VARIOUS MEANINGS OF "THEORY"*

ANATOL RAPOPORT

Mental Health Research Institute
University of Michigan

So many discussions go astray because the same words are used in different senses by adherents of different points of view that it seems imperative to start practically every discussion by clarifying the meanings of terms. Yet this problem is easier posed than solved. We in academic life owe understandable allegiance to erudition and to elegance of expression, and all too often we take a definition to be adequate (in the sense of clarifying meaning) if it sounds well. More is required, of course. Clarification of meaning (whether couched in formal definitions or in illustrative examples) takes place only if the terms defined are actually geared to the experience of the people concerned. This is a serious problem, because the experiences of people, although they overlap, can be widely disparate. Particularly among us in academic life the disparity may be quite wide. For our experience is very largely the experience of thinking, and thinking is tempered by language in the broadest sense, that is, by the way ideas are organized. And various ways of organizing ideas are imposed on us by our disciplines. Discipline means constraint. Discipline is essential for any organized activity. And so in academic disciplines, "discipline" means constraint on the mode of thought. It prescribes the repertoire of concepts, the patterns of classification, the rules of evidence, and the etiquette of discourse.

Cross-disciplinary endeavor, therefore, depends on the ability of the participants to think in terms of more than one language—a feat more difficult than the ability to think, say, both in English and French, because the languages of the disciplines vary not only in their vocabularies and grammars (as ordinary languages do) but also in deeper aspects, whose meaning I hope can be conveyed by the phrase, "principles of organizing thought."

I

Our concern here is the role of theory. I will first try to convey the experiences behind this term characteristic of the exact sciences (which I will define by their predominant thought patterns). Here I am on sure ground, because this is the field in which I myself have been "disciplined." Afterwards I will try to convey the experiences behind the term "theory" in other than exact sciences, particularly in the social sciences. Here I can only give my impressions. I am neither a political scientist nor a psychologist (to give two examples of "non-exact" sciences), and so I have not been properly disciplined to speak with authority on the way a term like "theory" is used there. My remarks are to be taken only

* A paper read at the 1958 annual meeting of the American Political Science Association, St. Louis, Missouri, September, 1958.

as an account of how the thinking of social scientists and others about theory appears to someone who thinks of theory in terms imposed by the discipline of exact science.

In the exact sciences, a theory is a collection of theorems. This concept of theory is also a partial definition of an exact science. It is the *entire* definition of any self-contained branch of mathematics. Some maintain, however, that mathematics is not a science, because it makes no assertions about the observable world. Whether one demands that a science must necessarily make assertions about the outside world is a matter of taste. I tend to accept this limitation on what is to be called a "science" and thus I agree to exclude mathematics from the sciences. What then is an exact science?

To begin with, it is a collection of theorems (to be presently defined); but the theorems have to be translatable into assertions about the tangible world and these assertions should be verifiable within certain limits of accuracy. The theorems are what makes a science exact; the accuracy of the assertions is what makes it successful.

Next we define a theorem. A theorem is a proposition which is a strict logical consequence of certain definitions and other propositions. The validity of a theorem, then, usually depends on the validity of other theorems. This tracing of antecedents goes on until the rock bottom is reached . . . assertions which are not proved but simply assumed, and terms which are not defined but simply listed. In a mathematical system it is unnecessary (in fact, impossible) either to prove these basic assumptions (the postulates) or to define the basic terms. This is what Bertrand Russell meant when he said that in mathematics we never know what we are talking about, nor whether what we are saying is true. In logic, the situation is exactly the same. Indeed logic is often taken to be a branch of mathematics or vice versa.

According to our criterion for science, however, we demand that some of the terms used in a science be related extensionally to referents and at least some of the assertions be empirically verifiable. I say some, not necessarily all, and this is an important point to which I will return.

Practically all the exact sciences we know are mathematical or at least highly mathematicized. This is by no means an indication of some supernatural power inherent in mathematics but rather of the propensity of mathematicians for preempting new territories and of adjusting and extending mathematical methods so as to be able to deal with different content areas. The central fact is that a necessary adjunct of an exact science is a set of completely rigid rules of deduction. It is the rigidity of these rules, not the accuracy of the assertions or precision of measurements, which makes an exact science. Now wherever there are such rules, a symbolism is invented as a purely mnemonic device. Wherever a symbolic notation occurs coupled with rules of deduction, *i.e.*, of manipulating the symbols, the mathematician steps in and assumes jurisdiction over the territory, or else the practitioners of the newly invented system of symbol manipulation are called mathematicians. This is what account for the all-pervasiveness of "mathematics" in the exact sciences.

II

I will now describe in greater detail by an illustrative example what is meant by a theory in an exact science. Specifically I will take a problem from mechanics, the earliest and one of the most successful of the exact sciences.

Consider a pendulum, that is, a weight supported by a string. The problem is to "explain" its motion. Immediately the question arises, what one means by "explain"? Obviously any explanation will contain statements introduced by "because" in reply to questions starting with "why." But what kind of questions are these? They are likely to be determined by what is observed about the pendulum. What is observed, in turn, will depend on what is singled out for observation. A question like "Why does the pendulum move around?" is so vague that it frustrates any attempt to answer it. Or, viewed in another way, it is trivially easy to answer, *because* it is so vague. One might say, for instance, that the pendulum moves around because there are forces acting on it.

The first task of an exact science, therefore, is to make the questions precise. The question "Why does the pendulum move *as it does?*" is more to the point. But the phrase "as it does" now lays the questioner open to a counter-question: "What do you mean, 'as it does'?" This is a challenge to describe how in fact the pendulum does move, and this temporarily turns the attention away from an "explanation" toward description. To explain anything we must first circumscribe just what we are going to explain.

The first problem, then, is to *describe* the motion of the pendulum. This immediately introduces a motivation to simplify the situation. If the pendulum is constrained only by the string, its bobbing about will appear at first too chaotic to yield to a systematic description. Let us therefore constrain the pendulum to move in one plane, the way clock pendulums move. Now it moves just "back and forth."

But "back and forth" is still too crude a description. How can we make it more precise? Here is where the fundamental orientation of mechanics (of motion) as an exact science begins to direct one's methods of observation. This orientation prescribes what shall be of prime interest in any investigation of motion. If a moving object can be specified at any given moment by its position, the description of motion consists of associating a sequence of positions with a sequence of moments of time. In the simplest case, the position can be uniquely specified by a single number, and, of course, time can also be so specified. The position of the pendulum bob, for example, can be specified by the angle of deflection (positive or negative) from the plumb line, or else equally well by the horizontal displacement of the center of the bob from the position of rest. A complete description of the motion, then, will be given by a table specifying the deflection at each moment of time. Such a table is called a mathematical function.

It can be demonstrated by experiment that the mathematical function $x = A \sin (mt)$ will very nearly, but not quite, describe the motion of the pendulum. Here x is the horizontal deflection of the bob, A is the maximum deflection, t is time, and m is a certain constant to be presently discussed. The qualification "not quite" is of paramount importance, as we shall see. First, however, let us see what we mean when we say that a mathematical function describes a set of data.

Actually a set of data in which values of two variables are related can be represented by a set of points in the plane, as many points as there have been readings. In particular, suppose one took only three readings of the pendulum position and got three points on a graph whose axes are displacement *vs.* time. A mathematical function will "describe" these data if the curve corresponding to the function can be passed through these points. Now a circle or a parabola can always be passed through any three (non-collinear) points, and so can many other kinds of curves. All of them, then, can be said to "describe" this limited set of data. Of course, if more readings are taken, the additional points may not lie on the same curve. However, given any number of points, a great many curves can be made to fit them, which is to say that a great many equations can describe the data, indeed exactly, not approximately. Why, then, is the particular function $x = A \sin (mt)$ chosen, even though it does *not* describe the data exactly?

This function is chosen because it does more than describe the data. It *explains* the data in the sense of "explanation" as it is used in an exact science. And the discrepancies between it and the data are accounted for by the inadequacies of the fundamental assumptions in terms of which the explanation is made.

The fundamental assumptions concern the general laws which supposedly govern the motion of a body subjected to forces. In the case of the pendulum, we can from the geometry of the situation analyze the forces acting on the bob, hence *derive* the mathematical form of the motion. The equation above gives the derived (idealized) motion of an idealized body in an idealized environment. The discrepancy between the idealized and the actual state of affairs is supposed to account for the discrepancies between the prescribed and the realized motion. Some of these idealizing assumptions are as follows:

1. The supporting string or rod is supposed to have no mass.
2. No friction or air resistance acts on the pendulum.
3. The bob is assumed to have mass but no extension.
4. The (small) horizontal displacement is supposed to be proportional to the angle of displacement.

Etc.

All of these assumptions are false. Yet the physicist continues to make them. Why? In return for sacrificing precision (precision must in any event be sacrificed wherever measurements are involved), he gains simplicity and, what is more important, he gets at the fundamentals (almost in the Platonic sense of the word) of the situation. This allows him to subsume a great many phenomena under a single scheme. For example, the assumed laws of motion and idealized properties of the pendulum allow the physicist to derive *from the same set of postulates* the following additional relations:

1. The period of the pendulum will be independent of the mass of the pendulum.

2. Within limits, it will be independent of the amplitude of oscillation.
3. For large amplitudes, the period will become dependent on the amplitude, and the precise nature of this dependence can be predicted.
4. The period will be directly proportional to the square root of the length of the pendulum.
5. The period will be inversely proportional to the square root of the acceleration of gravity. This relation explains the meaning of the constant m in the equation of the pendulum's motion. The constant involves the square root of the ratio of the acceleration of gravity to the length of the pendulum.
6. The horizontal velocity of the pendulum will be given by the function $v = mA \cos (mt)$.
7. The same scheme can be extended to the spherical pendulum, *i.e.*, one not constrained to a single plane. That is to say, from the same set of assumptions one can get many additional results, for instance, the precession of the plane of rotation of the spherical pendulum, etc. Indeed, all the observed motions of the heavenly bodies are derived from exactly the same three or four assumptions about motion which are supposed to underlie the swinging of a simple pendulum.

The story of the pendulum illustrates the power of the mathematical model. A mathematical model is much more than a description of events in terms of the mathematical relations among the variables. It is rather a set of assumptions often referring to a highly idealized situation, from which assumptions the relations to be observed are *derived*, to be compared with observations. Agreement with observations corroborates the model. Most important for corroboration is the prediction of *other* relationships, perhaps not hitherto observed. At any rate, the more are the relationships derived, and the fewer are the relationships assumed, the more powerful is the model. A trivial model does no more than lead to relationships which observations had suggested in the first place. Such a model penetrates no deeper than the observational level and is therefore purely descriptive, not explanatory. It simply restates in other terms what has been observed.

The method just outlined is applicable to many widely disparate exact sciences, though the content, of course, will be different in each case. In mechanics, for example, the fundamental assumptions have to do with the laws of motion. In mathematical genetics, another exact science in the sense defined, but not as successful as mechanics, the fundamental assumptions have to do with something quite different, namely the re-shuffling, segregation, and recombination in sexually reproducing organisms of entities supposed to be carriers of separate inherited traits. In this science, the laws governing these events are assumed to be not laws of mechanics or of electrodynamics or of thermodynamics but "laws of chance." For these too an exact mathematical theory exists. And on the basis of it, given certain combinations of genotypes, certain patterns of mating, certain linkages among the genes upon the chromosomes, certain selection pressures exerted by the environment, etc., the distributions of the genotypes and of the phenotypes in the succeeding generations can be computed. Here too discrepancies between prediction and observation are unavoidable, because of the idealizing assumptions which cannot be avoided. But the principle of investigation is the same as in theoretical mechanics. One derives a collection of theorems about how things should happen under idealized conditions. The discrepancies are attributed to the imperfections in the assumptions, and in the initial observations. The discrepancies provide the leverage for further refinements of the theory.

Economics may be viewed in the same light. The assumptions here have to do (in the classical picture) with the relations between price levels, supply, demand, maximization of profit (or expected profit) by the so called "economic man," etc. The mathematical scheme being given, predictions on the basis of initial observations can be made. In mathematical economics agreement between theory and observation is not often good, and many are led to dismiss mathematical methods in economics on this ground. In the light of what has been said, such an attitude is not justifiable. The predictive power of mathematical meteorology, especially in its early stages, was also quite poor. Yet there was never any question that air currents, temperature gradients, and all the other conditions studied in meteorology are subject to the strict laws of physics. The low success of primitive meteorology is entirely attributable to the complexity of the phenomena compared with the drastically simplified assumptions which it was still possible to handle mathematically. Crude meteorology was merely a stage in the development of more refined and more successful meteorology. The original conceptual scheme was correct. Only the tools had to be sharpened.

Thus the bluntness of the available mathematical tools is not a sufficient ground for rejecting them in principle. It is an entirely different matter if the question is raised whether the tools applied are the right *kind* of tools. As is conceded by all who understand the mathematical method, its power is enormous where it can be applied. The big question is, where can it be applied? Clarification is needed here. For many, mathematics means classical mathematics, that is, the mathematics of eighteenth-century physics, derived from the differential and integral calculus. As pointed out, the range of logical disciplines now called "mathematics" has enormously increased, both in techniques and in the variety of conceptualizations, so that any inadequacy of classical mathematics for dealing with problems of social science can by no means be taken as an indication of the inapplicability of mathematics in principle.

Nevertheless, let us first examine the conditions that must be fulfilled in order that classical mathematical methods may be utilized to advantage in theoretical social science.

III

First, there must be sharply defined, quantitative variables singled out for study. In the mechanics of motion, we have seen that position and time were fundamental. Actually there are three fundamental kinds of quantities in mechanics, from which all others are derived: length, time, and mass. Nor is there any question (in classical mechanics) how these quantities are to be measured. In other branches of physics, there are other quantities, for example heat and temperature in thermodynamics, electric charge current, and strength of electric and magnetic fields in electrodynamics, etc. Where probability theory is used, the fundamental quantities are frequencies of occurrence of events or of "types," and so on. In pursuing the question to what extent a quantitative variable can be sharply defined, a most important problem looms: the problem of recognition.

Note that the problem is trivial in mechanics. To determine the position of an object, we must, of course, recognize the object in all positions, but this is ordinarily so easy as to present no problem. As we pass from physics to chemistry, the problem of recognition becomes more important. For example, before quantities or concentrations of substances can be measured, the substances must be recognized. To be sure, there are unambiguous rules for recognizing substances by their "properties," and to the extent that the "properties" are defined in terms of classical physical measurements, the problem of recognition is solved. But it makes itself known.

As we pass to biology, the problem of recognition becomes even more serious. It now takes special training, sometimes quite prolonged, to tell one type of organism from another, one tissue from another, to interpret what is seen through a microscope, etc.

In the behavioral sciences, the problem of recognition becomes paramount. Since these sciences have only recently arisen from the humanities, their terms are derived largely from common sense and from intuitive notions at best, and from deeply rooted pre-scientific notions and prejudices at worst. Outside of science, no need may be felt to endow terms with operational meanings: one's intuitive meaning seems to suffice on the basis of the universal naive assumption that the other's perceptions are like one's own, or else something is wrong with the other's perceptions. Where there is no consensus on recognition, there can, of course, be no question of quantification or measurement, and so the first requirement of exact (or mathematical) science seems to be not fulfilled. Some workers in behavioral sciences feel that this difficulty precludes in principle the extension of mathematical methods to behavioral science.

The other condition usually assumed necessary for an exact science is this. Given a set of unambiguously measurable variables, one must be able to choose some assumptions about how they are related which reasonably reflect "reality." It is conceded, of course, that only an idealization of reality can be reflected in any finite set of assumptions, but it is maintained that the idealization should at least come close to reality. In mechanics, this condition is fulfilled. True, there is no such thing as frictionless motion, a perfectly rigid body, extensionless mass, etc., but these idealizations are in many instances well enough "approached" by reality. Once a set of assumptions is chosen, the mathematics required to derive theorems and conclusions from them must be amenable to being handled by the human mind. Again, it is maintained by some that even if unambiguous variables could be singled out (as they are in economics) and even if reasonably accurate assumptions could be made, the resulting mathematical system would be too unwieldy to be useful—there are too many relevant variables, and they are too intricately interwoven to permit treatment by existing mathematical techniques.

When it is proposed to simplify the situation by holding all but a few variables constant, it is pointed out (quite correctly) that, first, in many fields of investigation this is practically impossible: one cannot experiment with national economics or with real political systems; second, even where experimentation is possible, controlled conditions introduce distortions of such magnitude as to make extrapolations from controlled to natural situations, *i.e.*, from *in vitro* to *in situ* (to use the physiologist's terms), worthless.

Now I will develop what I want to say further along two lines. First I want to recognize to a certain extent the justice of the criticism of *premature* uses of mathematics in the social sciences and thereby the inevitability of an entirely different conception of "theory" and of the notion of a model, which have arisen in social science. But then I want to point out the limitations of this conception and its underlying assumption concerning the nature and function of mathematical methods and of mathematics itself.

IV

There are social scientists who understand the nature and the importance of the operational definition. Whether motivated by a hope of an ultimate possibility of mathematization of their disciplines (they are more likely to criticize premature mathematization, not mathematization in principle), or because they have been caught up by the spirit of positivism, which has become dominant in modern systematic thought, they undertook the task of creating a sound and consistent terminology of social science, particularly in sociology.

This task, however, would be trivial, if it were confined to a compilation of a glossary, no matter how "operational" the definitions in it. For definitions are arbitrary—they are no more than agreements on how to use terms. Implicit in the work of Parsons and of Levy (to name two workers who recognize the importance of systematization) are not only attempts to map observable events on terms (which is what is done in operational definitions). Their main efforts are directed toward selecting events and combining them in such a fashion as to make the terms applied to these combinations fruitful in the development of a social theory which eventually is to become a collection of theorems—statements in "if so . . . then so" terms.

They ask, for example, in effect, "What is a social action?" Given a certain philosophic orientation, questions like this can be appallingly misunderstood. Traditional philosophy is cluttered with questions of this type where the implicit assumption is that behind each word in use there must be a reality, and that the business of the philosopher is to discover it, so that making a "proper definition" is tantamount to establishing a truth. It has always been the curse of philosophy (until this curse was lifted by the logical positivists) to assume that entities called politics, society, power, welfare, tyranny, democracy, milieu, progress, etc., actually exist, just as cats, icebergs, coffee pots, and grains of wheat exist, and that each has an essence discoverable by proper application of reason and observation. I add observation, because I am speaking not only of the Platonists but also of the Aristotelians.

Now I certainly am not trying to say what is often said in vulgarized versions of the logical positivist position, namely that "concrete" objects certainly exist while "abstractions" don't. A "cat" is no less an abstraction than "progress," when you come to think of it. The problem is not one of existence but one of consensus. Not what *is* a cat, but what easily recognizable objects shall be *called* cats, is the first question. Because agreement is comparatively easy to reach on this question, we can pass immediately to the study of the cats themselves, their "nature," if you wish. But where agreement is not easy, that is, where one cannot immediately agree on an easily recognizable class of events which shall be subsumed under the term "democracy" or "status" or "power," it is futile to pass to the study of these supposed entities.

The systematizers understand in varying degrees the nature of this semantic problem, and they try to come to grips with it. They ask in effect, "What sort of thing shall we *call* a social action?" Consensus is not easy to reach, because the various definitions will presumably have different consequences. "Social action" once defined will presumably be a key term in some social science discipline. It will (hopefully) appear in the theorems of future theory. Therefore its particular definition serves to focus attention on the component events from which the definition is compounded. It may or may not be fruitful to focus attention on this or that combination of events. Hence the problem of definition becomes a "theoretical" problem, something which is often difficult for the natural scientist to recognize.

In the same spirit we can interpret the question "What is a political act?" "What is an economic act?" This search for primary, supposedly elemental, acts is itself inspired by the role of the atom concept in chemistry (as Easton points out). It is not so much a question of whether these "elementary particles" exist: just naming them does not confer existence. It is rather a question of whether our observations can be so organized that the *assumption* that they exist gives us a heuristic and predictive advantage. Incidentally, this is the only sense in which the so called "elementary particles" of physics can be said to exist.

We have, then, so far two distinct meanings of theory. For the natural, especially the physical scientist, theory, as we have said, is a collection of derived theorems tested in the process of predicting events from observed conditions. The physical scientist is able to address himself to problems of this sort, because for him the problems of recognition, of definition, and of meaningful classification either do not exist or have been largely solved. For the social scientist, all too often, the latter kinds of problems are central. The social scientist's aim, therefore, must be lower than that of the physicist.

It is often difficult to concede that one's aims are lower. So there are social scientists who will insist instead that their aims are "different." They are likely to say that they aim at "understanding" events, not at predicting them. For the physical scientist, however, "understanding" is synonymous with prediction. For this reason the physical scientist is likely to become impatient with the social scientist's distinction between the two. Moreover, the physical scientist often associates so called "understanding" (divorced from predictive ability) with ancient philosophical "explanations" which were unsurpassed in their vagueness or in tautological triviality. For this reason, too, the physical scientist often looks upon social science down his nose. And also for this reason, some investigators, usually those with physical science backgrounds, in an attempt to be constructive (that is, willing to extend to social science the power and respectability of the physical sciences) try to take the bull by the horns and to construct mathematical models of social behavior or of historical process wherever quantifiable variables can be found, not bothering too much with the question of whether these variables are germane to the sort of thing social science is trying to do.

The success of these attempts is spotty but by no means negligible. But because it is spotty, and because the relevance of the results to the important questions of social science is uncertain and (let's face it) because many social scientists do not read mathematics and for this reason develop defensive attitudes toward its methods, there is a reaction in the social science camp. This reaction is largely, I think, the source of the disclaimer to the effect that social science is not interested in prediction but only in "understanding." When asked what the proponents of "understanding" mean by it, they are in difficulties. It is as difficult to convey the meaning of "understanding" (in its subjective sense, as it is used here) as it is to convey the meaning of "appreciation" or of "perception." Yet these words are full of "meaning," of sorts. All of us "know" what they mean in the same sense that we "know" what vinegar smells like or how velvet feels. Pressing the issue of "meaning" of understanding is not fair in this situation. But it is fair to raise the question whether it is proper to give the name "science" to an activity which aims only at subjective understanding of this sort.

This is not a rhetorical question. I am not at all sure that the answer is categorically "no," although I suspect that I prefer "no" to "yes." Yet there is no denying that this intuitive organization of perception (akin to appreciation) is an important component in the psychology of science. Without it I doubt whether any but utilitarian motivation would exist for scientific activity, and I doubt whether science could get very far on utilitarian motives alone.

We have, then, a third meaning of "theory" in the attempts of social scientists (these attempts are no longer tolerated in physical science) to achieve and to impart intuitive understanding of social behavior, of the nature of institutions, of political systems, of cultures, and such matters. The language of such theory is largely metaphorical, although a great deal of factual material may be brought "in support." To "support" a theory in this context means at best to marshal factual material (historical and political events, case histories, etc.) in such a way that the reader who views this "evidence" through the metaphors, concepts, and definitions of which the "theory" is constructed will have the experience of "understanding." There is no need to say that even such attempts at concretization are often lacking in the writings of social scientists.

You may gather that as I mention concepts of theory farther and farther from those which enjoy hegemony in the physical science, I am becoming more and more sceptical about the scientific worth of such concepts. To a certain extent this is true, but I do not wish to draw a sharp line anywhere. The "worth" of a theory is not calculable by a set of cut and dried criteria any more than a man's worth as a member of the community is calculable in terms of how much he produces. In particular, metaphor and analogy, although they cannot be accepted as scientific "explanations," are sometimes important aids in the sense that they prepare the mind to make more precise investigations. It is in this sense that the so-called "models" of the non-exact sciences are to be appreciated. They are like the diagrams of geometry, neither necessary nor sufficient for the sort of proof that mathematical rigor demands, but often helpful for the eventual construction of such proofs.

There is also a branch of psychology which partakes in this sort of theorizing, namely, "depth psychology," to which the Freudian system also belongs. This branch of psychology is singularly poor in predictive capacity, either deterministic or statistical. Nor are there many attempts to make its terms operational, similar to the attempts of the systematizers in sociology. The aim is intuitive understanding of what makes up personality (another term which is only intuitively understood). It is strange for me, whose habitat is mathematics, to say this, but I think that depth psychology, particularly the contribution of Freud, is the richest area of behavioral science. I only regret that the disparity between the soft-heads and the hard-heads is so great that it is difficult for them to lay out a common program in which intuitive insights can be translated into strict deductions and verifiable generalizations.

V

There is still a fourth sense in which theory is used, in particular "political theory," namely in the normative, value-laden sense. In this sense, political theory would be concerned, for example, with the question of what is the best form of government. I have been specifically warned to avoid this issue on the gound that too much ink has already been shed over it. To some it seems that concern with what "ought to be" is farthest removed from science, which properly concerns itself with "what is." I will take serious issue with this position.

Whatever I know of the situation in political science is, of course, largely hearsay. In particular I take advantage of learning about the ideas of men whose books I have not read from the very few books on political science which I have read. Thus whatever I know about the ideas of James Bryce I have read in a book by Easton. Easton tells me that Bryce felt that generalization must be firmly rooted in "fact." He also shows how the idea is carried to extremes in some sections of American political science. This "hyperfactualism," as Easton calls it, is, of course, quite understandable. The passage from undisciplined speculation, left over from the times when all science was rooted in philosophy, to militant empiricism has occurred in so many sciences as to suggest the operation of a law. In physical science hyperfactualism died on the vine. It might have proliferated if Francis Bacon's recommendations were ever carried out. But the greatest scientist who was Bacon's contemporary happened to be Galileo Galilei, and he chose a different path. If he had taken "facts" too seriously and too meticulously, he could never have enunciated the general law of falling bodies, because it would never account for the falling of leaves from trees, nor for the fall of rain drops, which between them account for probably 99 per cent of the falling that ever occurs on this planet. Neither leaves nor rain drops follow Galileo's law even approximately. Therefore his law is factually false. But it is true nevertheless, in a deeper sense. Without such ideally true and factually false laws, mathematical physics would have never left the ground.

Galileo's was, in a way, a normative theory. It described not how bodies fell but how they ought to fall under idealized conditions. In this sense one can well see how a theory can be normative and yet truly scientific. The idea of a truly scientific normative theory of action is not to pontificate about morality but to prescribe a correct course of action on the basis of a given desiderata, and in certain (usually idealized) conditions. Such a theory may not have any "practical value," because the idealized conditions may never obtain, but it may have immense heuristic value. In particular, it may through its underlying analysis of the fundamentals of the situation impart to the social scientist just the part of intuitive understanding of an area of investigation that he is seeking.

An example *par excellence* of this sort is game theory, a mathematical structure, which for the most part deals with situations which seem exceedingly remote from the subject matter of social science. However, centuries of scientific experience should have taught us that remoteness of a theory from a particular content area is no indication of its relevance or irrelevance. The fever which derives its name from "bad air" was not really understood until the events in the life of a certain mosquito became known. The harnessing of natural forces owes a tremendous lot to the ancients' curiosity about the anomalous behavior of amber and to one man's logical analysis of an experiment, which had been designed for no other reason than to determine the earth's motion relative to "absolute space."

The relevance of game theory to social science, particularly to political science (although originally the most direct applications were thought to be to economics) resides in the circumstance that game theory distills the logical essence of the situation which Catlin has termed the political act, namely a desire to fulfill an act of will in a context where conflict with others' desires to fulfill their acts of will is to be expected.

Interpreted in a physical context, the metaphor "conflict of forces" calls for some sort of equilibrium theory. Such a theory can be and has been developed purely metaphorically. The concepts of "force," "pressure," "balance of power," "leverage," "stability," "instability," are mostly terms borrowed from physics. Descriptions of conflict situations in these terms sound like descriptions of physical systems. But of course the analogy is a metaphorical, not a logical, one, *i.e.*, the similarity is felt intuitively, not derived as a consequence of an isomorphism between two situations. Therefore metaphorical models of conflict, although they may be valuable for a variety of reasons, cannot be expected to yield logically compelling theorems, let alone theorems translatable into predictions.

There have been attempts to construct real mathematical models of conflict by utilizing the conceptual apparatus of classical mathematics. One such attempt, a very ambitious one, was undertaken by the late Lewis F. Richardson, who cast international rivalries in terms of differential equations and interpreted the stabilities and instabilities of the resulting systems of equations as the stabilities or instabilities of certain international situations. Since Richardson's theory is mathematical, its conclusions are definitive and compelling. In this they have an advantage over the metaphorical theories of the same sort. Its success as a predictive theory is, as would be expected, extremely limited. One set of data was fitted very well by the assumptions of the model, namely the arms expenditures of the two coalitions during the armament race preceding World War I. But even this extremely good fit is not impressive, since there are too many free parameters in the model and the points to be fitted are too few. One could at this point recall our previous argument that a mathematical theory always begins by treating an idealized situation and this beginning serves as a point of departure for greater refinements.

VI

The merits of game theory, however, are of quite another kind. It too deals with idealized situations; in particular it assumes "perfectly rational players." But it departs radically from earlier attempts to cast behavior into models of the mathematical physics type in one very important respect. The older models assumed the metaphysical basis of mechanics. That is, they described systems whose "states" were determined by a causality operating on the "here and now," the way physical states are determined. The state of any physical system is a consequence of the immediately preceding state in accordance with the laws governing infinitesimal changes of state.

In contrast, game theory is primarily a decision theory. It too casts situations into sequences of states. But each successive state is determined by a decision made by a rational being who foresees all possible outcomes and chooses a course of action, which in some way is likely to yield the best outcome under the circumstances. The phrase "under the circumstances" is crucial, for here game theory goes to the real heart of the matter. Each decison maker controls only a part of the situation. In making his decisions, he is aware that other decision makers whose "interests" may be opposed to his also make "rational decisions," and moreover take into account the decisions which *he* is likely to make. No physical theory treats of such situations.

Whereas the mathematical theories of behavior borrowed from the methods of physics (and often chemistry) depend for their success on the special assumptions concerning the interaction of variables and on the possibility of measuring certain key parameters, game theory is largely independent of special assumptions and measurements.

The independence from measurements is achieved by simply by-passing the problem. The only numerical variables are "utilities," *i.e.*, degrees of preference by the several decision makers for the various possible outcomes, and these are simply regarded as given. Independence from special assumptions obtains, because game theory is entirely "normative." It assumes "complete rationality" of the decision makers.

To some workers in game theory, particularly those who are oriented to social science applications, the most interesting results are paradoxically those which show up the inadequacies of game theory, *i.e.*, the indeterminacies of results based on the indeterminacy of the concept of "rationality" in all but the simplest of situations.

To illustrate, let us examine two game-like situations, in the first of which "rationality" can be satisfactorily defined, but not in the second. Suppose two decision makers have two courses of action each. Each of the four pairs of choices leads to an outcome which is denoted by a pair of "pay-offs," that is utilities accruing to each player. The situation can be represented by a 2×2 matrix. The first player chooses one of two rows; the second, one of two columns. The entries in the matrix are the pay-offs to the first and second players respectively, as in Figure 1.

$$\begin{bmatrix} 0, 0 & 3, -3 \\ 1, -1 & 2, -2 \end{bmatrix}$$

FIG. 1

Note that the pay-offs in this case all add up to zero, *i.e.*, what the first player wins, the second loses. These games are called zero-sum.

This game has a "solution" which prescribes a "rational" choice to each player, namely for the first player to choose the second row, and for the second player the first column. This is so, because the first player can obviously guarantee for himself a win of 1, and the second player can *prevent* him from winning more than one, even if each choice is made without knowledge of the other's choice. Here a true "balance of power" exists, but note that this balance is not analogous to any physical balance of "opposing forces." It is a balance based on the *logic* of the situation, involving rational decisions.

The "solution" is much less definite in our next example, shown in Figure 2.

$$\begin{bmatrix} 1, 1 & -2, 2 \\ 2, -2 & -1, -1 \end{bmatrix}$$

FIG. 2

Here the pay-offs do not add to zero. Such a game is called non-zero sum. If we follow the first player's "rational" reasoning, we must concede that he should choose the second row because *no matter which* column the second player chooses, he is better off by choosing the second row. By symmetry, the second column is the second player's "best" choice. But this pair of "best" choices results in a loss to both players $(-1, -1)$, whereas their "worst" choices would have given both a win. Obviously "self interest" is not a self-evident concept.

Agreement between the players to choose first row, first column would be considered "rational" in this case. But admitting agreements of this sort leads to questions of coalition formation, an immensely fertile field in game theory and a very difficult one, because of the ambiguities which plague the concepts of "power," "self-interest," "rationality," etc.—terms taken for granted in much discussion of individual and social behavior, but which game theory has undertaken to define with mathematical precision, and was led into formidable conceptual difficulties in consequence.

The central problem in game theory thus appears to be logical analysis, specifically logical analysis of situations in which common sense often fails. Where such logical analysis can be pushed to a definitive conclusion, the theory of games can be considered a normative theory. For example, the equilibrium solution of a two-person zero-sum game is essentially a prescription to two rational players how to choose their strategies.

But sometimes no such conclusion can be reached on the basis of existing concepts. The following extremely instructive example is taken not from game theory proper but from closely allied investigations. I shall cast it in terms likely to be of interest to political scientists.

Suppose three men, A, B, C, rank three courses of action, X, Y, Z, according to their preference 1, 2, 3. Let their rankings be displayed in the matrix shown in Figure 3.

	X	Y	Z
A	1	2	3
B	2	3	1
C	3	1	2

FIG. 3

What is a fair compromise? Suppose we apply the majority rule to paired comparisons. Noting that X is preferred to Y by A and B (a majority) and Y to Z by A and C (again a majority), we are tempted to assign to X, Y, and Z the ranks 1, 2, 3 respectively. However, this violates the majority rule in the case of X and Z, because Z is preferred to X by B and C (again a majority!).

Kenneth Arrow has shown that such impasses are *certain* to arise wherever more than 2 persons rank more than 2 alternatives; that moreover no decision rule except those which seem undesirable in a democracy, such as dictatorial or entirely arbitrary prescriptions, can be assigned which do not contain inherent contradictions of the sort just noted.

The methods of such analyses, like the methods of game theory, are those of exact science. The situations portrayed certainly remind us of political situations. They involve conflicts of interest, advantages of coalition, social decision rules, in other instances also arbitration schemes, calculation of power indices, etc. Here, then, is an exact science, seemingly applicable to politics (or perhaps, economics). But what does "applicable" mean? If it is taken to mean as involving the possibility of translating the theorems into predictions about human behavior, I am afraid it is stretching the imagination to call game theory "applicable."

The conditions of translatability are not met. Situations are seldom so clear-cut as to be describable in terms of a few "alternatives"; preferences are never so clear as to be measured in utilities; men are seldom rational. In this sense game theory fares no better than the classical mathematical models of behavior borrowed from physics and chemistry.

But in another sense, game theory and similarly oriented investigations are a genuine step forward. For they have burst through the framework of thought imposed by physical science on those who wished to apply mathematical methods to behavioral science. The floodgates have opened and a torrent of entirely new concepts has rushed by. I wish to note in passing that a flourishing terminology is more often a calamity in a rhetorically oriented discipline than an indication of its richness. Vagueness allows undisciplined proliferation and duplication, often providing a mask of erudition for lack of originality and insight. Not so in exact science. There concepts hold their own only if they provide points of anchorage for genuine theorems, not simply for rhetorical speculation.

But what good is it, the empiricist is bound to ask; which is his perfect right. No hide-bound answer can be given. Only if one believes that far reaching and deep-digging logical analysis is essential in any discipline which aspires to the status of science, will one be satisfied with the answer that game theory is an example of such an analysis, that it is of relevance to political science because its fundamental concepts are idealizations of what political science is about, namely decisions made amid partly conflicting and partly coincident interests of rational, calculating beings. Theory of this sort represents one of two poles. The other is meticulous empiricism, of the kind espoused by Bryce, by most historians, and by the positivists of Jerome Frank's persuasion in legal thought. The arguments of the empiricists are well known. They are worthy arguments. The arguments in favor of pure theory are much harder to present because social scientists do not often come into contact with really powerful pure theories, the kind that grow on mathematical soil. One can only say that their worth has been amply demonstrated elsewhere. The physicist might spend thousands of years studying the behavior of ocean waves on a beach in most meticulous detail; in the end he would be no wiser than before with regard to what is essential in wave motion.

The really profound understanding of waves is quite independent of observing any real waves. For that matter, the most important waves in our lives are not even observable directly—I am referring to the waves which underlie our entire telecommunication system and also subatomic events.

It is these kinds of essences which pure theory seeks. It goes without saying that *ultimately* the findings of theory must somehow be translated into real predictions and observations. But to demand this too soon is not wise. It would be like demanding cash payment in every business transaction. Such "hard-headedness," although aimed at security, would actually disrupt the system of credit on which any complex economy must depend to a great degree.

Theory, then, is like a system of credit. One has a right to demand that *somewhere* there are assets to back up the transactions. But, as often as not, these assets may be in the future, and the very act of questioning their existence may set in motion a chain reaction which will preclude their existence.

This is what I meant by the statement that not every conclusion of an exact theory has to be translatable into observation. In the case of the recent constructions of exact theories of presumed relevance to social science, we will do well to extend to them the most liberal credit. After all, in our society the thinker's time comes cheap.