

## Some Targets for Mathematical Psychology\*

W. K. ESTES

*The Rockefeller University, New York, New York 10021*

Concern among mathematical psychologists regarding the present status of the discipline arises from two sources. One is the growing disparity between the rapid development of mathematical psychology as an adjunct to research and the slower and more uncertain progress toward a cumulative body of theory. The other is the problem of adjusting to the encroachment of computers and computer simulation models into the traditional province of mathematical methods. It appears that these problems might be alleviated by recognition of the complementary aspects of mathematical and computer simulation approaches to psychological theory and by a shift of emphasis from tactics to strategy in the construction and evaluation of models of both types.

It is just about 25 years since mathematical methods and theory began to spread in a major way from their long established niche in psychological measurement and scaling into broader areas of experimental psychology, and a little over 10 years since the term "mathematical psychology" became current with the appearance almost simultaneously of the *Handbook* (Luce, Bush, & Galanter, 1963) and the first issues of the *Journal*. How can we characterize the current state of the field? In some respects very well. Certainly the volume and variety of research and the level of technical and methodological virtuosity far exceed anything that might have been reasonably anticipated a couple of decades ago. But at the same time it is clear that many investigators in our field are not entirely happy with their current situation.

A rather pointed documentation of this last remark is to be found in the introduction to the just published *Current Developments in Mathematical Psychology* (which was forced into two volumes by the current pace of research): "It is easy to point to excellent work in mathematical psychology, past and present. But in retrospect, cumulative progress is less easy to find ... without apologizing for the past, we do need to ask ourselves whether we can do better in the coming decades. Is it possible that the lack of cumulative progress is partly due to that goal being subordinated to others, such as originality or technical mastery? If so, then that goal needs to be formulated more directly, and seeking it needs to be encouraged" (Krantz, Atkinson, Luce &

\* This paper is based on an address given at the seventh annual Mathematical Psychology Meetings, Ann Arbor, MI, August, 1974.

Suppes, 1974, xiii–xiv). This view is doubtless shared by many investigators in our field, but even if it were not, it would need to be taken seriously since it evidently represents the consensus of four of our most eminent contributors.

A number of comments come to mind concerning this rather sobering self-appraisal. First of all, one might note that if we lack a sense of cumulative progress, the fault may to some unknown extent lie in our ability to appraise progress rather than in the state of the field itself. We are handicapped by not knowing what the end result of cumulative progress is going to be, and thus, we are not in an ideal position to evaluate the rate at which we are approaching this unknown goal. Second, we might note that cumulative progress need not necessarily be continuous. Robinson Crusoe had to fashion himself some tools before he could begin to build a house. In our case we need to accumulate a repertoire of small mathematical systems and devices that we understand before we can hope to proceed far in the building of theories. Third, and perhaps most important, I think that in discussions of this sort it is well to distinguish two aspects of mathematical psychology—development of mathematical techniques and models in conjunction with specific researches and development of mathematical psychology as a distinct specialty. The unevenness of progress in these two aspects may be a major source of some of the pains that give rise to critical self-examinations.

With regard to the first aspect, surely no one could complain about the accomplishments of the last quarter century. During the period of the early 1950's, when presentations of mathematical models were just beginning to appear as regular fare in the *Psychological Review* and the first workshops in mathematical psychology were just getting organized, what most of us wanted more than anything else was acceptance of our small subdiscipline as an integral part of experimental psychology or, more broadly, psychological research. The incorporation of mathematical psychology into various research areas, in fact, has proceeded so fast that already one finds it necessary to plod his way through a dozen journals to bring himself up to date on current mathematically oriented research in any of a large number of specific research topics, ranging from lateral interactions among receptor elements to the scanning of short-term memory, to the verification of propositions.

With regard to the second aspect, the editors of *Contemporary Developments* are not alone in looking for something beyond the ever-changing collection of methods and specific models that arise in particular research areas. It is here that we feel a lack of cumulative progress in any well-defined direction. We can note a number of symptoms of this mild malaise.

First, let us scan the overall picture of the contemporary literature. One is struck first of all by the succession of introductory textbooks with a wide range of special emphases (Atkinson, Bower, & Crothers, 1965; Coombs, Dawes, & Tversky, 1970; Restle & Greeno, 1970; Laming, 1973). It would appear that a good number of people must be learning about our field. But students or outsiders can evidently do so only at a fairly elementary level. There is a fairly steady appearance of advanced volumes

but these are suitable only for specialists in specific areas. Volumes of contemporary developments, not only ours but similar collections appearing in other countries (for example Poland—Kozielecki, 1971) present a wide variety of topics, but again the individual chapters tend to be addressed only to specialists. The conspicuous lack is any body of common problems in which any major fraction of investigators in the field share concern. A sampling of the *Journal of Mathematical Psychology* over its 11-year history yields a similar picture. Articles on measurement hold constant at 10–20% of the total output, contributions to methodology and treatments of rather general processes somewhat less; about 70–75% of the articles each year are devoted to models for specific experimental areas or other treatments of topics that require expert knowledge of the details of a particular research area for comprehension.

Either gaining a foothold on this literature or keeping up with it once aboard is a difficult task even for the specialist. But consider the plight of the outsider. I have rather frequently had occasion to talk to individuals with backgrounds in mathematics, computer science, or the like, who have heard a little about mathematical psychology and would like to participate. We certainly should welcome such potential contributors, but how can we get them started? All too often, one of these individuals attends a meeting or a few seminars and then gives up, realizing that he would need a Ph.D. in some specific research area to follow the conversation, let alone to begin to participate. Typically these people wander off into genetics or economics, two fields that have in common only the characteristic that they present rather well-defined problems for mathematical attack, problems that are comprehensible to one with only a modest acquaintance with the subject matter and research methods.

Finally, some of us are finding occasion to worry a bit about a problem that mathematical psychology shares with all behavioral science disciplines with respect to their potential contributions to problems of social welfare. Many of us have hoped that mathematical psychology would prove a major vehicle for developing theoretical interrelationships between psychology and the various social sciences, thus facilitating both theoretical developments and applications. But as things have actually gone, the flourishing of new mathematical models and methods has profited various specific research areas greatly but has contributed less than we would like toward bridging the gaps between disciplines or mediating applications of social science to social problems.

What we wish for is clear enough. We would like mathematical psychology to continue to aid in the direction and interpretation of various specific lines of research. But at the same time we would like to see more cohesion and more cumulative development of mathematical psychology as a distinct specialty among the behavioral science disciplines. We would like to see it begin to present a common body of methods and problems that are understood and appreciated by nearly everyone in the field. We would like to invite participation from talented outsiders who are better skilled than we in special aspects of mathematics, computer science, or linguistics, but who are

little acquainted with specific research procedures. We would like to see some common body of theory begin to take form at a more abstract level than that involved in the interpretation of memory scanning studies or lateral inhibition in the limulus—theory that might serve to promote useful interconnections between psychology and other social sciences. What could we hope to do that might help some of these wishes begin to come true?

Let me hasten to remark that I am not going to offer any firm prescriptions and most certainly I shall not propose anything of the nature of planning committees, or the turning of our meetings into sessions of agonizing self-appraisal. The substantial progress we can point to in our field, as in other areas of science, is attributable to individual enterprise and evolutionary processes. Thus, I shall suggest only that we might want to look a little harder as individuals for specific things that we might do to put our house in better order. I will offer a few comments under a number of categories.

#### THE DELINEATION OF PROBLEMS

Most of the time we quite naturally work on and report the results of research on relatively specific problems, having faith that in the long run generality will emerge. Faith is well enough, but it can stand supplementation on occasion by more direct efforts. One thought I have on this note is that our journals might well not only accept but even encourage articles that do not report either mathematical contributions or new empirical results, but rather report efforts to prepare problems for mathematical approaches. It is difficult enough for the specialist and impossible for the outsider to make his way through the typical welter of conflicting empirical findings and discover what pattern of quantitative relationships he should be attempting to account for in almost any of our research areas. But often it would be possible for an investigator with competence in a specific area to review the accumulation of studies, abstract the functional relationships that appear to stand up, and present an organized pattern of results that appears to call for mathematical interpretation. A timely example would be the accumulated work on spacing effects in studies of memory that was the topic of a recent symposium.<sup>1</sup> Although there are problems to be resolved at an empirical level with respect to various specific kinds of experiments on spacing, there also appears to be emerging a rather intriguing pattern of relationships which holds over a variety of materials and invites a more general mathematical interpretation (Glenberg, 1974; Murdock, 1974).

<sup>1</sup> Symposium on Spacing Phenomena in Memory, Mathematical Psychology Meetings, 1974. Participants: Robert A. Bjork, Arthur Glenberg, Douglas L. Hintzman, Thomas K. Landauer, Douglas L. Medin.

Reviews of this sort might on occasion lead, if not directly to theoretical efforts, to the carrying out of studies designed simply to provide the substantial body of reliable parametric data required to test and extend models. Referring again to the symposium on spacing effects, it is clear that we now have a major need for one or more studies designed simply to collect a substantial quantity of reliable data with the principal parameters varied within a single experimental situation in the fashion needed to provide adequate tests of theoretical efforts that are already under way. Further, just as is now done routinely with programs in computer science journals, the parametric data could readily be made available to mathematical psychologists. It is now so easy and inexpensive to reproduce and transmit quantities of data to interested users by means of tapes or cards that it seems a pity to continue to require every would-be quantitative theorist to serve also as his own data collector.

#### ON THE EVALUATION OF THEORIES AND MODELS

What problems, some will ask, can you see in the matter of testing mathematical theories? Does not one simply construct a model, apply it to data, and accept or reject on the basis of goodness of fit? Well, that is indeed a standard procedure—perhaps *the* standard procedure. The problem I wish to raise is that the procedure nonetheless may not be good enough.

It is true, up to a point, that models should describe data, and at one time achieving this purpose was no small task. I can remember the excitement when some functions derived from the early learning models of Bush, Mosteller, and myself turned out to provide reasonable descriptions of empirical learning curves for several situations with only one or two fitted parameters. Also, I can remember full summers of effort devoted to working out methods for estimating parameters in learning models, typically running to horrendous algebraic derivations, endless hours at a hand calculator, and whole chapters in books devoted to methods of estimation. All of this will seem quaintly antiquarian to students who have learned mathematical psychology during the last few years, for in major respects our field has been revolutionized by a development none of us could have anticipated 25 years ago, namely, the fantastically accelerated evolution, proliferation, and widespread use of digital computers. I think that learning to live with computers is perhaps the single most difficult and critical task facing mathematical psychology as a discipline.

**THE COMPUTER REVOLUTION.** The problems raised by computers are of two different types, both having to do with the evaluation of models and theories. The first is simply that the matter of fitting models to data has suddenly become so easy that it no longer constitutes a useful method of tracing theoretical progress. Virtually any learning experiment or group of learning experiments, for example, can be

described by a Markov chain model with a sufficient number of states. The fitting of the model to data is accomplished via computer programs with 4, 8, 12—virtually unlimited numbers of parameters being evaluated by computer search of the parameter space—almost completely bypassing both the opportunities for ingenuity and the quantities of blood, sweat, and tears that used to go into this enterprise. It has become almost literally the case that the fitting of a model to a set of data is primarily a test of the technical competence of the investigator and has little to do with the adequacy of the model for interpreting empirical phenomena. I do not propose that we stop fitting models to data; that we will never do. But I do believe that intensive efforts should be directed toward finding additional bases for evaluating quantitative theories.

The second aspect of the computer revolution which has raised new and, to say the least, challenging problems for us is, of course, the advent of computer simulation models—models that can be fitted to data just as readily as the more familiar mathematical models but that have no specifiable mathematical form and for which we are generally unable even to formulate, let alone solve, the problem of testing goodness of fit.

Some enthusiastic proponents of computer simulation models seem to have taken the view that if a program can be written that generates protocols closely mimicking those generated by actual experimental subjects, the problem of goodness of fit is solved and the program must be accepted as an adequate theory for the phenomenon. But despite the persuasiveness with which they are advanced, these arguments are not going to continue to satisfy us for long. If these premises had been accepted and the technology had been available, computer simulation models for the motions of the planets based on the conception of epicycles, or for the process of combustion in terms of phlogiston could never have been rejected. In learning to live with computer simulation models, it may be even more important than in the case of conventional mathematical models to find ways of augmenting the traditional conception of goodness of fit as a means of judging theoretical progress.

**PREDICTIVE POWER OF MODELS.** It is easier to pose the problem than to solve it, but I do think we can perceive some directions in which inquiry might be useful. The first suggestion is more serious exploration of an information-theoretic approach to model testing as distinguished from the conventional one based on statistical tests and goodness of fit. This approach has been outlined in an article by Hanna (1969) in which he brings out a number of interesting points, for example, that statistically rejected models may prove more adequate in the sense of providing information about the empirical variables influencing behavior, than statistically accepted models. The coefficient of predictive power of a model relative to a set of data that Hanna proposes is not ready for ubiquitous everyday use in that it can be directly applied only in situations where one can compute likelihood functions. Nonetheless, it might be useful to begin thinking in terms of predictive power as a complement to goodness of fit.

**SEQUENTIAL MODEL-FITTING STRATEGIES.** More generally, to the extent that our objective is a description of data in terms of models, we need to develop better means of assessing progress. A first step toward this end might be an attempt to characterize our overall strategies. It seems to me that two rather different strategies are apparent in current practice. The first of these constitutes an attempt to achieve explanatory and predictive power with respect to a class of phenomena by means of a nested sequence of experiments together with fits of a model, or more often of successive elaborations of a model, to the accumulating data. In the initial stage, a model is found that provides a satisfactory description of the data of a particular experiment. Then a new experiment is conducted involving some variation on the first and an attempt is made to fit the same model to the new data. When the attempt is unsuccessful, the model is modified or elaborated as necessary until a fit to the new data is obtained, usually with the constraint that a special case of the new model should fit the data of the original experiment.

This procedure seems to be motivated, at least in part, by the hope of reaching a point where further elaborations will be unnecessary and the model will prove capable of anticipating the results of new manipulations. In practice, however, this goal is rarely, if ever, attained; the cycle simply continues until the model becomes too cumbersome to be tractable or until the whole development is made obsolete by some new approach to the same material.

A more practical purpose of sequential model-fitting is to obtain a check on the adequacy with which our theoretical conceptions take account of the determiners of the phenomena we wish to predict. We cannot expect to achieve this end by fitting the data of a single experiment. We do research only when we lack complete understanding of a situation, and in the absence of complete understanding we may unknowingly hold constant some factors that would produce major effects if allowed to vary. In the hope of ensuring to some extent against this hazard, we do not rest with an account of one experiment, but follow the iterative procedure of introducing variations into the experiment and attempting to refit the model.

A major drawback to this sequential strategy is that we lack any fully satisfactory stopping rule. When is a description of data in terms of a model good enough? A common answer is "when deviations of observed from predicted statistics of the data fail to reach statistical significance at some conventional level." Unfortunately, this criterion depends so strongly on the power of the experiment that it is all but useless in practice. Almost any model can be accepted if the experiment is sufficiently insensitive and almost any model can be rejected if the experiment is sufficiently powerful.

**THE COMPETITION STRATEGY.** Realization of this evidently insoluble problem has led to increasing reliance on the second common strategy—competition in which a number of alternative models are compared with regard to their ability to describe

the same set of data (good examples would be Atkinson & Crothers, 1964; or Bush & Mosteller, 1959). This strategy certainly has its uses, and in many instances may be the best we have available. When alternative models are available, comparisons of them on the same data nearly always prove illuminating, but they rarely lead to a choice of a single model which then continues to prove superior over the subsequent course of investigation (as witness the two examples cited above among many others which could be cited).

The principal fruits of the competition strategy are to be found in the comparison process itself, which provides an invaluable aid to the testing of scientific hypotheses. Continually we find ourselves posing questions that cannot be answered directly because the processes or mechanisms referred to cannot be supposed to act in isolation. Do drive and incentive combine additively or multiplicatively in their effects on performance? Are the items of a free recall list organized in memory in accord with a chain association or a hierarchical schema? Is a set of items maintained in a short-term memory buffer searched by a self-terminating or an exhaustive scan? To deal with such questions, one must make assumptions concerning other factors that are operative in the test situations. These assumptions, taken together with alternative hypotheses concerning the question at issue, constitute alternative models for the situation. And by determining which model provides the better fit to the test data, we can gain some evidence bearing on the merits of the alternative hypotheses.

One of the drawbacks to the competition strategy is the extreme meagerness of the literature concerning statistical testing of the relative goodness of fit of alternative models to the same data. Wholly satisfactory solutions seem to be available only in a few rare instances where maximum likelihood tests are possible or in which the models being compared are nested in such a way as to permit  $\chi^2$  tests (see, e.g., Young, 1971). Further, as in the case of the nested sequence strategy, there is no rationale to lend confidence that a succession of applications of the strategy in a given area should be expected to lead toward models of increasing generality. Most acutely, we lack any methodology for statistical testing of the descriptive adequacy of alternative computer simulation models.

To the extent that our problems in the evaluation of models arise from deficiencies in statistical methodology, there may be little that most investigators in the field can do other than to point up these problems as targets for investigations by mathematical statisticians. However, we should recognize that even if the statisticians did all that could be asked of them our problems would not all be solved. For a statistical goodness of fit is not the only criterion of descriptive adequacy of models and descriptive adequacy is not the only measure of the fertility of a model or theory in relation to research.

CATEGORIZATION IN TERMS OF MODELS. Pursuing this line of thought I would like to suggest another concept, which we might term the *sharpness* of a model, that

is almost the antithesis rather than the complement of the standard concept of goodness of fit. We commonly speak of one of the major purposes of model construction as being that of describing data, but I do not believe that we really mean what we say. The purpose of constructing models is not to describe data, which must be described before models can be applied to them, but rather to generate new classifications or categorizations of data. It is true enough that a model that fits observations well may on occasion achieve a substantial amount of data reduction—many examples are to be seen for example in Coombs' theory of data (1964) and quite spectacular ones in applications of the constant ratio rule associated with Luce's choice model (1959). But compactness of description, however useful, is at most a facilitator of theoretical advance.

What we hope for primarily from models is that they will bring out relationships between experiments or sets of data that we would not otherwise have perceived. The fruit of an interaction between model and data should be a new categorization of phenomena in which observations are organized in terms of a rational scheme in contrast to the surface demarcations manifest in data that have only come through routine statistical processing. But it is a truth that seems obvious yet is not widely appreciated that models will not force us to new categorizations of phenomena if they are so flexible that they can always be made to fit each new experiment by suitable choices of parameter values.

Given a particular collection of observations, we generally advance our understanding but little if we find a model that will describe the entire collection with a suitable choice of parameters. But we feel that we have made a considerable advance if we find that when the collection of observations is categorized into two or more subsets along lines dictated by a particular model, we can achieve good fits to the observations within one or both categories by appropriate submodels, whereas the observations uncategorized or wrongly categorized from the standpoint of the theory remain refractory. A classical example from physics would be the categorization of phenomena having to do with the propagation of light into those described by wave versus those described by corpuscular models. A similarly familiar example occurred in the early development of the Mendelian models in genetics where it proved necessary to categorize data into those representing the principle of random assortment and those involving linkage. An illustration from psychology would be Stevens' (1957) categorization of prothetic versus metathetic sensory dimensions in relation to scales of measurement.

The immediate background for these fruitful categorizations, or recategorizations, of phenomena generally comprises a combination of conditions: (1) A model that yields definite testable predictions; (2) sufficient agreement with data to lend confidence that up to a point the processes involved are understood; and (3) specific errors of prediction that prove diagnostic of inadequacies in the conceptual structure and point the way to theoretical advance.

## ON MODELS FOR PROCESSES VERSUS MODELS FOR EXPERIMENTS

We have observed that a principal hazard associated with an overpreoccupation with the fitting of data is that our available strategies for the fitting and evaluation of models provide no assurance of progress toward models that are progressively more general and more useful in disclosing new relationships among previously isolated phenomena or theories. To move toward generality, we need deliberate and explicit efforts to increase our understanding of variables and processes one step more abstract than those involved in the interpretation of specific experiments. Perhaps the outstanding success story for this approach in contemporary psychology is the development of signal detectability theory, with its widespread applicability to problems of perception, judgment, and memory (see, e.g., Murdock, 1974; Swets, 1964; Tanner & Swets, 1954). A still more current example would be the flourishing of scanning models both for perceptual processing and for memory search (Estes & Taylor, 1964; Shiffrin, 1970; Sternberg, 1966) and the systematic investigation of these processes at an abstract level by Townsend (1974) with the resulting clarification of much-clouded issues concerning parallel versus serial scanning processes.

There is little point in speculating as to just where in the field the next major developments of this sort may appear, but we can point to one where such developments are acutely needed. I refer to research on learning and memory of verbal material, particularly in relation to models for the optimization of instruction. Efforts toward the development of theoretical bases for optimization of instructional schedules (Atkinson, 1972; Chant & Atkinson, 1973; Dear, Silberman, Estavan & Atkinson, 1967) have yielded useful and even impressive results, but the applicability of the models resulting from these efforts remains severely restricted to some rather simple varieties of list learning. The reason, I believe, arises from a corresponding limitation in our models for learning and memory that, with respect to dependent variables, are almost entirely limited to a few elementary empirical measures such as frequencies or latencies of correct responses to individual items of a list. However difficult the task may be, I suspect that the attainment of optimization models of much greater generality will have to wait upon our achieving more effective ways of measuring the amount of information stored and retrieved in the course of typical experiments in learning and memory and the formulation of models in terms of transformations of information rather than changes in state of representations of individual items in memory.

## MEASUREMENT, SUBSTANCE, AND ORGANIZATION

Undoubtedly the most elegant formal accomplishments of mathematical psychology are those having to do with models of measurement (see, e.g., Krantz, Luce, Suppes & Tversky, 1971). One motivation for the heavy concentration of attention on measure-

ment has been the intrinsic interest of the mathematical problems and the possibility of finding solutions, always a winning combination. A second motivation is more substantive. From the first efforts toward psychological measurement, investigators also seem to have had in mind the goal of making progress toward generality in psychological theory by developing quantities analogous to mass, charge, and the like in physics and showing that laws and principles formulated in terms of these derived quantities would have greater generality than those formulated in terms of observables. But although there have been some substantial accomplishments in the measurement of sensory magnitudes and some strenuous, though to date not conspicuously successful, efforts to do the same for such "psychological magnitudes" as utility and subjective probability (Lee, 1971; Luce & Suppes, 1965), by and large this approach has not yielded a rich harvest in psychological research.

One reason for the relative paucity of connections between measurement theory and substantive theory in psychology may arise from the fact that models for measurement have largely been developed independently as a body of abstract formal theory with empirical interpretations being left to a later stage. The difficulty with this approach is that the later stage often fails to materialize. A fertile interaction between measurement theory and research seems more likely to evolve when a measurement model is part and parcel of a theory developed for the interpretation of a process as, for example, has been the case with Luce's (1959) model for choice behavior and Krantz's (1972) approach to magnitude estimation.

Although the classical approaches to measurement deserve and will no doubt continue to receive attention, it seems likely that in the immediate future the center of action is going to shift from the measurement of simple magnitudes and dimensions to the measurement and representation of organization in behavior. The indicators in this direction are perhaps most conspicuous in the area of human memory and language where organization and structure are the order of the day (see, e.g., Tulving & Donaldson, 1972). But although there is much talk in the current literature about organization of processes in memory and many new techniques for the testing of hypotheses about organization, there have been as yet only the most rudimentary beginnings of formal theory.

Many lines of research point to the need for hierarchical models of the organization of material in memory (e.g., Anderson & Bower, 1973; Collins & Quillian, 1972; Estes, 1972; Mandler, 1972) and one would expect that in consequence we should soon begin to see the emergence of formal theories of the type needed to interpret hierarchical relationships. Tentative suggestions have been put forward (e.g., Allen, 1971; Greeno, 1972) that graph theory may provide a natural medium for the representation of structural aspects of memory, but the suggestions seem not to have taken.

One reason, I would surmise, for the slowness with which these suggestions have borne fruit is the tendency for separation of structure and function in models for organization in psychology. Simply assuming that certain types of structures exist

in memory does not in itself help to interpret processes of acquisition, retention, and retrieval of information. The challenge is to produce theories that include assumptions as to how elements of hierarchical memory structures are laid down and how the structures are transformed as a function of experience. So far, efforts of this sort have been confined largely to computer simulation models, but it does not seem that the harnessing of structure to process need be the sole prerogative of the computer simulation approach. A complementary approach that may prove fruitful is illustrated by the recent work of Hogan (1975) who assumes that properties of a directed graph representation of memory are subject to a Markov learning process and shows that the assumed process can be monitored in a free recall situation by means of overt rehearsal.

#### A CASE HISTORY: MATHEMATICAL AND COMPUTER MODELS FOR PAIRED-ASSOCIATE LEARNING

A number of points discussed in the preceding sections can be conveniently illustrated in terms of the development of models for paired-associate learning that flourished during the 1960's. This line of research was projected almost instantaneously into a position of high visibility by the discovery that extremely simple two-state Markovian, "one-element," models could provide strikingly accurate accounts, not only of the course of acquisition, but of numerous fine-grain statistics (e.g., variances, distributions of success and error runs, serial correlations) of the data of certain paired-associate learning experiments (Bower, 1961; Estes, 1961).

For the experiments initially treated, there was literally almost no room for improvement on the closeness of fit of models to data. However, the sequential strategy of modifying the experimental paradigm and refitting the model was set in motion and it soon became apparent that the models were adequate only for a limited class of experiments involving lists of dissimilar stimuli with exactly two response alternatives. It was shown, for example, by Estes and DaPolito (1965) that with more than two response alternatives, the one-element model provided an adequate account of recognition data but broke down in the case of recall data. Continuing efforts to achieve descriptive adequacy for a broader range of conditions then branched into two quite disparate courses, one utilizing mathematical models and the other computer simulation techniques.

In the first branch, a study by Polson, Restle, and Polson (1965) generated a useful categorization, based on a specific deviation of data from the baseline model. These investigators included pairs of confusable items in a list of distinctive items. They found that the data for the distinctive items were adequately handled by the one-element model but the confusable items required the addition of a discrimination process, which again could be described by a two-state Markov model.

Investigators who attempted to deal with the learning of longer lists with multiple response alternatives and varying degrees of confusability on both stimulus and response dimensions almost unfailingly found it necessary to elaborate the one-element model by incorporating one or more forgetting processes. Thus, there appeared the "long-short" models proposed by Atkinson and Crothers (1964), and a related model explored by Greco (1967), all including both short and long-term memory states deriving from the Atkinson and Shiffrin (1968) system; a "forgetting model" advanced by Bernbach (1965) which bypassed the long-short distinction but included an intermediate state of partial learning; and then a four-state "general forgetting theory" (GFT), first presented by Bjork (1966) and subsequently shown by Rumelhart (1967) to include all of the others as special cases.

The GFT enlarged the empirical scope of this family of models by providing for the representation of spacing effects. However, Young (1971), after satisfying himself that no variant of GFT could account both for the nonmonotone effect of interval between study trials and the advantage of interspersing study with test trials, brought this line of model construction to its presently maximal height of complexity. His five-state, seven-parameter model, incorporating multiple short-term memory states, does bring the two vagrant empirical relationships into the fold, though at the cost of a strenuous and intricate procedure of parameter estimation.

Now, how can we evaluate the theoretical progress associated with this sequence of elaborations on both experiments and models? If we take a full description of data in terms of a model as the primary goal, the answer is uncertain. On the one hand, the GFT and Young's extension provide rather impressive accounts of a considerable range of data. On the other hand, as the number of parameters has become large, the pattern of estimated values has tended to vary over experiments in an uninterpretable fashion and the tradeoff between parameters has lessened the value of discrepancies between observed and predicted data for diagnosing the contributions of underlying processes. Further, we can exhibit no tangible grounds for believing that we have reached a point of convergence of theory and observed phenomena such that we can expect the next variations on the experimental paradigm to be accommodated without still additional elaborations of the models.

If, however, we take the main purpose of developing models to be that of aiding in the identification of processes and generating significant classifications of phenomena, then the picture is brighter. We have mentioned the role of specific shortcomings of the one-element model in pointing to a separation of associative and discriminative processes. Similarly, the pattern of discrepancies between theory and data that was generated by the rather massive effort of Atkinson and Crothers (1964) to treat a collection of some eight experiments in terms of long-short models led to the hypothesizing of a trial-dependent forgetting process. The nature of the dependence is a direct relationship between the probability of forgetting of an item between two presentations and the number of intervening items that are in an unlearned state.

This idea was followed up by Calfee (1968), who showed that beyond a lag of zero items, spacing effects on paired-associate acquisition were negligible if the number of unlearned items currently being processed was taken into account. And on still another tack, Humphreys and Greeno (1970) utilized a special case of GFT effectively in conjunction with an experiment in which both stimulus and response difficulty were varied for the purpose of testing alternative hypotheses concerning the stages of paired-associate acquisition. Interpretation of their data within the framework of the model yielded evidence favoring the hypothesis that the subject first stores a representation of a stimulus-response pair as a unit in memory, then in a second stage learns a retrieval route to the stored representation.

The goal of encompassing a still wider range of paired-associate data within a single model certainly is not in sight, and begins to appear not so much unattainable as unattractive. The Markovian models have served rather effectively in the generation and interpretation of research on paired-associate learning, but their usefulness in these respects has not been obviously furthered by increasing complexity of the models.

There is less to be said about computer simulation models for paired-associate learning since the literature to date is largely confined to work associated with the EPAM model originated by Feigenbaum (1963). The basic model includes two learning processes, image building and discrimination [closely parallel to the associative and discriminative processes identified by Polson, Restle & Polson (1965)] and a sorting mechanism that generates recognition by sifting input stimuli through the network of images of stimulus-response pairs laid down during preceding acquisition trials. In general, the course of successive variation of experimental paradigms and elaborations of the model proceeds in much the same way as with the more traditional mathematical models. In one of the few studies that has been reported in detail, a third edition of EPAM was applied to several paired-associate experiments involving variations in similarity and meaningfulness of items (Simon & Feigenbaum, 1964) and statistics computed for simulated protocols proved to yield quite satisfactory fits to the group data.

As with the Markovian models it is difficult to give any formal evaluation of the goodness of fit that has been achieved, and the more interesting fruits of the approaches are to be found in byproducts of the data-fitting. Thus, Simon and Feigenbaum concluded that a satisfactory interpretation of results on intralist similarity requires the assumption that subjects recode CVC's in terms of auditory features. And in the course of fitting other data, Simon and Feigenbaum arrived at an interesting hypothesis (which seems not to have been definitively followed up) to the effect that both meaningfulness and familiarity of items, as manifest in paired-associate learning, are generated by the same familiarization process.

Finally, by means of an interesting variation on the usual protocol-fitting procedure, Hintzman (1968) showed that one can work within the framework of a computer simulation model, just as has been more commonly done with mathematical models,

to explicate properties of a single process that is assumed to be operative in many situations. Hintzman simplified the EPAM model, retaining only the discriminative mechanism and subsuming other aspects under random error. Then he was able to provide a useful evaluation of the adequacy of the assumed discriminative process by generating qualitative accounts of the pertinent empirical relationships in a rather extensive range of paired-associate experiments.

TRENDS IN PAIRED-ASSOCIATE MODELS. Judging by the dates in the reference list of this article, the wave of concentrated effort directed toward the interpretation of paired-associate learning in terms of formal models has somewhat subsided. How should we appraise the present state and the likely lines of continuation?

As a consequence of the various efforts toward model construction, it begins to seem clear that formal theory must continue to develop at several distinct levels—at the least those associated with individual items, lists, and vocabularies. Prior to the emergence of Hull's behavior theory, research on paired-associate learning was oriented toward measures of performance on a list of items, for example, trials to reach a criterion of an errorless cycle through the list. Progress toward theory was retarded by the lack of any rationale relating these measures to the systematic dependent variables of either association or conditioning theory.

The first major breakthrough was Gibson's (1940) interpretation of the acquisition of paired-associate items in terms of reinforcement concepts. The association between the stimulus and response terms of an item was the unit of analysis and the effects of similarity and interference among items were treated in terms of generalization and differentiation. The sequence of mathematical models running from the simplest one-element and linear models through GFT all were developed within the same framework, assumptions concerning both acquisition and retention being formulated with reference to memory and performance states of individual items.

Only with the advent of the EPAM model have we seen a major departure from the focus on the item as a unit. Paired-associate acquisition has come to be conceived of in terms of the growth of an associative structure in which the information represented concerning any item depends on the properties of the list in which the given item is embedded. This assumption of list-dependent accrual of item information has not been incorporated into mathematical, as distinguished from computer simulation, models, no doubt owing to considerations of tractability. Whether such a development is feasible at all may turn on the possibility of formulating process assumptions in terms of higher-order dependent variables, as, for example, some measure of the total amount of information stored or the amount retrievable concerning a list of items as a whole.

But new challenges to the construction of theories do not end here. Anderson and Bower (1973) have criticized the EPAM models on the ground that they do not take adequate account of the relationships of new items to the learner's current state of

semantic memory. In Anderson and Bower's simulation model (dubbed HAM) the theoretical frame of reference is again broadened and what is learned about a new item on a paired-associate acquisition trial depends on the entire associative network built up during the individual's previous relevant experience, for example, his entire vocabulary in the case of acquisition of words of a second language.

Although the concepts and assumptions of Anderson and Bower's model, or the related one of Rumelhart, Lindsay, and Norman (1974), present too complex an ensemble to invite attempts at comprehensive mathematical formulation, it is too early to predict the further course of theoretical development with much assurance. It may prove that aspects of the memory structures conceived in these models can be represented and studied effectively in terms of branches of mathematics, for example, topology or graph theory, that are so far largely untried in this context. More immediately, investigators may find that computer models sufficiently elaborate to encompass syntactic and semantic aspects of verbal learning will not provide an adequate substitute for mathematical models as a medium for reasoning about abstract relationships among phenomena. Thus, we may see increasingly frequent application of a strategy in which a simulation model provides the framework for investigation and then fragments or subsystems of the model are expressed in mathematical form to permit deeper analysis of the dynamic properties of real-time processes. An excellent illustration of this strategy is to be seen in the treatment of sentence memory by Anderson and Bower (1973, Chap. 10).

MODELS OR THEORY? Numerous rather impressive results have been achieved at the level of formulating models that provide close accounts of specific empirical relationships in situations where many other factors are held constant. Efforts toward a broader synthesis have largely been limited to the development of more complex models that include simpler predecessors as special cases. These efforts have taken two somewhat complementary forms. The stochastic models of the family dubbed "general forgetting theory" and the Atkinson-Shiffrin system incorporate processes having to do with the dynamics of acquisition and retention. Computer simulation models of the EPAM-HAM family embody processes having to do with the encoding and discrimination of information concerning similarities and semantic relationships among items.

In what direction should we look for further progress toward a theory, as distinguished from a collection of models? Still more complex models of either type can, and well may, be constructed, but it is not clear that these will effectively serve the purposes we have identified as the principal *raison d'être* for formal models in research. More might be accomplished by systematic efforts to organize the information that has been gained from model-oriented research concerning constituent processes and to generate a theoretically significant classification of the phenomena of paired-associate learning in terms of the combinations of processes implicated in various situations.

## CONCLUDING COMMENT

It might seem that mathematical psychology has had problems enough trying to gain a foothold in terrain that as yet offers only meager material for formal theory. But on top of these, our discipline finds itself confronted with a state of affairs which in clinical quarters might be termed an identity crisis. Suddenly, the all but omnipresent computers are taking over one after another of the functions formerly served only by mathematics. It seems that many of the things mathematical psychologists have learned to do with much effort the programmer can already do faster and better. Is mathematical psychology in danger of becoming obsolescent before reaching maturity?

A review of the current situation suggests rather that there are more than enough problems to go around and that mathematical and computer simulation models may prove more complementary than competitive. The computer program offers means of working with ideas that are insufficiently explicit for mathematical expression and techniques for simulating behavioral protocols that are too complex to be fitted by tractable mathematical models. But mathematics remains our principal vehicle for the flights of imagination that smooth our experiences and extract from varying contexts the relationships that would hold among events under idealized noise-free conditions. These abstract representations are not necessarily closely descriptive of data obtained in real noise-filled environments, but in the course of interactions with data they generate the reorganizations of our ideas that constitute new theory.

We cannot be sure that we have identified all of the reasons why mathematical psychology falls short of realizing our most ambitious aspirations with regard to cumulative impact on the field. But it does seem apparent that one major reason has to do with the fact that our long-term objectives are rarely kept sufficiently in mind during the course of our day to day research to have much influence on our choice of actions.

Although it is unrealistic to hope for changes on a grand scale, small measures that might yield tangible benefits are surely within our hands. We might, for example, cultivate just a bit more dissatisfaction with the assumption that if we attend devotedly enough to the fitting of models to data the problem of generality will take care of itself. Or with the assumption that, if we devote most of our efforts to dealing in isolation with measurement or with substance, with structure or with process, these strands will magically come together to form harmonious theories. Perhaps we need to shift the allocation of our processing capacity from virtually 100% concentration on tactics to a division that allows for some day to day consideration of strategy.

It is interesting to note that two investigators who have recently surveyed problems of mathematical models in psychology from a cultural standpoint somewhat different than that of the editors of our *Contemporary Developments* have arrived at a rather similar diagnosis and prescription:

An analysis of the present situation shows that contemporary psychology and contemporary mathematical instruments are still not compatible enough with one another to allow mathematization to assume a central place in the development of psychological knowledge; the reason for this is not merely the low level of sophistication of the latter but also the lack of sufficiently sophisticated mathematical instruments that are especially adapted for use in psychology. What is required is a continual interaction between mathematics and psychology, an interaction that on the one hand would lead to a restructuring of psychological theories into forms more amenable to the proposed mathematical instruments and, on the other, to a revision of existing mathematical methods into forms more amenable to the proposed mathematized conceptual systems. (Leont'ev & Dzharov, 1973, p. 20).

## REFERENCES

- ALLEN, M. Graph theory, organization, and memory. Paper given at the meetings of the Eastern Psychological Association, New York, April, 1971.
- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. Washington, D. C.: V. H. Winston & Sons, 1973.
- ATKINSON, R. C. Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 1972, **96**, 124-129.
- ATKINSON, R. C., BOWER, G. H., & CROTHERS, E. J. *An introduction to mathematical learning theory*. New York: John Wiley & Sons, 1965.
- ATKINSON, R. C., & CROTHERS, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. *Journal of Mathematical Psychology*, 1964, **2**, 285-315.
- ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: a proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation: advances in research and theory*. Vol. 2. New York: Academic Press, 1968. Pp. 89-195.
- BERNBACII, H. A. A forgetting model for paired-associate learning. *Journal of Mathematical Psychology*, 1965, **2**, 128-144.
- BJORK, R. A. Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals. Technical Report #106, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.
- BOWER, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255-280.
- BUSH, R. R., & MOSTELLER, F. A comparison of eight models. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford, CA: Stanford University Press, 1959. Pp. 293-307.
- CALFEE, R. C. Interpresentation effects in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1030-1036.
- CHANT, V. G., & ATKINSON, R. C. Optimal allocation of instructional effort to interrelated learning strands. *Journal of Mathematical Psychology*, 1973, **10**, 1-25.
- COLLINS, A. M., & QUILLIAN, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972. Pp. 309-351.
- COOMBS, C. H. *A theory of data*. New York: John Wiley & Sons, 1964.
- COOMBS, C. H., DAWES, R. M., & TVERSKY, A. *Mathematical psychology*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1970.
- DEAR, R. E., SILBERMAN, H. F., ESTAVAN, D. P., & ATKINSON, R. C. An optimal strategy for the presentation of paired-associate items. *Behavioral Science*, 1967, **12**, 1-13.

- ESTES, W. K. An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory*. New York: V. H. Winston & Sons, 1972. Pp. 161-190.
- ESTES, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73-84.
- ESTES, W. K., & DAPOLITO, F. Independent variation of information storage and retrieval processes in paired-associate learning. *Journal of Experimental Psychology*, 1967, **75**, 18-26.
- ESTES, W. K., & TAYLOR, H. A. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964, **52**, 446-454.
- FEIGENBAUM, E. A. Simulation of verbal learning behavior. In E. A. Feigenbaum and J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963. Pp. 297-309.
- GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, **47**, 196-229.
- GLENBERG, A. M. Retrieval factors and the lag effect. Ph.D. dissertation, University of Michigan, 1974.
- GREENO, J. G. Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. *Journal of Mathematical Psychology*, 1967, **4**, 430-472.
- GREENO, J. G. On the acquisition of a simple cognitive structure. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972. Pp. 352-377.
- HANNA, J. F. Some information measures for testing stochastic models. *Journal of Mathematical Psychology*, 1969, **6**, 294-311.
- HINTZMAN, D. L. Explorations with a discrimination net model for paired-associate learning. *Journal of Mathematical Psychology*, 1968, **5**, 123-162.
- HOGAN, R. M. Inter-item encoding and directed search. *Memory and Cognition*, 1975, **3**, 192-209.
- HUMPHREYS, M., & GREENO, J. G. Interpretation of the two-stage analysis of paired-associate memorizing. *Journal of Mathematical Psychology*, 1970, **7**, 275-292.
- KOZIELECKI, J. (Ed.), *Problems of mathematical psychology*. Warsaw: PWN, 1971.
- KRANTZ, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, **9**, 168-199.
- KRANTZ, D. H., ATKINSON, R. C., LUCE, R. D., & SUPPES, P. (Eds.), *Contemporary developments in mathematical psychology. Vol. 1. Learning, memory, and thinking*. San Francisco: W. H. Freeman, 1974.
- KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. *Foundations of measurement*. New York: Academic Press, 1971.
- LAMING, D. *Mathematical psychology*. New York: Academic Press, 1973.
- LEE, W. *Decision theory and human behavior*. New York: John Wiley & Sons, 1971.
- LEONT'EV, A. N., & DZHAFAROV, E. N. Mathematical modeling in psychology. *Soviet Psychology*, 1974, **12**, 3-22.
- LUCE, R. D. *Individual choice behavior*. New York: John Wiley & Sons, 1959.
- LUCE, R. D., BUSH, R. R., & GALANTER, E. (Eds.), *Handbook of mathematical psychology. Vol. 1*. New York: John Wiley & Sons, 1963.
- LUCE, R. D., & SUPPES, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology. Vol. 3*. New York: John Wiley & Sons, 1965. Pp. 249-410.
- MANDLER, G. Organization and recognition. In E. Tulving and W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972. Pp. 139-166.
- MURDOCK, B. B., JR. *Human memory: theory and data*. New York: John Wiley & Sons, 1974.

- POLSON, M. C., RESTLE, F., & POLSON, P. G. Association and discrimination in paired associate learning. In E. D. Neimark & W. K. Estes (Eds.), *Stimulus sampling theory*. San Francisco, Holden-Day, 1967. Pp. 231-239.
- RESTLE, F., & GREENO, J. G. *Introduction to mathematical psychology*. Reading, MA: Addison-Wesley, 1970.
- RUMELHART, D. E. The effects of interpresentation intervals on performance in a continuous paired-associate task. Technical Report #116, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967.
- RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972. Pp. 197-246.
- SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970. Pp. 375-447.
- SIMON, H. A., & FEIGENBAUM, E. A. An information processing theory of some effects of similarity, familiarity, and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 385-396.
- STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, 153, 652-654.
- STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, 64, 153-181.
- SWETS, J. A. *Signal detection and recognition by human subjects*. New York: John Wiley & Sons, 1964.
- TANNER, W. P., & SWETS, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, 61, 401-409.
- TOWNSEND, J. T. Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: tutorials in performance and cognition*. Potomac, MD: Erlbaum Associates, 1974.
- TULVING, E., & DONALDSON, W. (Eds.), *Organization of memory*. New York: Academic Press, 1974.
- YOUNG, J. L. Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology*, 1971, 8, 58-81.

RECEIVED: December 12, 1974