

ANATOL RAPOPORT

## EXPLANATORY POWER AND EXPLANATORY APPEAL OF THEORIES\*

In everyday life, we usually demand explanations only of *unexpected* events. Thus, when I come home from work and find my family at home, I do not ask why they are there. I would ask for an explanation if I found my family absent and the house occupied by strangers. We do not ask why the streets are crowded at 5 PM. We would be inclined to ask for an explanation if we saw a traffic jam at 5 AM. An explanation would satisfy us if we were told of an event as a consequence of which we would *expect* to find heavy traffic at 5 AM. In other words, the acceptance of an explanation of an event depends on our comparing two estimated likelihoods, namely the a priori and the a posteriori probability of the event in question.

In mathematical information theory, this comparison is made precise, and defines the *quantity of information* in a message. Suppose the a priori probability of an event (that is, the probability before the message was sent) is  $p(E)$ , and the a posteriori probability (after the message is received) is  $p(E/m)$ . That is to say,  $p(E/m)$  is the conditional probability of  $E$ , given the receipt of the message. Then the amount of information *about the event* contained in the message is defined (in bits) as

$$I(E;m) \equiv \log_2 \frac{p(E/m)}{p(E)}$$

This quantity may be positive, zero, or negative; and it is instructive to examine the respective conditions. It is positive if  $p(E/m) > p(E)$ ; that is, if the a priori probability of the event is smaller than the a posteriori probability. Indeed, the amount of information increases (though not proportionately) with the ratio  $p(E/m) / p(E)$ . Suppose, for example, we think it is unlikely that a legislator will vote for a certain measure. Suppose, further, that we learn about certain business connections of his, and as a result of this knowledge we make a larger estimate of the likelihood that he will vote for the measure. Then if he does vote for the

measure, we have received a positive amount of information about the way he will vote when we have learned about his connections.

The amount of information about an event in the message is zero if  $p(E|m) = p(E)$ , for in that case the ratio equals 1, and  $\log_2 1 = 0$ . For example, if we are told that the legislator voted for the measure because the vote was called on a Tuesday, and if we fail to see the relevance of this circumstance, we have received no (explanatory) information. In this case, the a posteriori probability of the event was not increased as a consequence of the message.

The amount of information about an event may also be negative. This happens if the a posteriori probability of the event *decreases* with the receipt of the message. This would be the case, for example, if we were told that the passage of the measure would be against the interests of the legislator. If he voted for the measure nevertheless, we have received a negative amount of information about 'his reason for doing so'. It is important to keep in mind that this amount depends on *both* the a priori and on the a posteriori probability of the event. Hence, if the a priori probability is already large, not much information can be given in a message to increase it.

The *explanatory power* of a hypothesis or a theory can be conveniently related to the above mentioned definition of information. The greater the increase in the a posteriori probability of an event (given the assumptions embodied in the theory) relative to the a priori probability, the greater the explanatory power of the theory.

The so-called hypothetico-deductive method of the mathematicized sciences illustrates this principle. A mathematical theory is essentially a device for deducing (by mathematical reasoning) the consequences of certain assumptions. These consequences are generally expressed in quantitative relations. In the ideal case, the deduced mathematical expression predicts the observation of some magnitudes; for example, in astronomy the time of occurrence of the next several solar eclipses. Now, the a priori probability that an eclipse will occur at any *particular* time, is, of course very small. This can be directly tested by asking people ignorant of astronomy to guess when the next eclipse will occur. Accurate guesses will be rare indeed. The a posteriori probability of a definitive prediction is, on the contrary, equal to 1. (If we believe the prediction, we are, by definition, certain that the event will occur.) Therefore the discrepancy between the

a priori and the a posteriori probabilities is very large, and this is the meaning of the assertion that the newtonian theory of celestial mechanics (on the basis of which the prediction is made) is a *powerful* theory.

To take an example of the opposite kind, consider the 'theory' which explains the principles of finding underground water by means of the divining rod. The water-diviner is successful about 70% of the time. This looks like a pretty good score. But to estimate the power of the theory, this score is not enough. We must also know the a priori probability of finding water in an arbitrarily selected location. Some years ago skeptics took the trouble to find out. They selected a large number of sites at random, and found water in about seven cases out of ten. Consequently, the a priori and the a posteriori probabilities of finding water turned out to be equal, and the amount of information in the diviner's message ('Here be water'), hence the explanatory power of his theory, turned out to be zero.

In short, a theory is the more powerful the more successfully it predicts and the less probable (a priori) is the event predicted. We feel this intuitively when we accept an explanation. We find the explanation most satisfying when an event, which we had thought baffling, becomes self-evident after the explanation. This feeling is called the 'Aha!' phenomenon.

This conception of explanatory power is attractive, because it is both intuitively acceptable and rigorous. However, care must be exercised in the way the definition is applied. Suppose, for example, we ask for an explanation of the seasons and get the usual one referring to the inclination of the earth's axis, etc. If we apply our criterion 'mechanically,' we may be led astray. For, although the variation of seasons follows from the explanation, i.e., acquires a posteriori probability equal to 1, the a priori probability of, say, spring following winter is also 1 (since spring is a regularly occurring event). According to our criterion, therefore, the usual explanation of seasonal variation seems to have no power at all. This conclusion is not warranted because we did not consider the entire range of events consistent with the explanation. This inclination of the earth's axis explains not only the seasons but also the apparent path of the sun among the stars. It also predicts the result of an experiment where a sphere revolves on an inclined axis around a source of heat, demonstrating the relation between the inclination of the axis and the amount of heat absorbed as a function of the angular distance from the poles. Hence

the theory explains also the climatic zones. Therefore it is not necessarily the predictive power of a theory in a single instance which makes the theory acceptable, but rather its 'integrative potential', the extent to which many apparently unrelated events are seen in the light of the theory to be related. We shall call this aspect of a theory its *explanatory appeal*.

Note that both explanatory power and explanatory appeal have to do with the degree to which, in the light of the theory, unexpected events become expected ones. The two criteria differ in that explanatory power can be made an *objective* criterion if the a priori and the a posteriori probabilities of an event can be objectively estimated. Explanatory appeal, on the other hand, remains a subjective criterion, rooted in the feeling we have that we have 'understood' an event.

Explanatory power and explanatory appeal are valued in different degrees by people with different epistemological outlooks. Specifically, logical positivists and 'hard' scientists put a value only on explanatory power. Social scientists of the humanist persuasion, on the other hand, place much more value on explanatory appeal. Myths and superstitions are essentially theories devoid of explanatory power but with explanatory appeal in the cultural settings where they occur. To avoid misconception, I should like to emphasize that the coincidence of the criteria used in accepting theories by the humanist social scientists and by the superstitious and naive in no way puts the former in the category of the latter, nor the latter in the category of the former.

A superstition is an imagined causal link between events, for example, between seeing a black cat cross one's path and a mishap. One who believes in the link 'expects' the mishaps in the sense that he is not surprised when they occur. In his estimation, therefore, the probability of a mishap increases following the encounter with a black cat. The formal criterion of 'explanatory power' seems to be satisfied by the superstition. However, the explanatory power of this 'theory' is illusory, as can be demonstrated by comparing the estimated probabilities of mishaps following and not following encounters with black cats with the actual frequencies of such mishaps.

Myths are essentially analogies. Thunder used to be explained by the peasants of Eastern Europe as the rumbling of Elijah's chariot, because the sound resembled the noise made by wagons on bumpy roads. In this way a connection was established (in the minds of the peasants) between

a mysterious phenomenon and a familiar one. Political myths serve a similar purpose. When industrial setbacks in Stalin's Russia were explained in terms of sabotage by traitors and foreign agents, bewildering or threatening events were reduced to 'expected ones'. The same purpose is achieved when our Secretary of State explains social upheavals in impoverished countries in terms of a world communist conspiracy. Once the explanation is accepted, when an event of that sort occurs, one can say to oneself, 'Of course! There it is again.' The *familiarity* of the presumed cause (because it was given a name) obviates the necessity to search further.

The explanatory appeal of naming is rooted in the cognitive function of language. We read in Genesis "...and whatsoever the man would call every living creature, that was to be the name thereof." The new-born man is shown acquiring 'knowledge' about the world he was put in by simply naming the things around him. This conception of knowledge was satirized by Molière in *Le Malade Imaginaire*, where a physician explains that opium puts people to sleep 'because of its dormative property'. It is embodied in word magic to the extent that among some people knowledge of a person's name is believed to confer power over that person.

Strictly speaking, knowledge of names confers no predictive power. The reason naming has explanatory appeal is that it facilitates *recognition*, which is a transformation of something unexpected into something expected. Classification (the naming of names) facilitates recognition still further. For this reason, theories which are little more than systems of classification (hence devoid of explanatory power in the sense here defined) often have great explanatory appeal. We feel we understand the world if we can put our experiences into proper pigeon holes.

In the disciplines where I come from (mathematics and mathematicized sciences), a system of classification is not considered to be a theory. Naming and classification (i.e., defining terms) is seen at most as laying a groundwork for a theory, creating a language in which to develop the logical structure of a theory. As for the theory itself, the meaning of the term in the 'hard' sciences derives from its etymology. A theory is a collection of theorems, assertions derived from postulates by mathematical deduction. The truth of a theory is contained entirely in its predictive power, i.e., to the extent to which the derived assertions are verified by observations.

Critics of the social sciences, especially those of logical positivist persuasion, are wont to demand that 'theory' in the social sciences should have the same meaning. In my opinion, this is impractical for two reasons. First, theories with genuine explanatory (i.e., predictive) power are hard to come by in the present stage of the social sciences. Hence, if theories with predictive power are demanded of the social scientist, and he cannot produce them, he may turn away from theory altogether. An anti-theoretical attitude inhibits the development of a science. Second, naming and classification in the social sciences have much more far-reaching consequences than in the 'hard' sciences. Therefore systems of classification, so-called concept-generating theories, which pervade the social sciences and which are valued mainly for their explanatory appeal cannot be summarily dismissed, as they would be in the context of a 'hard' science. At the same time, reliance on theories characterized by only explanatory appeal is fraught with serious danger.

We shall return to this point. For the moment let us take a closer look at theories with explanatory power, as they have been developed in the 'hard' experimental sciences where the hypothetico-deductive method has come fully into play. Here the conditions under which observations are made are *manipulated* (in controlled experiments). The sciences which make use of controlled experiments are the 'hardest' because they offer the opportunity of predicting events never before observed; that is, events with very small a priori probabilities of occurrence. Note that according to this definition, celestial mechanics is a science of this sort. This may seem strange at first thought, since the heavenly bodies cannot be manipulated. However, the instruments of observation can. Thus the discovery of Neptune, following its prediction on the basis of calculations, was, in effect, a controlled experiment. The 'experimental set-up' was the aiming of the Berlin Observatory telescope at the point in the sky indicated by the calculations. The outcome of the experiment might have been 'planet observed' or 'no planet observed'. Since the a priori probability of the positive result was extremely small, and the prediction was definitive, the result established the truly astonishing explanatory power of the newtonian theory.

Repeated successes of a theory with high explanatory power eventually endow it with explanatory appeal, even if the theory lacked such appeal when it was first proposed. Usually this happens as a consequence of a

conceptual reorganization, a replacement of one framework of thought with another. Newtonian celestial mechanics at first lacked explanatory appeal, because the idea that a force (gravity) could act across 'empty space' was foreign to experience. People were accustomed to think of a 'force' exerted by one body on another only when the bodies were in contact. However, the explanatory power of newtonian mechanics was so great that eventually physicists accepted the notion of action at a distance in their conceptual scheme, and the theory acquired also explanatory appeal. There were several instances of this sort. Clerk Maxwell's equations governing the propagation of electro-magnetic waves had explanatory power but lacked explanatory appeal, because physicists were accustomed to explain phenomena by 'mechanical models'. Maxwell attempted for a long time (in vain) to 'explain' his equations in terms of a mechanical model. Eventually the conceptual framework of the physicists was enlarged, and the need to 'understand' all physical phenomena in terms of mechanical models was dissipated. Again, physicists at first found it difficult to accept both relativity theory and quantum mechanics, the former because it violated the intuitive notion that a time interval had an 'objective' meaning, that is, a value independent of the frame of reference of the observer; the latter because it violated the notion (derived from classical mathematical physics) that physical events obeyed deterministic laws of causality. However, the explanatory power of those theories forced a reorganization of basic concepts and they eventually lost their 'bizarre' character. In the new framework the events explained by the theories became 'understandable' as well as predictable.

The phenomenal success of physical science has been attributed (justly, I believe) to the physicists' preference of explanatory power over explanatory appeal as a criterion for accepting a theory. In the physical sciences verification of predictions is 'the final court of appeal,' as it were. Whenever the derived consequences of a theory result in predictions that consistently fail to be corroborated, the theory is modified or discarded. Thus a 'feedback circuit' is established between theory and experiment, between logical deduction and observed fact. The process propels theories toward greater generality and precision. When it becomes clear that even modifications of existing theories will not eliminate the discrepancies, that is, when what is observed seems to violate the very logical structure of the entire conceptual scheme (the 'paradigm', as Thomas Kuhn called it), a

scientific revolution occurs: the paradigm is replaced by another. As a result, scientists acquire a new insight into the nature of their world, and a period of extremely rapid theoretical development is initiated.

Because of this history of unqualified successes and (let's face it) because science now enjoys unprecedented prestige as a source of power, there has been considerable longing among the social scientists to 'harden' social science, to make social science, like physical science, a source of genuine, progressive enlightenment or, in the thinking of some, a source of power. Thus there has been a transplantation into social science of the 'hard' science paradigm – the hypothetico-deductive method. The fundamental instrument of this approach is the *model*. Here a word of explanation is in order, because outside the physical sciences the term 'model' is used in two different ways.

Sometimes 'model' is used to denote a representation that has only explanatory appeal. Such, for example, are the 'models of the mind' used by psychoanalysts. They are little more than diagrammatic representations of what are purported to be the components of the psyche; for example, the id, the ego, and the super-ego. Such 'models' are sometimes represented as areas with boundaries and arrows between them to indicate 'flow of energy' or of 'control'. These diagrams are essentially analogies or metaphors. Their purpose is to induce certain mental images, so that if one thinks in terms of these images one gains an 'understanding' of how the psyche operates. Descriptive models of this sort are not confined to abstruse matters. A model of the organizational structure of an institution may well represent components and relations which have counterparts in reality, for example, lines of authority and communication. A geographical map is also a model in this sense. Such purely descriptive models help us visualize the object of investigation, either real or imagined. They have, however, no substantial predictive power.

The models used in the 'hard' sciences are different. A model of this sort is essentially a set of assumptions. In particular, a mathematical model is a set of postulated relations among variables. From these assumptions other relations are deduced mathematically. The point of the matter is that, while it may be impossible to verify the original assumptions, it is possible to verify (or to refute) the *consequences* of these assumptions.

For example, one model of population growth states that the rate of change of a population is a sum of two terms, one positively proportional



to the size of the population already attained (the Malthusian factor), the other negatively proportional to the square of the population (the crowding factor). As stated, the assumption is difficult or impossible to verify. But the mathematical consequence of the assumption is that the growth of the population will follow a so-called logistic curve. This consequence can be easily tested, for example, in a bacterial culture.

If observations fail to corroborate the consequences, the model is thereby refuted. But if the consequences are corroborated, the model is not thereby 'proved' to be a representation of reality; it only acquires more credence. We can continue to use it, drawing *additional* consequences from it (if we can), thus putting it to more and more severe tests, perhaps modifying or generalizing it in the process. This is the 'feedback circuit' mentioned earlier.

How are the assumptions chosen? They may be the result of 'operationalizing' certain hypotheses already current. Usually hypotheses proposed by social scientists are not stated in a way that permits us to draw consequences which can be subjected to unambivalent empirical tests. The first task of the model builder is to re-state the hypothesis so that at least its derived consequences are testable.

We shall examine two examples of mathematical models inspired by hypotheses about social behavior. The first was proposed by William Riker, the second by Lewis F. Richardson. Both seem pertinent to political science.

The starting ('raw') hypothesis in the first model (Riker's) asserts that 'in a political setting, people seek to increase their power'. As stated, the hypothesis cannot be subjected to an unambivalent test. Clearly, anecdotal material about individuals who acted thus and so, presumably in order to increase their power, will not do. We have no way of knowing whether the illustrative incidents were or were not selected because of their salience, no idea of how many counter-examples there may be, etc. In order to state the assertion as a testable hypothesis we need an operational definition of power, actually a measure, because the assertion mentions 'increase of power'. An operational definition is one which describes *operations to be performed on observables* to determine what is defined. Many convincing *sounding* definitions of power do not satisfy this criterion. For example, it has been suggested that the power which one individual wields over another is proportional to the probability with which the latter complies

with the demands of the former. Definitions of this sort are abundant in any textbook on a social science discipline. I am afraid that, in most of these statements, elegance of grammar is mistaken for precision. There is an unfortunate tendency among social scientists to engage in polemics about definitions, the arguments being based on the extent to which this or that definition captures the 'essentials' of the thing defined. For example, objections to the definition of power just mentioned may be made to the effect that it neglects the nature and the frequency of the demands. Specifically, the first party may be careful to demand only what he is reasonably sure to get from the second party, and so the degree of compliance does not properly represent his 'power'. Again, the setting in which demands are made may not be one in which considerations of power are relevant. For example, a passenger may never have had a request refused by an airline hostess, but he would be foolhardy to conclude that he had complete power over airline hostesses, etc.

The point I wish to make is that arguments of this sort are vacuous. Before we can say anything at all about the usefulness of the definition offered, we must have an idea of how to determine the key criterion, in this case the 'probability of compliance'. What is the universe of demands? Clearly, 'probability of compliance' (if it can ever be estimated) will vary widely from one setting to another even with the same two individuals. And even if such numbers were somehow computed, what is one to do with them?

Any definition of power to be of service in a testable hypothesis must relate to a specific setting in which systematic observations can be made. There are such settings, for example, legislatures. In fact, legislatures lend themselves very well for the sort of observations that can be fitted into models with predictive content. They contain aggregates of individuals large enough to yield statistical indices (typically some hundreds), yet not too large to preclude examining the whole aggregate. Next, legislators engage in a formal procedure which can be reasonably interpreted as that of pitting the power of one group against that of another. Finally, accurate records of these activities are readily available.

Now, there exists a strictly formalized model of conflict known as the mathematical theory of games. In particular, the branch of the theory dealing with 'games' involving more than two players (so-called N-person games) devotes much attention to the strategic aspects of coalition for-

mation. It must be stressed that 'strategy' in this context is completely stripped of all its content except what pertains to the exercise of power as defined in a specific context. For example, in legislatures the coalition which can pass or defeat a measure gives evidence that the coalition is at that time in possession of power. Here we can define a unit of power in its simplest context, namely what it takes to pass a measure. In other words, there is a 'unit of power' that the legislature as a whole possesses. The problem is to estimate how this unit is distributed among the members of the legislature. Note that the formal nature of the definition requires that the criteria of power distribution be equally formal. This means that only what is included in the definition of power should be included in the criterion of power distribution.

A definition of the 'amount of power' accruing to a member of a legislature (or a coalition) has been proposed by L. S. Shapley in the context of N-person game theory. Suppose the voting units of a legislature have different 'weights'. This can be easily imagined if political parties vote in blocs: each bloc has a weight proportional to the number of votes in it. One way of defining the power of a bloc is as a number proportional to its weight. It turns out, however, that another somewhat more involved definition captures more of what in politics is intuitively understood by 'power'. Assume that every issue is decided by a roll call vote, and that the order in which the members (or blocs) are called is random. Then every member (or bloc) has a certain chance of casting the *decisive* vote, that is, the vote which gives the majority to one or the other side of the issue. A member or a bloc in this position is called the *pivot*. For example, at the 1964 Republican nominating convention, South Carolina was the pivot when her delegation cast the decisive vote for Goldwater.

Naturally, the more votes a bloc has the greater the probability of its being a pivot; but this probability is not necessarily proportional to the number of votes. To illustrate, imagine a legislature with five blocs, having 11, 6, 6, 5, and 3 votes respectively. If a measure passes by majority vote, it is clear that the bloc with 3 votes can never be the pivot, since, in order for this to happen, some combination of blocs must have 13, 14, or 15 votes, whereas no combination in the example cited has any of these. Therefore the bloc with 3 votes has no power at all in spite of the fact that it controls almost 10% of the votes.

An opposite situation is illustrated by a fictitious legislature (or a stock

company) where three blocs control 50, 49, and 1 vote respectively. It is easily verified that, in the six possible orders of voting, the 1 vote bloc is the pivot once, the 49 vote bloc once, and the 50 vote bloc four times. The distribution of power, then, according to Shapley's criterion is in proportion  $4 - 1 - 1$ , which is not reflected in the number of votes.

It is apparent that this definition of power is a generalization of the 'swing vote' concept. Those who have the swing vote hold the 'balance of power', which in the case of three blocs makes itself felt especially strongly if the other two blocs always find themselves on opposite sides. Shapley's 'power' extends this notion to any number of blocs (or parties).

Having thus established a precise operational definition of power, let us return to the original quasi-hypothesis, namely that in political contexts people act so as to increase their power. We find that we can now turn the assertion into a genuine hypothesis if we can find a situation where people act so as to change the amount of power accruing to them. Such a situation presents itself in the French National Assembly, which William Riker studied with a view of testing the hypothesis.

As is known, the French National Assembly consists of several voting blocs (parties). The power of each bloc can be calculated in the manner indicated. In the absence of data to the contrary, we assume that the power *within* each bloc is apportioned equally among its members. The circumstance that makes the French National Assembly suitable for an investigation of this sort is that migrations of members from one party to another are not infrequent. In Riker's study, 61 members of the Assembly changed their party affiliation. It is, of course, natural to postulate some ideological factors underlying such changes of party affiliation. However, the proposed hypothesis is silent on this matter. We are concerned *only* with the power (as defined by Shapley's index) that accrues to a member of the Assembly before and after he migrates from one party to another. If we calculate this change for the several migrations, will the average change be positive? Will it be statistically significant; i.e., will it be difficult to ascribe this result to chance alone? If so, we have found a *factor* in the phenomenon of party affiliation changes.

From Riker's study it appears that the power of the migrators was not increased on the average by the migrations, and that consequently the hypothesis was not corroborated. However, it also turned out that the

migrators were predominantly those members who initially had less power (as here defined) than the average amount accruing to a member of the entire Assembly. It is therefore still possible to conjecture that an intuitive feeling of 'power deprivation' motivated the members to change party affiliation, although they did not on the average increase their power by migrating.

Our next example will be Lewis F. Richardson's mathematical model of an arms race between two rival blocs. Again, the nature of the model demands that the situation examined be stripped of all content except what pertains directly to the variables singled out for study.

The variables chosen by Richardson are 'amounts of hostility' which each of the blocs manifests toward the other. The ground rule in constructing mathematical models is that the variables (at least those to be used in the test of the model) be operationally defined. Therefore some index must be selected to justify an operational definition. Usually such selection requires that we hold in abeyance all our misgivings about the 'realism' of the index, that is, about whether it 'really' captures the essentials of what is defined. Richardson's indices of hostility are essentially the military budgets of the rival blocs. The model states that the rate of change of each budget is a sum of two terms, one positively proportional to the *other's* armament budget, the other negatively proportional to one's own. That is to say, each bloc is stimulated to increase its military budget by the level of the military budget of the other, while its own military budget serves as an inhibiting factor (either because of the economic burden or because the threat is seen in terms of the *difference* between the levels of armaments).

The solution of such a system of equations is in the form of two 'trajectories', that is, the time courses of the armament budgets of the two states. These trajectories are essentially expressions for the military budgets as mathematical functions of time. A trajectory of the combined military budgets of the two blocs derived from Richardson's model was compared with the actual course of the combined budgets of the Entente and the Central Powers in the period 1908-14. The agreement between the predicted (actually 'post-dicted') and the actually observed course was almost perfect.

Here, then, are two examples of a 'transplant operation'. The 'heart' of mathematicized science (the hypothetico-deductive method) was trans-

planted into social science. Will the transplant 'take' or will it be 'rejected'? Let us examine some pros and cons, taking the criticisms first.

(1) It is maintained that mathematical models are useless in social science because they are too simple to capture the intricacies of human motivation and behavior.

It is certainly true that mathematical models are extremely drastic simplifications of reality; but this property is not confined to mathematical models of human behavior. The mathematical theories used in the physical sciences are also often gross simplifications; nevertheless the success of mathematical theories in the physical sciences cannot be disputed.

(2) It is maintained that mathematical models seldom come to grips with the essential determinants of human behavior, in particular social processes. Some have argued, for example, that the arms race of 1908–14 was only a symptom of the rising tensions in Europe; that the real causes of the war should be sought in national rivalries, old grievances, economic interests, intrigue in the chancelleries, etc. Similarly, the game-theoretic model of 'political migrations' seems fatuous. Changes of political affiliation, it is said, are consequences of ideological considerations, possibly of the formation and dissolution of personal contacts, behind-the-scenes bargains among politicians, etc.

Again there is no denying that all of these matters play a role. But just listing them gives us at best a 'feeling of recognition'; it does not help build a systematic science. It must be kept in mind that a model, even though it may have explanatory power, is not an 'explanation'. It is rather a *question* put to nature of the following sort: Is an observed phenomenon a strict logical consequence of *this* particular hypothesis? If the answer is no, the hypothesis must be discarded; it does not qualify as an explanation. If the answer is yes (with a reasonable degree of certainty), the hypothesis *may be an* explanation of the phenomenon. The corroboration of a hypothesis does not prove it to be true and does not establish it as *the* explanation. Rather it tends to open up further questions to be investigated in a similar manner; and this permits the *building up* of a systematic science.

(3) Mathematics is precise and definitive. Predictions of human behavior derived from mathematical formulae imply the assumption that human behavior is determined, whereas it is often a consequence of free choice, which is by definition unpredictable.

This argument is metaphysical and so beyond the scope of scientific discussion. Besides, it manifests a misconception of mathematics. Mathematical prediction can be probabilistic as well as deterministic. Also there is good evidence that, although the individual may exercise 'free will', large masses of individuals (with which social science is typically concerned) often exhibit impressive regularity in their behavior. If this were not the case, insurance companies, traffic control systems, and businesses dependent on more or less stable markets could not operate. Actually, even the practically perfect determinism of physical laws is now known to be a consequence of the fact that gross bodies (to which classical physical laws apply) are immensely large aggregates of particles which individually are not subject to the same (deterministic) laws.

Even on the metaphysical level, the reality of 'free will' is open to question. There is much wisdom in Schopenhauer's remark "Man can do what he will, but he cannot will what he will." Freud's more recent exploration of this theme should not be ignored.

So much for the arguments that, in my opinion, can be easily refuted.

The really serious arguments against 'hardening' the social sciences are of a different sort.

In selecting a research problem to be treated by 'hard', especially mathematical, methods, the social scientist is naturally motivated to choose *tractable* problems; that is, those that lend themselves to precise formulation, in particular to mathematization. Now, the physical sciences owe their success to the circumstance that the tractable problems were also the fundamental problems. The simplest physical experiments (measuring velocities of falling bodies, the expansion of a heated gas, the deflection of a magnetic needle in the presence of a current, the course of a simple chemical reaction) revealed the fundamental laws governing the behavior of inert matter. On this foundation the entire edifice of physical science was systematically constructed. These simple but fundamental experiments do not seem to have analogues in the realm of human behavior. Psychophysics (the 'hardest' part of psychology, where simple definitive experiments can be performed and good mathematical theories can be used) has so far told us next to nothing about the human psyche. Small group behavior (which in certain of its aspects can be investigated by 'hard' methods) has so far told us very little about the nature of social classes, political systems, or international relations. As for mass behavior,

'hard' methods can be applied quite easily in this area (because of the relatively deterministic behavior of large aggregates). But the more tractable problems are of questionable relevance to what the critics of positivism consider to be of fundamental importance in social science.

C. Wright Mills was especially severe in his criticism of American sociology (in which he could have included political science). Although Mills' critique was directed more at empirical sociologists than at model builders, it applies to the latter as well, because the sort of data which are likely to be of immediate use in tractable, testable models would also likely be of the sort that Mills considered trivial. In Mills' view, sociology became trivialized because sociologists abandoned substantive sociological theory in favor of indiscriminate data collection and routine statistical testing of arbitrarily chosen, disconnected 'hypotheses', that is, essentially 'crank-turning' research. Worse, Mills believed that American sociologists, because of pressures of conformity acting on them, were deliberately or unconsciously *avoiding* the really important sociological problems, such as the investigation of power relations in American society.

Implied in Mills' criticism is an attack on pseudo-objectivity in social science. Objectivity is, of course, essential in the pursuit of any scientific investigation. By definition, such an investigation must be directed toward the discovery of truth and toward accepting whatever truth may be discovered regardless of whether it is palatable or bitter and whether it fits or does not fit into our pre-conceptions. However, this stance can be assumed only *after* a problem has been selected and formulated. There can be no objectivity in *selecting* a course of investigation. Such choice must necessarily be guided by one's interests and one's estimates of the 'payoffs' to be gained from the solution of the problem. (The payoffs need not, of course, be material gains. They can be status awards, or, in a purely scientific context, advances in a particular field.) Pseudo-objectivity, then, is an attitude that denies the importance of these subjective factors in the *choice* of problems for investigation. In particular, the pressures to select problems that can be solved by routine procedures (and those that do not come to grips with sensitive issues) are not recognized, and social science drifts along lines of least resistance, both methodologically and politically.

The danger, then, is that routine, massive reliance on quantitative methods makes for mediocrity in social science, regardless of how sophisticated the methods may be and how meticulous may be the observations



and the data processing. The humanistic social scientists seem to feel this strongly, and they see their role as that of counter-acting this tendency by developing 'insight theories' aimed at providing enlightenment instead of exercise in the use of manipulative techniques.

In my opinion, neither the hard-liners nor the soft-liners have a monopoly on mediocrity; nor do the ones or the other have a monopoly on creativity. Ideally, speculative concept-generating theories and rigorous, technically competent hypothesis-testing theories should complement each other in the development of social science. I do not join the positivists who demand that every theory must be immediately translatable into testable hypotheses, and that every hypothesis is formulated only to be tested and discarded if found wanting. This would be like demanding that all commercial transactions be conducted on a strict cash basis (actually gold, to pursue the analogy to its conclusion). Concept-generating theories are essential in social science, because social science simply does not have a catalogue of ready-made concepts with which the physical scientist operates in full confidence that his concepts reflect the essential features of the world with which he is concerned.

A moratorium on concept-generating speculation would leave us stuck with the concepts we have; and it is highly unlikely that the most sophisticated mathematical techniques would get us much farther if the models to which they were applied were built from the same old concepts. Speculative theories, therefore, ought to provide the *raw materials* for the 'hard' scientist to operationalize and to use in building theories with explanatory power in the hard sense.

The concepts of political science are of crucial importance in this regard. Today we find political scientists working in advisory capacities in institutions where decisions are made affecting the lives (and deaths) of millions of people over the face of the globe. Indeed, the life of mankind itself may depend on these decisions. The advisors, consultants, and 'intellectuals-in-residence' are expected by those who hire them to assume a stance of 'realism'. Unfortunately, this stance is all too often tacitly identified with conventional wisdom, especially in the conceptualization of international relations.

I am totally in accord with Mills' critique of American sociology (and by implication of American political science). The issue, however, as I see it, is not between the hard-liners and the soft-liners, or between theoreticians

and empiricists, but rather between those who coast along and those who actively seek new conceptualizations. The crucial difference between natural and social sciences must be constantly kept in mind. In the former, how we conceptualize the world makes a difference in what we can learn about the world but does not directly affect the object of our investigation. In the social sciences, on the contrary, the object of study is ourselves. A change in our conceptions changes us, the objects of investigation. Hence, when a political scientist, while admitting that the *realpolitik* conception of international relations is regrettable, nevertheless accepts it, because, he says, he must see the world 'objectively, as it is', he states only a part of the truth. He omits the important part, namely that believing *realpolitik* to be the content of international relations *makes* it so. This idea is hard to accept, because the success of science has so far depended on the separation of fact from value, belief, and pre-conception. Our values, beliefs, and pre-conceptions, however, *are* the facts of our social life.

At the present time, the United States is seen by at least one half the people on this planet as a formidable obstacle to their aspirations, at times a threat to their very existence. No doubt, this is a feeling engendered by their conceptions of the United States and of its intentions. It is equally true that most people in the United States consider at least one third of the world and the ideas prevailing therein as a threat to *their* aspirations, at times to their existence. This feeling is also engendered in the conceptions of history, of man, and of society prevailing in the United States. It makes no sense to ask whether the threats are 'real'. Perceptions of threats *make* them real. The more 'real' they become the more 'realistic' becomes the stance which takes them into account. And since a 'realistic' foreign policy (at least as the United States power elite perceives it) equates the security of a nation with a preponderance of destructive power and with being willing, able, and ready to use it, the perception generates its own reality. The United States *really* becomes a threat to humanity.

The self-realizing nature of threats has been talked about so much (among others, by professional civilian and military strategists) that to bring up the subject is only to belabor the obvious. Nevertheless, the balance of terror persists, and in the last six years has already erupted into mass murder. Every one recognizes the threat inherent in the combination of Clausewitzian perception of international relations and nuclear capa-

bilities, but few social scientists and fewer political scientists do anything about it.

Where then shall we look for a source of a radical change in our perceptions? I think that concept-generating theories with high explanatory appeal can make a significant contribution. Here is where the humanist social scientist can regain the stature of which the onslaughts of the positivists have all but deprived him. From what has been said, it should be clear that social science theories with only explanatory appeal (unlike physical science theories of this sort) are not necessarily vacuous. They are vacuous only if they are no more than re-runs of old mental images. But if they generate new images of the world, they can contribute mightily to the remaking of man-in-society, because man-in-society is, after all, a composite of beliefs.

At the same time, the dangers inherent in theories with purely explanatory appeal should not be minimized. Paranoid delusions have a tremendous 'explanatory appeal': every event, even the most inconsequential, falls into the paranoiac's pattern and so seems to have been expected. Thus every observation 'corroborates' the theory. Examples of this pathological epistemology can be found everywhere, in demagogic politics as well as in caricatures of psychoanalysis.

To guard against 'paranoid degeneration' of theories with large explanatory appeal, the social scientist must arm himself with 'hard' methods. Still, the present weakness of these methods in the social sciences cannot be denied, therefore they must be nurtured. They should not be abandoned just because they have not yet paid off in predictions, the legal tender of 'hard' research. Specifically, the mathematical model in social science should be valued for having 'broken ground' and for its concept-generating potential.

To illustrate, let me return to the two examples discussed above.

Recall that when I mentioned the good agreement between Richardson's 'trajectory' of the 1908-14 arms race and the observed course of the military budgets, I did not say that the model had high explanatory power on that account. In fact, the explanatory power of Richardson's model is quite low. The reason is that the solution of Richardson's equations involves two free parameters, to which numerical values must be assigned before the derived trajectory can be compared with actual observations. The number of 'points' on the trajectory fitted to observations was only four. Had the

number of free parameters been equal to the number of points, the model would have no explanatory power at all, since in that case *any* set of data could be made to agree with the trajectory by assigning proper values to the parameters, and the model would be irrefutable. (In 'hard' science an irrefutable theory is worthless.) The low explanatory power of Richardson's model together with the almost embarrassingly drastic simplifications on which it is based makes the underlying theory of very questionable value to the political scientist seeking to discover the 'true' causes of World War I or, more generally, the dynamics of early twentieth century international relations.

What has been left out, however, is the concept-generating potential of Richardson's model. The 'trajectory' is by no means its most important output. The point is that Richardson's equations are of a type which characterize both stable and unstable physical systems. Indeed, one of Richardson's conclusions was that if the parameters of mutual stimulation and self-inhibition had been assigned different values, the resulting model would have predicted not an escalation but a stabilization of the arms race. Nor is this all. With the same values of the parameters but with different initial conditions, the 'system' (which turned out to be an unstable one) would have gone at an accelerated pace in the *opposite* direction; that is, toward disarmament and increasing cooperation, perhaps a United Europe, instead of toward war. In Richardson's model, inter-bloc trade volumes enter as 'negative hostility'. Examining the actual arms budgets and trade volumes at the start of the arms race in 1908, Richardson concluded purely formally (i.e., as a consequence of the model) that if the inter-bloc trade volume at that time had been just five million pounds sterling larger (or, correspondingly, the arms budgets that much smaller), the system would be below the ignition threshold and the trend would have been reversed.

I would not ask any one to take this conclusion seriously. I would, however, suggest that it is food for thought. There *may* be situations in international politics which are stable or unstable depending on the values of certain parameters. The idea has already motivated students of international relations to try to discover such parameters. The international system *may* be driven by forces impervious to decisions made presumably on the basis of 'rational calculations'. It may be worthwhile to inquire into the nature of such forces. The model calls our attention to these

possibilities. It *illustrates* how a miniature system with *some* of the characteristics of the international system behaves; and it stimulates thought along paths far removed from conventional wisdom.

Similar remarks apply to Riker's model. Its explanatory power is even smaller than Richardson's, since one of the hypotheses was actually discredited. However, the model is thought-provoking. To what extent is the behavior of modern politicians governed by ideological considerations and to what extent by pure power considerations? Most people agree that politics has become progressively less ideological and more like a game played by professionals for concrete stakes (patronage, advancement along the political ladder, etc.). This theory has explanatory appeal (it agrees, for example, with our impressions about American politics). But how can one 'capture' this idea so as to formulate specific hypotheses to be corroborated or refuted? How can one separate ideological from power components as determinants of political behavior? The mathematical model requires the theoretician to use observable indices, and brings the apparatus of mathematical reasoning to bear on logical deduction. It helps to clarify thinking. Thus Riker's model of how French politicians are supposed to behave is *not* to be taken seriously as a theory of how they actually behave. The model is essentially an exercise, an 'etude', if you will. Its most valuable result is that it accustoms *us* to the techniques, reveals to us their potentialities and weaknesses, suggests other approaches; in short, lights the way in our search for new conceptualizations of social behavior. The model, like most of the still primitive mathematical models in social science, has *heuristic* value.

It will be a long time before we can look to the social sciences for the sort of knowledge about ourselves that the natural sciences have provided us about the world outside of ourselves. Indeed, we may never obtain it. We must constantly keep in mind, however, that the most valuable by-product of social science may not be 'objective knowledge' (which has been traditionally thought to be the principal product of pure science), much less the 'power to control' (which has been the principal product of applied science), but rather the transformation of ourselves through a persistent search for self-knowledge. To the seeker of self-knowledge, rewards accrue even if he never reaches the goal.

*Mental Health Research Institute,  
The University of Michigan, and The University of Toronto*

## NOTE

\* This paper was presented at the Conference on Explanatory Theory in Political Science, Department of Government, University of Texas at Austin, Feb. 19-23, 1968.