What Are Mathematical Models of Behavior Models of?[1]

Introduction

The form of the question suggests an answer (and a very short paper). But behavior, even in its most narrow scientific sense let alone in its ordinary, everyday meaning, is a grossly misleading answer. If for no other reason, it is inadequate because the social and behavioral sciences fail in any serious scientific sense to treat the whole range of human behavior and its concomitant emotional states, but more significantly, it is inadequate because our models pertain only to a restricted class of the best formulated areas of these sciences and, even there, they are only partially effective. Models of behavior are our goal, not a claim of accomplishment. So the question is not trivial. This does not mean that it is especially subtle: any specialist can answer it readily for himself, although he may agree that it can be vexing to formulate it in a way that communicates satisfactorily to his less mathematical colleagues. To a degree, the somewhat forbidding and demanding nature of mathematical discourse raises a barrier;

however, I suspect that the main difficulty is not the mathematics, but rather the scattered and none-too-systematic aspects of our present literature. If so, a system of classification may be of some help.

The scheme I use is neither deep nor entirely satisfactory. It fails many of the usual criteria for a good classification: the categories are neither sharply defined, exhaustive, mutually exclusive, nor do they form a simple hierarchy. The best I can claim for it is that the list is short enough to be remembered, that most models seem to fall reasonably comfortably into just one of the categories, and that I have failed to think of a better one. The six main headings are: models of variables (or, perhaps better, of attributes), simple models of phenomena, more complex models of phenomena, models of experiments, models of interactions among individuals, and models of social institutions and mechanisms. Throughout, I shall focus on behavioral and social processes and exclude all purely physiological and biological ones.

Models of Attributes

Physical scientists quickly become uneasy about the behavioral sciences when we fail to answer clearly the question: What are your fundamental variables and how do you measure them? Often this suggests to them that we have none that are uniquely ours (you may recall that an international commission once declared fundamental measurement to be impossible in psychology), in which case our sciences must be some admixture of applied biology, physics, genetics, etc. But such a conclusion flies in the face of common sense: all our talk of intelligence, love, hate, aggression, beauty, power, loudness, brightness, utility, and the like surely is not wholly idle. To deny the existence of such concepts because we cur-

rently do not know how to deal with them systematically is pseudo-scientific arrogance—what we can't formulate now can't be formulated—and to suppose that soon they will be reduced to terms from other sciences is simply scientific naïveté. We must assume that we speak, however imperfectly and vaguely, of something ultimately capable of careful analysis, just as 500 to 1,000 years ago men meant something close to what we now mean by concepts such as force, work, weight, length, heat, etc., even though they lacked satisfactory theories for any of them, could not measure many of them, even approximately, and partially misunderstood all of them.

Aside from subjective scaling in psychophysics (loudness, brightness, etc.), the careful theoretical analyses of utility stemming originally from economics and statistics, and the none-too-satisfactory but widely used attempts to measure abilities and intelligence, psychology as a whole has tried to bypass the problem of analyzing its variables by substituting so-called physical indicators or indices for them. We really wish to control and manipulate hunger, but instead we control and manipulate hours of deprivation. It is doubtful that they are monotonically related, but, what is worse, we tend to drop hunger from our kit of scientific concepts in order to be entirely, if a bit inappropriately, operational. We wish to control and manipulate aversiveness, but instead we control and munipulate milliamperes of current. And on and on. The simple fact is that we do not yet understand the structure of most of the attributes we believe affect and accompany behavior, and so we substitute for them what we hope are partially correlated physical measures that we do understand.

Rather than deny our variables, we must learn how to isolate and purify them, to measure them, and to relate them one to another in systematic theories. One class of behavioral

models is addressed to this task. It is probably the least understood and least popular of our efforts, but it is doubtful if much of great generality is possible in the behavioral sciences until some of these problems are cracked.

The first sign that a serious examination of an attribute has started is the appearance of empirical exchange relations for that attribute. An exchange relation simply tells what may be substituted for what without altering the amount of the attribute exhibited. At the very least, this requires some means to decide whether two entities—stimuli, events, outcomes, etc.—exhibit the attribute to the same degree, and often it is useful to be able to order the entities according to which has more of the attribute. One cannot say a priori how this is to be done; indeed, discovering a qualitative method for ordering them according to the attribute is usually the heart of the problem. At present, psychologists frequently depend upon a subject's judgment: he tells us which outcome he prefers, which tone seems louder to him, which event he believes to be more probable, etc. Whether more refined and stable methods can be evolved remains to be seen.

However the data may be obtained, the model-builder then attempts to isolate properties—consistent patterns that reflect constraints imposed by the subject—that are (approximately) true for all subjects (of some population). If he uncovers enough of these properties, he may then be able to establish some sort of representation (usually in the real numbers or in an Euclidean n-space) that provides a compact summary of the data and a method for making ready deductions about them. The representation is merely a convenience (albeit, a considerable one); the substance of the theory is not the representation, but rather the qualitative properties that have been found in the data.

Two general types of exchange relations have been studied.

In the first, the attribute is manipulated simply by forming combinations of entities where both the entities and their combinations exhibit the attribute; physical mass and length are prototypes. Various consistencies are observed, which in the case of mass may be summarized by the ordinary representation: to each object a number, called its mass, is assigned in such a way that numerical inequality reflects the qualitative ordering determined by, say, an equal-arm pan balance, and the mass of a collection of objects is numerically equal to the sum of their individual masses. Little has been done with this particular theory (of extensive measurement) in psychology, but somewhat similar theories have been proposed for the bisection of pairs of stimuli in psychophysics and for the measurement of utility (HBMP, 1, 6, 19).[2]

In the second type of exchange relation, two or more independent variables each affect the same attribute and the exchange is a statement of how much one of the independent variables has to be altered to compensate for a given change in another. A physical example is momentum (the attribute) and the exchange is between mass and velocity (the independent variables). An economic-psychologic one is utility (the attribute) and the exchange is among amounts of difference commodities or, in expected utility theory, between commodities and the probabilities of receiving them. The utility exchange relation is called an indifference curve. Many exchange relations, usually called equal-name-of-attribute curves, are known from psychology: the exchange between frequency and intensity to maintain equal loudness of pure tones, between amount and delay of reward to main-

tain equal incentive, between hit and false alarm rates to maintain equal detectability (ROC curves), between resistance and voltage to maintain equal aversiveness of shock, etc.

A family of detailed algebraic theories, known collectively as conjoint measurement theories, is being developed as a set of possible formal structures within which to analyze exchange relations of this sort. Perhaps the best known examples are expected utility theory and additive conjoint measurement (HBMP, 19).

Because the subjects' responses often are inconsistent in some way, the algebraic theories are not always easy to apply. To overcome this difficulty, a variety of probabilistic models, more-or-less closely related to the algebraic ones, are also being elaborated (HBMP, 19). Characteristically, they postulate, among the response probabilities, constraints which go beyond those of probability theory itself. Sometimes, although by no means always, a numerical scale similar to those in the algebraic theories is either assumed or deduced, and the probabilities depend upon those scale values in some systematic way.

Simple Models of Phenomena

Aside from attributes which adhere to stimuli, responses, and outcomes, and for which ordering concepts seem appropriate, there are many behavioral events that are best thought of as discrete, qualitative phenomena. A stimulus is or is not detected, a problem is or is not solved, an association is or is not learned, an item is or is not remembered, a decision is or is not taken, etc. To be sure, in many cases some notion of degree may also be appropriate: a stimulus may be detected with some degree of confidence or with some proba-

bility, a problem may be solved partially, an association may be learned gradually or with some probability, an item may be remembered partially, or a decision may be taken with a certain dispatch. Nevertheless, a distinction between attributes and occurrences seems useful, even though any model of a phenomenon, such as for the detection of stimuli or for learning, will usually have a place in it for the attributes of the stimuli, responses, and outcomes that contribute to the phenomenon.

Unlike the models for variables, which merely state low level abstractions from the observable data, those for phenomena are usually cast in terms of hypothetical happenings as well as in terms of observables. They explicitly acknowledge a conceptual structure within the organism—not the moist red and grey structures of the physiologist, but a hypothetical one of representations of stimuli, of storage and decay, of associations and random sampling, of counting mechanisms, and of elementary comparisons and decisions. It is a simple internal world, composed of processes simpler than those built into even the most primitive digital computer, but one whose consequences, which generally are worked out in considerable detail by mathematical methods, often are of the same order of complexity as the experimental data now available.

A few well-known examples serve to illustrate this type of model. Several models for psychophysical detection and recognition (HBMP, 3) postulate that each stimulus leads to an internal representation that we can treat as a random variable (with values in the continuum of real numbers in the Thurstone and closely related signal detectability models and in a discrete set of numbers in the threshold models). The subject is assumed to partition the set of possible values into classes that he makes correspond to possible response alternatives,

and the boundaries of the partition are affected in systematic ways by other experimental variables such as payoffs and presentation probabilities. The random variables and the partitionings are, of course, unobservable.

Stimulus sampling theory of learning (HBMP, 10) supposes that an unobservable set of elements partially reflects something of the stimulus conditions of the experiment, that each element is conditioned in some way to one or another of the possible responses, and that the decision as to which response to make on a given trial is reached by counting in a random sample the number of elements that are conditioned to different responses. Following the response and outcome, the conditioning of some of the elements may be changed. Models of this general type are sometimes called urn schemes, and they have been postulated for a variety of psychological processes.

Lately there has been a resurgence of interest in short term memory which has led to a number of clever experiments and several plausible, if difficult to test, models. Among the hypothetical phenomena embodied in the models are abstract, somewhat imperfect representations of the stimuli, temporary storage during which information about the order of arrival of stimuli may be lost, decay in the representations, and various forms of elementary decision-making.

It is no exaggeration to say that the sizable majority of all psychological models are of this general type. At least four reasons underlie their popularity. First, the phenomena in question are undoubtedly important and have been so thought throughout the history of psychology. Second, considerable data are already available, and well established experimental paradigms make it relatively easy to collect additional data that are pertinent to the models. Third, in spite of their unobservable constructs, these models usually predict

observable, or nearly observable, behavior and so they bear on data quite directly. And fourth, such simple mechanisms are easy to think about and to formulate in familiar and comparatively elementary mathematics, although deriving conclusions from them can sometimes be quite a formidable task.

Such models as these exhibit two major difficulties, neither of which is necessarily insurmountable. First, the "observable" predictions of most are response probabilities that depend upon at least one, and often more, free parameters that must be estimated from data. The data are mostly, but not entirely (e.g., latencies), discrete responses. This incompatibility leads both to complex problems of estimation—often requiring more analysis than the model itself—and to subtle issues of evaluation when we try to decide if the model is adequately consistent with the data. Second, as one passes from models for the simplest experiments—usually those with one or two stimulus conditions and one or two responses—to more complicated ones, the model-builder faces a multitude of decisions about what is supposed to be going on in the "mind" of his hypothetical subject. Since none of these postulated processes can be observed directly and since often it is not readily apparent how different choices affect the predictions, one usually begins to feel as if he has entered into a never-never land of choices. This may lead him to turn experimenter and to undertake an elaborate, time-consuming series of studies designed to ferret out each choice point more or less independently of the others. Although progress can be made along these lines—the work in discrimination learning is a good example—it is usually very painstaking. Perhaps the most vivid example of these conceptual difficulties and the corresponding attempts to isolate different subprocesses is to be found in the current work on short-term memory.

The obvious, and possibly the only, alternative tack is di-

rect physiological observation of what are now hypothetical processes. Were this possible, the development of these models would be greatly accelerated. I doubt, however, that we have much hope of aid from this quarter in the near future. The level of abstraction of the events postulated in the models does not seem to correspond at all well with the phenomena that physiologists are now able to isolate. Of course, physiological psychology has seen a number of impressive developments in the past two decades and is changing rapidly, so perhaps we will be surprised, but it would be unwise to count on it.

To close the section, a few words about the interrelation between models of attributes and phenomena are appropriate. Models of phenomena almost always end up with free numerical parameters that must be estimated from data, and sometimes the parameters seem to have a natural, intuitive interpretation as numerical scales of one sort or another—of stimulus intensity, utility, learning rate, etc. Thus, in principle and occasionally in practice (e.g., signal detectability theory), models of phenomena provide indirect scales of attributes. This fact has not been as effectively exploited as one might have hoped. Such scales, if worked out in detail, should provide significant hints to the measurement theorist about the types of scales needed and their dependence on experimental variables. In the opposite direction, as measurement improves, it should become possible to incorporate direct measurements into the theories of phenomena, thereby reducing or eliminating the free parameters. Nothing along this line has yet been done.

More Complex Models of Phenomena

Although the logical distinction between simple and more complex models of phenomena is fuzzy, two simple opera-

tional criteria suffice at the moment. We say that a model is of the more complex category if either it is too complex to be stated mathematically and is only formulated as a computer program or if it is one that involves several intermediate processes which have been studied in some detail by themselves. A few words about each type will suffice.

In computer simulations of decision-making and learning (HBMP, 7), the "mind" of the hypothetical subject is embodied in a computer program which is analogous to the mathematical models of phenomena, but is vastly more complex. The innumerable choice points of a program, with its possibilities for complex comparisons and detailed combinatorial explorations, far exceed in complexity anything that one would be willing to formulate mathematically. Which choices to make, when so many exist, seems beyond any hope of sensible resolution; yet those who have written such programs seem to hold that this freedom is illusory. One has to be a bit ingenuous to believe this claim which, I suspect, either reflects some over-enthusiasm for the method or a failure to acknowledge the numerous implicit choices that have been made. Some of these choices are probably buried in the programming language that happens to be used. The problems in evaluating these models are at least as severe as for the mathematical ones. Some programs predict specific responses on each trial (of course, it is trivial to modify them so that they do not), in which case we surely do not want to reject a program on the basis of one erroneous prediction. If not one, how many? Since subjects do do different things, how do we alter the program to handle their differences? These and related issues of evaluation have not been dealt with very effectively in the literature.

Our second class of more complex models includes those concerned with specific peripheral sensory processes, mainly in the eye and the ear, for which it is possible to get detailed

physiological, mechanical, and chemical information about various of the steps (HBMP, 15, 16). The models describe the transduction of energy through these organs. An attempt is made to take into account data obtained at each interface, thereby reducing appreciably the number of unguided choices that have to be made. The resulting over-all models, which often are quite complex, are remarkably good in describing even the fine detail of the transduction. As we noted, similar methods are not currently available when we are concerned with phenomena that occur in the central nervous system.

Models of Experiments

Since the models just discussed are of interest only to the extent that they pertain to experiments, they could all be classed as models of experiments. But I don't mean that. Rather, I refer to models that are relatively atheoretic, that apply to just one or to a very limited class of experiments, and that are, to be blunt, little more than an elaborate form of curve fitting. The last charge is not likely to be well received by the authors of such models.

The attempt—admittedly not as successful as one would like—in the models previously discussed is to isolate and describe phenomena that take place within the organism and that in some way constrain his possible behavior. From these postulates we deduce what such a hypothetical organism will do when confronted with the boundary conditions established by this or that particular experiment. The idea is to parallel the approach taken in classical physics in which certain (usually differential) equations describe the constraints that hold among physical variables in all situations, and any particular situation is specified by boundary conditions for the equations. Together, the boundary conditions and the equations lead to specific predictions for the particular situation. Such a

division into theory and boundary conditions has proved an extremely powerful technique since having once evolved the theory a model may, in principle, be constructed for any new situation provided only that it is adequately described. Something like that, feeble though it may be in comparison to what is done in physics, is being attempted, for example, in detection theory, in stimulus-sampling theory, and in the work on memory.

A model of an experiment may be defined negatively: it is one that fails to separate clearly the postulated properties of the organism from the boundary conditions that represent a particular situation (experiment). At present, most if not all of our work fails to some degree to make this separation and so, to that degree, our models are of experiments; but some are considerably more satisfactory than others. Much of the work using linear and nonlinear stochastic operators and Markov chains (HBMP, 9, 10) to analyze simple learning data suffers badly from this failure of separation. If one of these models accounts well for one experiment, we rarely know what to predict about a closely related one: there is insufficient underlying theoretical structure to venture much beyond what we already know.

Atheoretic models may, of course, be extremely valuable in predicting (extrapolating) things of practical importance. The input-output models of various industries and national economies is one example. When, however, the models are for experiments whose only conceivable interest is the possible insights they may give into basic phenomena, their usefulness is less clear.

Models of Interactions among Individuals

Were our models of individuals adequate, models for their interactions would, in principle, be easily constructed. All of the other individuals would form part of the (time-varying)

boundary conditions of any one, and it would be a purely mathematical (or, more likely, computational) problem to deduce predictions. Judging by the troubles physicists have had in solving the equations for small numbers of interacting particles, the working out of these deductions would be dreadfully difficult. Although this approach seems somewhat fanciful at our present level of development, limited examples actually exist: Markov learning models to describe two individuals interacting in a simple game-like situation, several computer simulations of interacting individuals, and game theory in which each individual is assumed to make a rational analysis of the rational behavior of the others which, together with their individual utility functions, sometimes lead to (prescriptive) decisions (HBMP, 14).

Since, however, this approach is not yet suited to the analysis of most small group processes of interest, other authors have attempted to abstract various global aspects of interacting groups and to construct models in these terms. In one type of model (HBMP, 14), the group is assumed to be described in terms of time-varying variables such as pressure to communicate, cohesion, and the like, and certain differential equations are assumed to interrelate the variables. The properties of these equations have a certain intuitive plausibility, but the fact of the matter is that little can be done to test them since no one has the slightest idea how such variables should be measured. In another type of model (HBMP, 14) the time course and many other details of the interaction are abstracted away until all that remains are certain discrete structural links between some pairs of individuals. Depending upon the focus of interest, they may represent lines of authority, possible communication channels, affective relations, etc. The hope is that the structure of these graphs, as they are called, will reveal something of the social psychology involved. Many mathematical properties of these structures are known, and

some of them (e.g., balance) are thought to correspond to socially important notions. It has, however, not proved easy to relate the mathematical definitions and theorems to empirical observations. Some of the difficulty may stem from the static nature of the abstraction, but probably more important is the fact that the abstraction does not really make any explicit assumptions about the participating individuals.

I think that it is safe to say that, so far, models of small group processes have contributed but little to our understanding of these processes. The fault does not, I think, lie with the model-builders, but with the basic intractability of the problem at the present time. On the one hand, there are no remotely adequate models of the individual behaving in a social environment and, on the other hand, there is no real opportunity to aggregate over sufficiently large collections of individuals so that statistical smoothing, as in some economic models, comes to our rescue.

Models of Social Institutions and Mechanisms

Although this category is quite extensive and includes some of the most successful models in the social sciences, I shall deal with it only briefly because the models do not refer to the behavior of individuals as such. Their basic terms are not individuals or their actions, but rather abstractions such as price, quantity, rate of interest, rate of growth, and so on. Of course, it is the actions of people that determine these variables, but the models make no attempt to analyze them from that point of view—and for good reason. The best examples are from economics, but somewhat similar ones are beginning to be developed in parts of sociology (e.g., for voting patterns within large groups of people) and in political science (e.g., some of the work on coalition formation).

I would class as similar the qualitatively quite different

models that use concatenation algebras and the theory of recursive functions to analyze the underlying grammatical structure of languages (HBMP, 11, 12, 13). Here the basic terms are linguistic rather than economic or behavioral. It is possible that such models will come to play a significant role in the development of theories of behavior since universal aspects of languages undoubtedly reflect certain deep-seated human constraints. In particular, the study of language learning and concept formation should be affected.

Concluding Remarks

Of the above categories, only the ones for attributes and phenomena include models directly concerned with the behavior of individuals. Other models may well bear upon behavior, but the core of mathematical psychology is in these two areas. Confining our attention to them, it may be useful to conclude by citing a few of their present failings.

Work in the theory of measurement has not yet begun to make clear how many inherently different variables there are. If one is to judge by the scaling of psychophysical attributes, the number is fantastically large; however, I suspect that this is much like treating the energy of chemical reaction A as a measure distinct from the energy of B. Even though the modalities involved are inherently different, could it be that there is a single notion of subjective intensity of which, for example, loudness and brightness are just two special cases? For this to be possible, a subject should be able to say whether a given sound is more or less intense than a given light, which at first hearing sounds pretty silly. Yet experimentalists using the method of cross-modality matching have found that subjects can do just this and that they do it consistently. So, perhaps, ultimately there will be a grouping of some attributes into a single theory of measurement of subjective intensity.

Similarly, there appears to be a group of attributes that can all be called affective. Already, economists have been willing to group together all sorts of disparate economic goods under the common affective measure called utility. To extend this sort of measurement to other things that we like and dislike in varying degrees, such as shocks and other stimuli used as reinforcers, may be possible. It should be noted that many stimuli can be viewed as possessing both intensity and affect, and so at least two types of measures are associated with them.

A third general class of measurements that seems to hang together are those that may be called predictive and for which some sort of probability concept is the common attribute. Here the unity seems already to have been achieved, and one has no hesitancy in comparing events of quite different types by the same measure.

How many other categories of measurement will be needed is unclear; it is doubtful whether some of the personality concepts, if they can be measured at all, fall among these three. Nevertheless, considerable simplification will have been achieved if it can be shown that many apparently distinct attributes are special cases of a single variable and can be measured on the same scale.

Turning to models of phenomena, two limitations are striking. First, the models are mostly concerned with choices (or decisions) from prescribed sets of (basically, non-verbal) responses, and with the effect of input information, rewards, and risk on these choices. At least part of the reason for this limitation is the use of probability theory which strongly invites the choice formulation, but whatever the reason, the result is that much of importance has escaped the model-builder. We do not make choices every instant. When do we view situations as partitioned into alternatives? Why do we partition them as we do when we do? In part, this must be related to certain perceptual problems of how energy distri-

butions over time are structured into unitary concepts such as people, chairs, etc. How do we create these concepts and how can they be modified? What is the nature and role of various emotional states? How do motor skills develop and what influences their acquisition? And so on. The point is simple; choices, although important, are really only a small part of our behavior and it is unlikely that they will be well understood without at least some understanding of other aspects of behavior.

A second, and related, limitation is that both the models and the experiments involve a vast amount of repetition of the same or nearly the same choice situation, usually structured in the form of discrete trials. The reason for this is that the models talk about probabilities but the experiments provide choice data, and only by making repeated observations can we estimate probabilities from data. Since these experiments are a far cry from any natural environment of the subjects, animal or human, one cannot help being a bit edgy about just what it is that we are studying. If learning requires one hundred or more identical trials, what organism would ever have the opportunity to learn in his usual environment? I believe that it is important that we try to break out of some of these self-imposed limitations, although I confess I do not see clearly how to do it.