symptoms such that it is virtually certain that he has either ailment A or ailment B, with respective probabilities of 0.4 and 0.6, where these probabilities derive from the statistical record of past cases. Thus the entire body of explicit symptomatic evidence is (by hypothesis) such as to indicate a margin in favor of the prediction that the patient suffers from disease B rather than A, and thus may respond positively to a corresponding course of treatment. But it is quite possible that an examining physician, taking into consideration not only the explicit indicators that constitute the symptoms (temperature and blood pressure, for example) but also an entire host of otherwise inarticulated background knowledge with regard to this particular patient (such as the circumstances of the case) may arrive at a diagnosis of disease A rather than B. Thus the use of background information, in a way that is not systematized but depends entirely on the exercise of informal expert judgment, may appropriately lead to predictive conclusions in the face of *prima facie* evidence that points in the opposite direction.

Quite similar in its conceptual structure to the foregoing medical example are various other cases of predictive expertise in the economic sphere. The advice of an expert investment counselor, for example, may exhibit essentially the same subtle employment of nonarticulate background knowledge that characterized the prediction of the diagnostician.

Again, in such essentially sociological predictions of public reactions as are involved in the advertising and marketing of commercial products, the same predictive role of expert judgment comes into play. When the production of a motion picture is completed, a decision must be made regarding the number of prints to be made. There are economic reasons for an accurate prediction of the need: if too few prints are ready to meet the immediate demand, film rental income will be lost; on the other hand, the prints are costly, and an oversupply leads to considerable excess

[12] An extensive and useful discussion of medical prediction is contained in P. E. Mehl's book on *Clinical vs. Statistical Prediction*, Minneapolis, 1954.

expenditure. Here again, as in the medical or economic examples, certain limited predictions can be based wholly on the record of past statistics in analogous instances. The presence of certain actors in the cast, the topic, the theme and setting of the film, perhaps even its reception by preview audiences, may suggest a probability distribution for its demand. However, the major studios involved in motion-picture production are not content to rely on these explicit indicators alone. They are aware of the potential influence of a host of subtle intangibles such as so-called audience appeal, of the dependence on competitive offerings, and of other factors that are difficult to predict such as timeliness with reference to current events—all of which are susceptible of explicit statistical treatment only with the greatest difficulty, if at all. Therefore they prudently rely on the forecasts of professional experts in the field, who have exhibited a demonstrated ability to supplement the various explicit elements by appropriate use of their capacities for an intuitive appraisal of the many intangible factors that critically affect the final outcome.

Other examples drawn from the applied sciences, engineering, industry, politics, etc., will easily suggest themselves. What they have in common is the reliance, in part or wholly, on an expert, who here functions in an intrinsic rather than an extrinsic role. By extrinsic expertise we mean the kind of inventiveness, based on factual knowledge and the perception of previously unnoticed relationships, that goes into the hypothesizing of new laws and the construction of new theories; it is, in other words, the successful activity of the scientist qua scientist. Intrinsic expertise, by contrast, is not invoked until after a hypothesis has been formulated and its probability, in the sense of degree of confirmation, is to be estimated. The expert, when performing intrinsically, thus functions within a theory rather than on the theory-constructing level.

THE ROLE OF PREDICTION AS AN AID TO DECISIONMAKING

The decisions that professional decisionmakers—governmental administrators, company presidents, military commanders, etc.—are called on to make inevitably turn on the question of future developments, since their directives as to present actions are invariably conceived with a view

to future results. Thus a reliance on predictive ability is nowhere more overt and more pronounced than in the area of policy formation, and decisionmaking in general.

For this reason, decisionmakers surround themselves with staffs of expert advisers, whose special knowledge and expertise must generally cover a wide field. Some advising experts may have a great store of factual knowledge, and can thus serve as walking reference books. Others may excel through their diagnostic or otherwise predictive abilities. Still others, such as operations analysts and management consultants, may have a special analytical capacity to recognize the structure of the problems at hand, thus aiding in the proper use of the contributions of the other two types of experts. The availability of such special expertise constitutes for the decisionmaker a promise of increased predictive ability essential to the more effective discharge of his own responsibilities. Thus the ultimate function of expert advice is almost always to make a predictive contribution.

While the dependence of the decisionmakers on expert advisers is particularly pronounced in social-science contexts—in the formulation of economic and political policies, for example—such dependence on expertise ought by no means to be taken to contradistinguish the social from the physical sciences. In certain engineering applications, particularly of relatively underdeveloped branches of physics (such as the applied physics of extremes of temperature or velocity) the reliance on know-how and expert judgment is just as pronounced as it is in the applications of political science to foreign-policy formation. The use of experts for prediction does *not* constitute a line of demarcation between the social and the physical sciences, but rather between the exact and the inexact sciences. And, as we have already said, certain areas of applied physics are in the present state of our knowledge just as inexact as much of the social sciences, if indeed not more so.

Although we have held that the primary service of expert advisers to decisionmakers is in functioning as "predictors," we by no means intend to suggest that they act as fortunetellers, trying to foresee specific occurrences for which the limited intellectual vision of the nonexpert is insufficient. For the decision-supporting uses of predictive expertise, there is in general no necessity for an anticipation of particular future

occurrences. It suffices that the expert be able to sketch out adequately the general directions of future developments, to anticipate—as we have already suggested— some of the major critical junctures ("branch points") on which the course of these developments will hinge, and to make contingency predictions with regard to the alternatives associated with them.

While the value of scientific prediction for sound decisionmaking is beyond question, it can hardly be claimed that the inexact sciences have the situation regarding the use of predictive expertise well in hand. Quite to the contrary, it is our strong feeling that significant improvements are possible in the predictive instruments available to the decisionmaker. These improvements are contingent on the development of methods for the more effective predictive use of expert judgment. In our final section we shall give consideration to some of the problems involved in this highly important, but hitherto largely unexplored, area.

JUSTIFICATION OF THE INTRINSIC USE OF EXPERTISE

We come back to the problem of preserving objectivity in the face of reliance on expertise. Can we accept the use of intrinsic expert judgment within the framework of an inductive procedure without laying ourselves open to the charge of abandoning objective scientific methods and substituting rank subjectivity?

To see that explicit use of expert judgment is not incompatible with scientific objectivity, let us look once more at the medical-diagnosis example of the preceding section. Consider the situation in which a diagnostician has advised that a patient be treated for ailment A (involving, say, a major surgical operation) rather than B (which might merely call for a special diet). Our willingness, in this case, to put our trust in the expert's judgment surely would not be condemned as an overly subjective attitude. The reasons why our reliance on the expert is objectively justified are not difficult to see. For one thing, the selection of appropriate experts is not a matter of mere personal preference but is a procedure governed by objective criteria (about which more will be said in the ensuing section). But most importantly, the past diagnostic performance record makes the diagnostician an objectively reliable indicator (of diseases), in the same sense that one of any two highly correlated

physical characteristics is an indicator of the other. ("If most hot pieces of iron are red, and vice versa, and if this piece of iron is red, then this piece is probably hot.")

Even if the expert's explicit record of past performance is unknown, reliance on his predictions may be objectively justified on the basis of general background knowledge concerning his reputation as an expert. The objective reliability of experts' pronouncements may also be strongly suggested by the fact that they often exhibit a high degree of agreement with one another, which—at least if we have reason to assume the pronouncements to be independent—precludes subjective whim.

Epistemologically speaking, the use of an expert as an objective indicator—as illustrated by the example of the diagnostician—amounts to considering the expert's predictive pronouncement as an integral, intrinsic part of the subject matter, and treating his reliability as a part of the theory about the subject matter. Our information about the expert is conjoined to our other knowledge about the field, and we proceed with the application of precisely the same inductive methods that we would apply in cases where no use of expertise is made. Our "data" are supplemented by the expert's personal probability valuations and by his judgments of relevance (which, by the way, could be derived from suitable personal probability statements), and our "theory" is supplemented by information regarding the performance of experts.

In this manner the incorporation of expert judgment into the structure of our investigation is made subject to the *same* safeguards that are used to assure objectivity in other scientific investigations. The use of expertise is therefore no retreat from objectivity or reversion to a reliance on subjective taste.

CRITERIA FOR THE SELECTION OF PREDICTIVE EXPERTS

The first and most obvious criterion of expertise is of course knowledge. We resort to an "expert" precisely because we expect his information and the body of experience at his disposal to constitute an assurance that he will be able to select the needed items of background information, determine the character and extent of their relevance, and apply these insights to the formulation of the required personal probability judgments.

However, the expert's knowledge is not enough; he must be able to bring it to bear effectively on the predictive problem at hand, and this not every expert is able to do. It becomes necessary also to place some check on his predictive efficacy and to take a critical look at his past record of predictive performance.

The simplest way to score an expert's performance is in terms of "reliability": his *degree of reliability* is the relative frequency of cases in which, when confronted with several alternative hypotheses, he ascribed to the eventually correct alternative among them a greater personal probability than to the others.

This measure, while useful, must yet be taken with a grain of salt, for there are circumstances in which even a layman's degree of reliability, as defined above, can be very close to 1. For instance, in a region of very constant weather, a layman can prognosticate the weather quite successfully by always predicting the same weather for the next day as for the current one. Similarly, a quack who hands out bread pills and reassures his patients of recovery "in due time" may prove right more often than not and yet have no legitimate claim to being classified as a medical expert. Thus what matters is not so much an expert's absolute degree of reliability but his relative degree of reliability, that is, his reliability as compared to that of the average person. But even this may not be enough. In the case of the medical diagnostician discussed earlier, the layman may have no information that might give him a clue as to which of diseases A and B is the more probable, while anyone with a certain amount of rudimentary medical knowledge may know that disease A generally occurs much more frequently than disease B; yet his prediction of A rather than B on this basis alone would not qualify him as a reliable diagnostician. Thus a more subtle assessment of the qualifications of an expert may require his comparison with the average person having some degree of general background knowledge in his field of specialization. One method of scoring experts somewhat more subtly than just by their reliability is in terms of their "accuracy": the *degree of accuracy* of an expert's predictions is the correlation between his personal probabilities p and his correctness in the class of those hypotheses to which he ascribed the probability p. Thus of a highly accurate predictor we expect that of those hypotheses to which he ascribes, say, a probability of 70 per cent,

approximately 70 per cent will eventually turn out to be confirmed. Accuracy in this sense, by the way, does not guarantee reliability,[13] but accuracy in addition to reliability may be sufficient to distinguish the real expert from the specious one.

THE DEPENDENCE OF PREDICTIVE PERFORMANCE ON SUBJECT MATTER

Not only are some experts better predictors than others, but subject-matter fields differ from one another in the extent to which they admit of expertise. This circumstance is of course in some instances due to the fact that the scientific theory of the field in question is relatively undeveloped. The geology of the moon and the meteorology of Mars are less amenable to prediction than their mundane counterparts, although no greater characteristic complexity is inherent in these fields. In other cases, however, predictive expertise is limited despite a high degree of cultivation of a field, because the significant phenomena hinge on factors that are not particularly amenable to prediction.

In domains in which the flux of events is subject to gradual transitions and constant regularities (such as, say, astronomy), a high degree of predictive expertise is possible. In those fields, however, in which the processes of transition admit of sharp jolts and discontinuities, which can in turn be the effects of causal processes so complex and intricate as to be "chance" occurrences for all practical purposes, predictive

[13] For instance, suppose experts A and B each gave 100 responses, assigning probabilities 0.2, 0.4, 0.6, and 0.8 to what in fact were the correct alternatives among 100 choices of H and $\sim H$, as follows:

p	A	B
0.2	10	0
0.4	20	0
0.6	30	60
0.8	40	40

Then A is perfectly *accurate* (for example, exactly 60 per cent, or 30, of the 20 + 30, or 50, cases to which he assigned 0.6 were correct), but he is only 70 per cent *reliable*; B, on the other hand, is 100 per cent reliable, but his accuracy is quite faulty (for example, 100 per cent, rather than 60 per cent, of the 60 cases to which he assigned 0.6 were correct).

expertise is inherently less feasible. The assassination of a political leader can altogether change the policies of a nation, particularly when such a nation does not have a highly developed complex of institutions to ensure gradualness of policy change. Clearly no expert on a particular country can be expected to have the data requisite for a prediction of assassinations; that is, his relevant information is virtually certain not to include the precisely detailed knowledge of the state of mind of various key figures that might give him any basis whatsoever for assigning a numerical value as his personal probability to the event in question. This situation is quite analogous to that of predicting the outcome of a particular toss of a coin; only the precise dynamic details of the toss's initial conditions might provide a basis for computing a probability other than one-half for the outcome, and these details again are almost certainly unavailable. We may here legitimately speak of "chance occurrences," in the sense that an expert, unless he has the most unusual information at his disposal, is in no better a position than the layman to make a reliable prediction.

In the inexact sciences, particularly in the social sciences, the critical causal importance of such chance events makes predictive expertise in an absolute sense difficult and sometimes impossible, and it is this, rather than the quality of his theoretical machinery, that puts the social scientist in a poor competitive position relative, say, to the astronomer.

However, when the expert is unable to make precise predictions because of the influence of chance factors, we can expect him to indicate the major contingencies on which future developments will hinge. Even though the expert cannot predict the specific course of future events in an unstable country, he should be able to specify the major branch points of future contingencies, and to provide personal probabilities conditionally with respect to these. Thus, for example, while it would be unreasonable to expect an expert on the American economy to predict with precision the duration of a particular phase of an economic cycle (such as a recession), it is entirely plausible to ask him to specify the major potential "turning points" in the cycle (such as increased steel production at a certain juncture), and to indicate the probable courses of development ensuing upon each of the specified alternatives.

Such differences in predictability among diverse subject-matter fields

lead to important consequences for the proper use of experts. One obvious implication is that it may clearly be more profitable to concentrate limited resources of predictive expertise on those portions of a broader domain that are inherently more amenable to prediction. For example in a study of long-range political developments in a particular geographic area, it might in some cases be preferable to focus on demographic developments rather than the evolution of programs and platforms of political parties.

However, the most important consideration is that even in subject-matter fields in which the possibility of prediction is very limited, the exercise of expertise, instead of being applied to the determination of *absolute* personal probabilities with respect to certain hypotheses, ought rather more profitably be concentrated on the identification of the relevant branch points and the associated problem of the *relative* personal probabilities for the hypothesis in question, relative, that is, to the alternatives arising at these branch points.

Even in predictively very "difficult" fields—such as the question of the future foreign policy of an unstable country—the major branch points of future contingencies are frequently few enough for actual enumeration, and although outright prediction cannot be expected, relative predictions hinging on these principal alternative contingencies can in many instances serve the same purposes for which absolute predictions are ordinarily employed. For example, it would be possible for a neighboring state, in formulating its own policy toward this country, to plan not for "the" (one and only) probable course of developments but to design several policies, one for each of the major contingencies, or perhaps even a single policy that could deal effectively with all the alternatives.

PREDICTIVE CONSENSUS TECHNIQUES

The predictive use of an expert takes place within a rationale which, on the basis of our earlier discussion, can be characterized as follows: We wish to investigate the predictive hypothesis H; with the expert's assistance, we fix on the major items of the body of explicit evidence E that is relevant to this hypothesis; we then use the expert's personal probability valuation $pp(H, E)$ as *our* estimate of the degree of confirmation of H on the basis of E, that is, as our estimated value of $dc(H, E)$.

This straightforward procedure, however, is no longer adequate in those cases in which *several* experts are available. For here we have not the single value $pp(H, E)$ of only one expert, but an entire series of values, one for each of the experts: $pp_1(H, E)$, $pp_2(H, E)$, etc. The problem arises: How is the best joint use of these various expert valuations to be made?[14]

Many possible procedures for effecting a combination among such diverse probability estimates are available. One possibility, and no doubt the simplest, is to select one "favored" expert, and to accept his sole judgment. We might, for example, compare the past predictive performance of the various experts, and select that one whose record has been the most successful.

Another simple procedure is to pool the various expert valuations into an average of some sort, possibly the median, or a mean weighted so as to reflect past predictive success.

Again, the several experts might be made to act as a single group, pooling their knowledge in round-table discussion, if possible eliminating discrepancies in debate, and the group might then—on the basis of its corporate knowledge—be asked to arrive at one generally agreeable corporate "personal" probability as its consensus, which would now serve as our *dc* estimate. (A weakness in this otherwise very plausible-sounding procedure is that the consensus valuation might unduly reflect the views of the most respected member of the group, or of the most persuasive.)

One variant of this consensus procedure is to require that the experts, after pooling their knowledge in discussion, and perhaps after debating the issues, set down their separate "second-guess" personal probabilities, revising their initial independent valuations in the light of the group work. These separate values are then combined by an averaging process to provide our *dc* estimate. The advantage of such a combination of independent values over the use of a single generally acceptable group value is that it tends to diminish the influence of the most vociferous or influential group member. Incidentally, in any consensus method in which separate expert valuations are combined, we can introduce the refinement of weighting an expert's judgment so as to reflect his past performance.

[14] Compare A. Kaplan, A. L. Skogstad, and M. A. Girshick, "The Prediction of Social and Technological Events," ***Public Opinion Quarterly***, Spring, 1950.

Another consensus procedure, sometimes called the "Delphi Technique," eliminates committee activity altogether, thus further reducing the influence of certain psychological factors, such as specious persuasion, the unwillingness to abandon publicly expressed opinions, and the bandwagon effect of majority opinion. This technique replaces direct debate by a carefully designed program of sequential individual interrogations (best conducted by questionnaires) interspersed with information and opinion feedback derived by computed consensus from the earlier parts of the program. Some of the questions directed to the respondents may, for instance, inquire into the "reasons" for previously expressed opinions, and a collection of such reasons may then be presented to each respondent in the group, together with an invitation to reconsider and possibly revise his earlier estimates. Both the inquiry into the reasons and subsequent feedback of the reasons adduced by others may serve to stimulate the experts into taking into due account considerations they might through inadvertence have neglected, and to give due weight to factors they were inclined to dismiss as unimportant on first thought.

We have done no more here than to indicate some examples from the spectrum of alternative consensus methods. Clearly there can be no one universally "best" method. The efficacy of such methods obviously depends on the nature of the particular subject matter and may even hinge on the idiosyncracies and personalities of the specific experts (such as their ability to work as a group). Indeed this question of the relative effectiveness of the various predictive consensus techniques is almost entirely an open problem for empirical research, and it is strongly to be hoped that more experimental investigation will be undertaken in this important field.

SIMULATION AND PSEUDO-EXPERIMENTATION

We have thus far, in discussing the intrinsic use of experts or of groups of experts, described their function as being simply to predict, in the light of their personal probabilities, the correctness or incorrectness of proposed hypotheses. This description, while apt in principle and often in practice, may in some instances be overly simplified.

For one thing, in situations concerned with complicated practical

problems, no clear-cut hypothesis to which probability values could be meaningfully attached may be immediately discernible. For another, several kinds of expertise, all interacting with each other, may have to be brought to bear simultaneously in anything but a straightforward manner. Examples of such cases are provided by questions such as "How can tension in the Middle East be relieved?" "What legislation is needed to reduce juvenile delinquency?" "How can America's schools improve science instruction?" Here, before even a single predictive expert can be used intrinsically, some at least rudimentary theoretical framework must be constructed within which predictive hypotheses can be stated. This, of course, calls for expertise of the extrinsic kind. Generally, the process involved is somewhat as follows. The situation at hand (say, current crime patterns) is analyzed—that is, it is stated in terms of certain specific and, it is hoped, well-defined concepts. This step usually involves a certain amount of abstraction, in that some aspects of the situation that are judged irrelevant are deliberately omitted from the description. Then either some specific action is proposed, and a hypothesis stated as to its effects on the situation in question or, more typically, a law or quasi-law is formulated, stating that in situations of the kind at hand, actions of a certain kind will have such-and-such consequences.

A variant of this is of considerable epistemological significance. Instead of describing the situation directly, a *model* of it is constructed, which may be either mathematical or physical, in which each element making up the real situation is simulated by a mathematical or physical object, and its relevant properties and relations to other elements are mirrored by corresponding simulative properties and relations. For example, any geographical map may be considered a (physical) model of some sector of the world; the planetary system can be simulated mathematically by a set of mass-points moving according to Kepler's laws; and a city's traffic system can be simulated by setting up a miniature model of its road net, traffic signals, and vehicles.[15] Now, instead of formulating hypotheses and predictions directly about the real world, it is possible to do the same thing

[15] On simulation in traffic research, see W. H. Glanville, "Road Safety and Traffic Research in Great Britain," *Journal of the Operations Research Society of America*, Vol. 3, 1955, pp. 283–299, reprinted in J. F. McCloskey and J. M. Coppinger, *Operations Research for Management*, Vol. 2, Baltimore, 1956, pp. 82–100.

about the model. Any results obtained from an analysis of the model, to the extent that it truly simulates the real world, can then later be translated back into the corresponding statements about the latter. This injection of a model has the advantage that it admits of what may be called "pseudo-experimentation" ("pseudo" because the experiments are carried out in the model, not in reality). For example, in the case of traffic-system analysis, pseudo-experimentation may produce reliable predictions as to what changes in the time sequence of traffic signals will ease the flow of traffic through the city.

Pseudo-experimentation is nothing but the systematic use of the classical idea of a hypothetical experiment; it is applied when true experimentation is too costly or physically or morally impossible[16] or—as we shall discuss next—when the real-world situation is too complex to permit the intrinsic use of experts. The application of simulation techniques is a promising approach, the fruitfulness of which has only begun to be demonstrated in documented experiments.[17] It is particularly promising when it is desirable to employ intrinsically several experts with varying specialties in a context in which their forecasts cannot be entered independently but where they are likely to interact with one another. Here a model furnishes the experts with an artificial, simulated environment. Within this environment they can jointly and simultaneously experiment, responding to the changes induced by their actions and

---

[16] Classical examples of such pseudo-experimentation are found in atomic physics and in military analysis. Atomic physics deals with particles so small as to preclude experimentation requiring direct manipulation and observation; what has been done here is to construct a mathematical model of certain atomic and nuclear processes and then to use Monte Carlo sampling techniques to conduct not real, but paper, experiments, in which the paths of fictitious particles of the model are "observed" as the latter go through a series of random collisions, deflections, or what not; in this "experiment" the features supposedly not under the control of the "experimenter" are assumed to be subject to given probability distributions, so that the chance fluctuations that would naturally occur can be simulated in the model by the operation of an artificial chance device. Similarly, the military, in order to evaluate the effectiveness of alternative weapon systems (which clearly cannot be fully tested directly), conduct pseudo-experiments on simulation models.

[17] See, for instance, J. L. Kennedy and R. L. Chapman, *The Background and Implications of the Systems Research Laboratory Studies*, Symposium on Air Force Human Engineering, Publication No. 455, National Academy of Sciences, National Research Council, 1956; and R. L. Chapman, "Simulation in RAND's Systems Research Laboratory," and W. W. Haythorn, "Simulation in RAND's Logistics Systems Laboratory," both in *Report of Symposium on System Simulation*, Waverly Press, 1958.

acquiring through feedback the insights necessary to make successful predictions within the model and thus indirectly about the real world.

This technique lends itself particularly to predictions regarding the behavior of human organizations, inasmuch as the latter can be simulated most effectively by having the experts play the roles of certain members of such organizations and act out what in their judgment would be the actions, in the situation simulated, of their real-life counterparts.[18] Generally it may be said that in many cases judicious pseudo-experimentation may effectively annul the oft-regretted infeasibility of carrying out experiments proper in the social sciences by providing an acceptable substitute that, moreover, has been tried and proved in the applied physical sciences.

OPERATIONAL GAMING

A particular case of simulation involving role-playing by the intrinsic experts is known as operational *gaming*, a special case of which is war gaming. A simulation model may properly be said to be gaming a real-life situation if the latter concerns decisionmakers in a context involving conflicting interests. In operational gaming, the simulated environment is particularly effective in reminding the expert, in his role as a player, to take *all* the potentially relevant factors into account in making his predictions; for if he does not, and chooses a tactic or strategy that overlooks an essential factor, an astute "opponent" will soon enough teach him not to make such an omission again.

Aside from the obvious application of gaming to the analysis of military conflict—of which there are numerous examples, ranging from crude map exercises to sophisticated enterprises requiring the aid of high-speed computing equipment—gaming has been used to gain insights into the nature of political and economic conflict. In the political field, cold-war situations have been explored in this manner; and in the economic field, inroads have been made into analyses of bargaining and industrial competition.[19]

---

[18] See, for example, the American Management Association study of *Top Management Decision Simulation*, New York, 1957.

[19] Compare C. J. Thomas and W. L. Deemer, Jr., "The Role of Operational Gaming in Operations Research," *Journal of the Operations Research Society of America* (*JORSA*), Vol. 5, 1957, pp. 1–27; A. M. Mood and R. D. Specht, *Gaming as a Technique of Analysis*,

We note in passing that operational games differ greatly in the completeness of their rules. These may be complete enough so that at each stage the strategic options at the players' disposal are wholly specified and the consequences resulting from the joint exercise of these options are entirely determined; this would mean that the model represents a complete theory of the phenomena simulated in the game. On the other hand, neither of these factors may be completely determined by the rules, in which case it is up to an umpiring staff to allow or disallow proposed strategies and assess their consequences. Umpiring in this sense is yet another important device for the use of expertise intrinsically within the framework of a scientific theory (namely, the model in question).

REVIEW OF THE MAIN THESES

Before considering certain recommendations that seem to us to emerge as conclusions from the analysis just presented, it is appropriate to pause briefly to review the main points of the foregoing discussion. Our starting point was the distinction between the "exact" and the "inexact" areas of science. It is our contention that this distinction is far more important and fundamental from the standpoint of a correct view of scientific method than is the case with superficially more pronounced distinctions based on subject-matter diversities, especially that between the social and the physical sciences. Some branches of the social sciences (certain parts of demography, for example), which are usually characterized by the presence of a formalized mathematical theory, are methodologically analogous to the exact parts of physics. By contrast, the applied, inexact branches of physical science—for instance, certain areas of engineering under "extreme" conditions—are in many basic respects markedly similar to the social sciences.

The RAND Corporation, Paper P-579, October 19, 1954; W. E. Cushen, "Operational Gaming in Industry," in J. F. McCloskey and J. M. Coppinger, *Operations Research for Management*, Vol. 2, Baltimore, 1956, pp. 358–375; R. Bellman *et al.*, "On the Construction of a Multi-stage, Multi-person Business Game," *JORSA*, Vol. 5, 1957, pp. 469–503; F. M. Ricciardi, "Business War Games for Executives: A New Concept in Management Training," *Management Review*, 1957, pp. 45–56; J. C. Harsanyi, "Approaches to the Bargaining Problem Before and After the Theory of Games," *Econometrica*, Vol. 26, 1955; F. M. Ricciardi *et al.*, "Top Management Decision Simulation," *American Management Association*, 1957; and G. K. Kalisch *et al.*, "Some Experimental *n*-Person Games," in Thrall, Coombs, and Davis (eds.), *Decision Processes*, John Wiley & Sons, Inc., New York, 1954.

This similarity applies both to methods of explanation and to methods of prediction. Partly because of the absence of mathematically formalized theories, explanations throughout the area of the inexact sciences—within the physical- and the social-science settings alike—are apt to be given by means of the restricted generalizations that we have called "quasi-laws." The presence of such less-than-universal principles in the inexact sciences creates an asymmetry between the methods of explanation and those of prediction in these fields. This suggests the desirability of developing the specifically predictive instrumentalities of these fields, for once the common belief in the identity of predictive and explanatory scientific procedures is seen to be incorrect, it is clearly appropriate to consider the nature and potentialities of predictive procedures distinct from those used for explanation. As for predictions in the inexact sciences (physical as well as social), these can be pragmatically acceptable (as a basis for action) when based on methodologically even less sophisticated grounds than are explanations—grounds such as expert judgment.

These general considerations regarding the methodology of the inexact sciences hold particularly intriguing implications for the possibility of methodological innovation in the social sciences. Here the possible existence of methods that are unorthodox in the present state of social-science practices merits the closest examination. This is particularly true with respect to the pragmatic applications of the social sciences (in support of decisionmaking, for example), in which the predictive element is preponderant over the explanatory.

One consideration of this sort revolves about the general question of using expertise. We have stressed the importance in the social sciences of limited generalizations (quasi-laws), which cannot necessarily be used in a simple and mechanical way, but whose very application requires the exercise of expert judgment. And more generally, when interested in prediction in this field (especially for decisionmaking purposes), we are dependent on the experts' personal probability valuations for our guidance. A systematic investigation of the effective use of experts represents a means by which new and powerful instruments for the investigation of social-science problems might be forged.

Further, the use in social-science contexts of a variety of techniques borrowed from other, applied, sciences that are also inexact (such as engi-

neering applications and military and industrial operations research) deserves the most serious consideration. For there are numerous possibilities for deriving leads on methods that are also potentially useful in the social sciences; in particular the use of simulation as a basis for conducting pseudo-experiments comes to mind. Finally, there is the important possibility of combining simulation with the intrinsic use of expertise, especially by means of the technique of operational gaming. This prospect constitutes a method whose potential for social-science research has hitherto gone virtually wholly unexplored. It is our hope that this neglect will soon be remedied.

SOME TASKS FOR METHODOLOGICAL RESEARCH

The thoughts contained in this study are intended to represent a challenge to those who would like to see the applied social sciences narrow the gap that has been created between them and the applied physical sciences by the explosive progress of technology in the first half of this century. A particularly promising prospect, it seems to us, is a pragmatic reorientation of social-science methodology along some of the lines that have proved successful in the fellow inexact sciences in the applied physical field.

In order to achieve this, much work has yet to be done. Aside from the need for further conceptual analysis, especially with regard to the status of quasi-laws in theory construction and the expansion of the degree-of-confirmation concept to languages containing relations and functions, numerous empirical studies are suggested by the approach that we are recommending.

The following is a list of areas of such research. It is intended to be a suggestive, rather than an exhaustive, list of the type of effort required to implement this new approach.

1. *Methodology of expertise*:
   - *a*. Performance of the individual expert—for example, selection and training of experts, aids to their performance, scoring systems for predictions.
   - *b*. Performance of groups of experts—for example, methods

of consensus formation, "Delphi" techniques for interrogation with feedback, investigation of other multi-expert structures.

c. Psychological problems in the use of expert groups—for example, reduction of respondents' bias in conference situations and as game players, simulation of motivation in pseudo-experiments requiring role-playing, man-machine interaction.

2. *Methodology of pseudo-experimentation*:

(Here our suggestions are, on the whole, only obliquely methodological, since—in our opinion—a firm methodology will evolve gradually from a process of prolonged trial and error.)

a. Simulation techniques in the social sciences—for example, simulation of industrial or business processes or of the operation of some sector of the national economy, or of some governmental activity.

b. Gaming techniques in the social sciences—for example, gaming of industrial competition, cold-war games, foreign-trade and foreign-investment gaming.

c. Problems inherent in pseudo-experimentation—for example, the question of "controlled" experiments, problems of scaling arising in the translation of results from a simulation model to the real world.

In all of these areas some preliminary studies have been carried out, which, while insufficient in themselves, lend great promise to more extensive efforts. Along such lines, it seems to us, there lie valuable and as yet only fragmentarily exploited hopes of augmenting the range of the methodological instrumentalities of the applied inexact sciences, and, in particular, those of the applied social sciences.