35

Memory Models

ROGER RATCLIFF & GAIL MCKOON

The study of models of memory often seems like a backwater in the overall study of memory. Models do not have a prominent place in experimental studies of memory and they are not used or examined by most researchers in the field. This review examines the various questions that models can address, discusses why theory is not as prominent in the memory domain as in other domains of science, and presents an overview of current models. The aim is to show why models should have greater prominence and wider use.

Mainstream models for "long-term" memory take as their database the results of experiments in which subjects are asked to study and learn lists of items (words, nonsense syllables, letters, numbers, sentences, or pictures). Memory is tested in one of a number of ways: asking subjects whether or not an item occurred on the study list (recognition), asking for recall of the items on the study list, asking what item on the study list was associated with a cue, asking for the recency or frequency of appearance of an item on the study list, and so on. The dependent measures are usually accuracy, confidence ratings, or a combination of reaction time and accuracy. The eventual aim is to account for the effects on all the dependent variables for a range of experimental tasks and for a range of experimental manipulations, including the length of the study list, the strengths of the items in memory, the type of material, the similarity among study and test items, levels of processing, rehearsal methods, and so on.

Recent development of models of long-term memory has proceeded relatively independently of other areas of memory research. For example, there has been little contact between the long-term memory models and the findings of implicit memory experiments and there has been little explicit theoretical work in the domain of implicit memory. Over the last 20 years, the domains of reaction time research and memory have not interacted in strongly productive ways, although there has been a recent resurgence of interest in random walk and diffusion reaction time models and so there may soon be more fruitful interactions. The one domain of research with which there is some sharing of representation and process assumptions is categorization. In this domain, subjects are presented with exemplars and through feedback learn how to assign the exemplars to categories. Some models of categorization are essentially long-term memory models in that they assume a representation of a category is built up with learning and the category decision process depends on retrieval from this memory representation.

Short Historical Background

The attempt to produce models of memory that can account for data both qualitatively and quantitatively has had a history of working from small models with very restricted ranges of application to more comprehensive models with wider ranges of application. Early memory models (in the 1960s to '70s) used the dichotomy between short-term memory and long-term memory as a reason either for restricting the domain of the model to short-term or long-term memory or for explaining how information progressed from short-term to long-term memory. In many of the models, component processes paralleled developments in mathematical psychology, borrowing the mechanisms of Markov chains (used in models of associations), signal detection theory (used in perception and strength based models), or serial search processes (derived from the computer metaphor and used in the influential Sternberg, 1966, serial search model). Two excellent sources for the state of the art in the early 1970s are the books, *Models of Human Memory* (Norman, 1970) and *Human Memory: Theory and Data* (Murdock, 1974). As is shown in these books, many of these were models of a particular task (e.g., paired associate learning and cued recall, or memory search and recognition) and the models were usually applied only to a restricted range of experimental data.

In the late 1960s and early 1970s, there were two more comprehensive models that were particularly influential. One was Atkinson and Shiffrin's (1968) model of short-term and long-term memory. The model provided a qualitative/theoretical basis for the separation of short- and long-term memory, although the mathematical structure of the model was not widely applied to experimental data. The model served as a focus for testing hypotheses about how the two kinds of memory might be differentiated and it was the "modal" memory model in that it represented the pinnacle of development of models of its class. It was also the stepping-off point for attacks on the simple rigid bipartite division of memory into short- and long-term components.

The second influential model was Anderson and Bower's (1973). Because the model was explicitly applied to memory both for the standard paradigms and for sentences, it was influential at the intersection of traditional memory research and the rapidly developing domain of language research. This union of memory and language theory evolved in the early 1990s into a rational model (Anderson, 1990, 1993) of memory that is intended to apply over a range of domains that includes memory, sentence processing, and categorization. The notion of rationality has become a critical component of the newest memory models, as will be discussed below. Anderson, Bothell, Lebiere, and Matessa (1998) recently applied the Anderson (1993) model to the standard list-learning experiments targeted by more traditional models, but the model was fit only to selected aspects of data so, until more comprehensive evaluations have been carried out, it is too soon to tell how the model will fare.

Beginning in the 1980s, models were developed that were aimed at being comprehensive in both the range of tasks and the range of data to which they were applied. Murdock's (1982) TODAM model was designed to apply across the categories of information that Murdock used to classify memory, item information, associative information, and serial order information (see Murdock, 1974, p. 16), and the processes that operate on those kinds of information. Around the same time, Raaijmakers and Shiffrin's (1981) search model for free recall (SAM) accounted for many of the experimental findings from free recall experiments, and Gillund and Shiffrin (1984) extended the model to data from recognition experiments. A little later, Hintzman's (1986, 1988) MINERVA2 model was applied to recognition and categorization and then extended to judgments of recency and frequency. The next sections of this review describe these models in more detail, then describe the empirical phenomena that were inconsistent with the models, and then finally describe the next generation of models.

Global Memory Models

The SAM, MINERVA2, and TODAM models are called the global memory models for two reasons. First, for each of the models, it is assumed that a test item contacts a great deal of information in memory, possibly all stored memories. Second, the models were intended to explain data from a range of experimental tasks and a range of experimental manipulations within those tasks.

MINERVA2, SAM, and TODAM each make different assumptions about how information is represented in memory. SAM stores strengths between cues to memory (test items) and items in memory. TODAM assumes that items are vectors with random values as the elements of a vector. Memory is assumed to be a single vector into which all the item vectors are stored. MINERVA2 assumes that items are vectors with elements +1, 0, or −1, and that items are stored in separate vectors in memory. For recognition memory, the three models share the same assumption about retrieval: at test, a test item matches all of memory and a single value of familiarity is produced that serves as the basis for the decision about whether the test item was or was not on the study list. Recognition memory is the main area in which the coverage of the models overlaps, so it has become a central focus for testing and evaluation, and the two models of the next generation, to be described below, focus almost entirely on recognition memory.

A good introduction to the global memory models is presented in Neath (1998) and this would be a useful starting point for readers new to memory models.

The SAM Model

The SAM model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) represents information in a cued dependent structure. That is, what is stored are the strengths of the connections between cues (test items) that interrogate memory and items stored in memory ("images"). These strengths are built up by an encoding process that uses a simple short-term buffer. During the time an item to be studied is in the buffer, c units of "self strength" per unit of time are built up between the item as a cue and the image of the item in memory and a units of strength are built up between the general study context and the image of the item in memory. "Interitem strengths" (b units per unit of time) are also built up between the item as a cue and the images in memory of each of the other items that are in the buffer at the same time. There is also assumed to be some residual strength between an item as a cue and the images of each of the other items in memory; this serves the role of pre-experimental strength of connection between items.

At retrieval, a cue interrogates memory. In recognition, the cue is assumed to be the test item plus the context in which it is presented. The strength from each of these to an item in memory is calculated, and the product of these two strengths is formed. This calculation is performed for all items in memory and all the products are summed. The sum is a measure of the global familiarity of the test item. This value of familiarity is used in a signal detection analysis to produce hit and false alarm rates. A criterion is set on familiarity, and if the computed value is greater than the criterion, then the decision is that the item was on the study list and an "old" response is produced; if the computed value is lower than the criterion value, then the decision is that the item was not on the study list and a "new" response is produced. Free recall is a two-phase process, sampling and recovery. First, the context cue is used to probe memory and for each item in memory, a "sampling probability" is produced; this probability is a function of the item's strength relative to other items and it is the probability that the item will be selected for recovery. If an item is selected, then it has some probability of being recovered that is a function of its strength, so the stronger the item, the greater the probability of recovery. If recovery fails, the process resamples and attempts another recovery. An item that is successfully recovered is used as a cue along with context for an attempt to retrieve another item. This process continues until a criterion number of attempts is made without a successful retrieval. Cued recall works like free recall except that there are two cues with which to probe memory, the recall cue and the context.

SAM has been successfully applied to a wide range of experimental data. The recognition process has changed in the transition to the new generation of models, but the recall assumptions still provide the most successful explanation of recall data.

The MINERVA2 Model

In MINERVA2 (Hintzman, 1986, 1988), items are represented as vectors of features. At study, a new memory vector is created for each studied item, with the each feature of the item copied into the vector with a probability that varies as a function of study time. Features have values of plus or minus 1 or, if a feature is not copied into memory, the value for that feature is set to zero. For recognition, a test item's vector is compared with each vector in memory and a dot product (divided by the number of features) is formed to give the degree of match—the "activation" value. This value is cubed and then the activation values for all the items in memory are summed to provide an overall measure of match called "echo intensity." If this value is greater than a criterion, an "old" response is produced; if it is less, a "new" response is produced.

For recognition, the model does a good job of predicting many of the standard experimental results, including the effects of repetition and study time on performance. It also has been successfully applied to frequency judgment data and to categorization. Processes have been suggested that would allow the model to explain recall data but there have not been any specific applications.

The TODAM Model

In TODAM, an item is a vector of attributes and each attribute has a value derived from a normal distribution with mean zero. Memory is a single vector, a composite memory trace of all studied items. A study, for each studied item, the probability that an attribute is encoded is a function of study time. The vector for each studied item is simply added to the single memory vector. If two items are studied together in a pair (e.g., paired associate learning), the association between the two items is the convolution of their vectors (see Neath, 1998) and this convolution is stored in the single memory vector along with the vectors of the individual items.

For recognition, the vector for a test item is compared with the single memory vector using a dot product (i.e., multiplying corresponding attributes together and summing over attributes). The resulting sum is compared with a criterion value just as in the SAM and MINERVA2 models. The mathematics of the model would actually work in exactly the same way if it was assumed that the items are stored in separate vectors in memory instead of a composite, so the differences between this model and MINERVA2 are less than they might seem at first examination.

For cued recall, one member of a studied pair is presented and the vector for this item is correlated with the memory vector. This produces a noisy version of the other member of the pair and this is cleaned up to produce a response. Recent empirical work on the interactions of associative and item information has required a revision of TODAM; TODAM2 is the most recent version of this model (Murdock, 1997).

The Data that Challenged the Global Memory Models

It is often thought that the global memory models are unfalsifiable, that they can be manipulated to produce any pattern of data by varying parameter values or adding new parameters. Recently, Slamecka (1991) voiced this view directly: "If the models have anything, they have resilience, or to put it more precisely, their inventors have resilience, and I suspect that after some skillful patching of assumptions and/or fine-tuning of some parameters, these veterans will lumber down the runway and lift off again." If this were correct, the enterprise would truly be of little interest. Fortunately for science, and unfortunately for the global memory models (and Slamecka's argument), the models are solidly grounded in the phenomena to which they were addressed.

To understand what kinds of tests falsify the models, it is necessary to understand which are the fundamental assumptions of the models that are not subject to patching and parameter twiddling. We discuss two such assumptions. One has to do with the way the output quantities specified by the models (e.g., familiarity values) get larger as the strengths of the items stored in memory increase and the second concerns variability in the match between a test item and memory.

The easiest demonstration of the models' inabilities to handle the effects of increasing strength of items in memory is the "mirror effect." According to the models, increasing the strength of an item in memory should increase the probability that an item is called "old" for items that are indeed "old" because they were on the study list *and* for items that are not "old"—that is, items that were not on the study list. But often this is not the empirical result. For example, low-frequency words are more likely to be called old if they were studied but they are more likely to be called new if they were not (e.g., Glanzer & Adams, 1985; Glanzer, Adams, Iverson, & Kim, 1993).

One solution the models could offer for this problem is that the retrieval process begins by determining whether a test item is low versus high frequency and then alters the decision criterion for old versus new responses, but everyone agrees this is an unsatisfactory solution. Some of the problems with this solution (besides the fact that it is exactly the kind of solution that Slamecka uses to attack the modeling enterprise) are, first, that responses can be made just as quickly when frequencies are not mixed in the study and test lists as when they are; second, subjects are not very good at determining the frequencies of words; and third, in experiments with a number of different frequency values, the multiple criteria required would make the task very difficult.

The second falsifiable assumption of the models is that as the strength of an item increases, variability increases. In SAM, variability is introduced at encoding. For each unit of time that an item remains in the encoding buffer, the units of strength that are built up are not the fixed values of the strength parameters (the c, a, and b values), but instead they are these values with probability one third, .5 times these values with probability one third, or 1.5 times these values with probability one third. It follows that, as the amount of time in the buffer increases, both the encoded strength values and the variability in those values increases. Similar behavior is produced by the variability assumptions in MINERVA2 and TODAM. The assumptions about variance underlay all predictions of the models and only fundamental alteration of the basic structures of the models could change these assumptions.

For each of the models, the variance assumptions lead to the prediction of a "list strength effect." Consider a mixed study list in which some items are studied for a long time (e.g., 5s) and some for a short time (e.g., 1s) versus a pure list in which all items are studied for a short time. For SAM, when a recognition test item is matched against the items in memory, variance in familiarity is increased for the mixed list relative to the pure list because of the increased variance in the strength values of the long-study items. This increase in familiarity variance increases the chance that the familiarity value will be below criterion for an "old" response for short-study items. This results in lower accuracy for the short-study items in the mixed list relative to the pure list. This is the predicted "list strength" effect and data almost never show this effect in recognition (Murnane & Shiffrin, 1991; Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990), though an effect is predicted and obtained in recall (Tulving & Hastie, 1972; Shiffrin et al., 1990). MINERVA2 and TODAM make the same prediction for recognition and so they are also disconfirmed by the data.

The mirror effect and the list strength effect are two examples of predictions of the models that are testable, contradicted by experimental data, and inalterable by minor modifications to the models. These are the kinds of phenomena and tests that are at the heart of testing and evaluating the models.

New Generation Global Memory Models

Two new models have been developed to deal with the phenomena that the older global memory models could not explain, as well as all the phenomena that they could explain. The models, REM (Shiffrin & Steyvers, 1997) and McClelland and Chappell's (1999) model, have much in common. In both, items are represented as vectors of features and, in recognition, the degree of match between a test item and memory is compared with a criterion value to make the old/new decision.

At encoding, the features of each studied item are stored in vectors in memory, a separate vector for each item. In each unit of time, some proportion of an item's features are stored, some with their accurate values and some with incorrect values. The vector for a test item is matched against all the vectors in memory. For each vector in memory, the probability is calculated that the memory vector was generated from a study item identical to the test item. A likelihood ratio is produced by dividing this probability by the probability that the memory vector was generated from some different item. For REM, the likelihood ratios for all the items in memory are summed and the result is compared with a criterion to decide whether the test item is old or new. For McClelland and Chappell's model, the maximum of the likelihood ratios is compared to the criterion. As the proportion of features stored for a study item increases, the probability of a match increases and the probability of a nonmatch decreases, essentially differentiating between old and new test items (Shiffrin et al., 1990).

The key feature of the two new models is the use of the likelihood ratio for evaluating the degree of match between a test item and memory. The models account for the mirror effect because as the likelihood ratio for a match increases (because there are more matching features), the likelihood ratio for a mismatch decreases (because there are more mismatching features). The models account for the list strength effect because strengthening an item only slightly affects the likelihood ratios for mismatches.

There are two main assumptions of these models that allow them to produce likelihood ratios. The first is that feature storage is fallible; incorrect values of features are stored with some probability. If storage were completely accurate, then there would be no mismatches when the test item is the same as the study item that produced the memory vector and the likelihood ratio could not be computed. The second is that items are represented as vectors of individual features. This makes it possible to calculate the probability of a match between a test item and an item in memory that is not the same as the test item.

The two new models are quite similar to Hintzman's MINERVA2 model in terms of their representational assumptions. They differ from MINERVA2 in their assumption about how the degree of match is computed (likelihood ratio versus echo intensity).

The new models are at the stage of development where they have been shown to account for the data that was problematic for the older memory models, but as yet no critical tests of the models have been developed. What will be needed is a clear understanding of which are the critical underlying structures of the models that cannot be altered without completely changing the predictions. These models leave the field at an exciting point, awaiting critical new challenges, applications to wider domains, and the next cycle of testing and evaluation.

Reaction Time and Memory Models

All of the global memory models, including REM and McClelland and Chappell's model, predict response accuracy; little attention is paid to the behavior of the other dependent variable they eventually will have to model, reaction time. Although there are many regularities in the relationship between reaction time and accuracy (see Atkinson & Juola, 1973; Hockley & Murdock, 1987; Ratcliff, 1978), substantive research has not been done to integrate the mechanisms that predict accuracy in these models with mechanisms to predict reaction time.

One of the important problems that will have to be faced is how to translate a dimensionless quantity, familiarity or likelihood, into a quantity with a time-related dimension (e.g., rate of processing). In the two new models, the distributions of likelihood values are highly skewed and there would be no reasonable linear translation from likelihood values to rate of accumulation of evidence. What would be needed would be a transformation (e.g., log likelihood; McClelland & Chappell, 1999) to make the distributions less highly skewed. Considerable research is needed to determine what kinds of transformations might work to correctly predict reaction time data.

What the Models Have Not Been Used For

If the memory models were having an impact on empirical research, then we would expect to see the results of experiments interpreted in terms of model parameters. But this is almost never done; the question is, why not? One reason is that many phenomena will be explained in the same way with or without a model. For example, obtaining a fit of a model to the increase in accuracy that occurs as a function of study time would not be news (even though such explicit fitting has rarely been done). Second, there are technical difficulties in fitting the models to data that have not been explored in detail. For example, in the SAM model, the effects of study time could be modeled by varying any one of several parameters. This means that the studies that show what kinds of data will constrain the models have not been presented in a way that allows nonexperts to use the models. Third, there have been few compelling demonstrations that the models are needed to interpret experimental data in domains where the experimental questions concern simple hypotheses.

The big payoff for actually fitting the models to experimental data will come with applications of the models to research in memory as a function of development in children, aging, head injury, varieties of amnesia, and so on, where the parameters of a model can be used to explain what aspects of processing and/or memory lead to the differences in performance observed across subject groups and between the individuals in a group. Variations in parameter values across individuals of different classes could be used to evaluate hypothesized explanations of their differences in performance. The models could also be used to examine performance across a range of tasks to determine if the same components or processes are affected in the same ways across tasks. This kind of evaluation of common mechanisms across tasks is not possible without theory. In ways like these, the models could be expanded from their narrow focus on standard list learning experimental procedures to questions about the larger domain of human memory.

Uses of the Memory Models: Compound Cue Models for Associative Priming

One of the phenomena that is not explained in the same way with the global memory models as it is without the models is priming in lexical decision and recognition. This kind of priming is the speedup in processing for a word that results from processing a related word just before it. For example, in making word/nonword decisions about strings of letters, if a target test word *doctor* is preceded by a related prime word, *nurse*, then the "word" response to *doctor* is speeded by 30 to 50 ms. The usual theoretical interpretation of priming is based on spreading activation: the prime word activates other words related to it in memory and this advance activation leads to a speedup on the target (see McNamara, 1992a, 1992b, 1994a, 1994b). An alternative explanation of priming is the compound cue model (Ratcliff & McKoon, 1988, 1994, 1995; McKoon & Ratcliff, 1992), which is based on the global memory models' accounts of pair recognition. In pair recognition, a pair of words is presented simultaneously and the task is to decide whether both were on the study list. If the two words were studied in a pair together, the degree of match between the test probe and memory is greater than if both words were studied but in different pairs. Ratcliff and McKoon use the same mechanism to explain priming, adding some assumptions about how reaction time is derived from degree of match. There has been considerable controversy over whether spreading activation (ACT*; Anderson, 1983) or compound cue mechanisms give the best account of priming, with no clear winner.

Models of Categorization

Categorization research has been dominated by two main classes of models, exemplar based models and decision bound models (see contrasts between the models in Ashby & Maddox, 1993; McKinley & Nosofsky, 1995). Exemplar-based models assume that a category membership decision is based on stored exemplars of the category, exemplars that were learned in the process of performing the experimental task. This means that the models are essentially memory models. Nosofsky (1991) examined the relationship between categorization and recognition memory and found that both could be explained for his experimental data with the same exemplar based memory representation but different retrieval processes for the two tasks. Apart from a few examples like this, the categorization models have largely operated independently of the memory models, and more theoretical interaction between models in the two classes is overdue.

Implicit Memory Models

The memory models examined so far deal with what has been called explicit memory, but there is another domain of research that is concerned with the effects of prior study on performance in tasks that do not require recollection of prior study; this has been called implicit memory. Much research on implicit memory has centered on the experimental finding that repetition of a stimulus produces a benefit to performance even when conscious memory of the prior episode with the stimulus is not required. A key result is the finding that on many tasks, this repetition priming effect is unimpaired in amnesics even when their explicit memory, as shown by recognition or recall tasks, is severely impaired. It has been claimed that this priming is produced by a separate memory system from the system that performs explicit memory tasks. For example, Squire (1992) proposed a hierarchy of separate systems and Schacter and Tulving (1994) produced a taxonomy of multiple memory systems. The problem with this approach is that it is driven by hypothesis testing; at no point are the hard theoretical questions asked about how information is represented within each memory system, how processing works within each system, or how processes interact among the systems. Crucially, there is no discussion of how processing works for the tasks in which repetition effects are found.

Consider, for example, what must happen in a multiple memory systems account of priming in word identification. In the tasks used to show this phenomenon, a test word is flashed briefly, then masked. A prior presentation of the word increases the probability of correctly identifying it. If the earlier encounter is stored as a new representation in a separate memory system from that used for word identification, then when the test word is presented it must contact this representation and the representation must become available to the processes that are standardly used for word identification in time to facilitate them. It seems unlikely to us that any reasonable mechanism could be constructed to work this quickly to both identify the test word in the implicit memory system and use the resulting information to aid identification.

Ratcliff and McKoon (1996, 1997) reported data that provided the basis for a different interpretation of implicit priming effects. Using experimental procedures for which costs as well as benefits could be examined, they found that the facilitative priming effect for an exact repetition of an item was accompanied by inhibition in processing when a closely similar item to the test item had been presented earlier. Ratcliff and McKoon argued that this result shows a bias in processing, not the operation of a separate memory system. They explained bias with a model for word identification (Ratcliff & McKoon, 1997), a modification of Morton's logogen model (Morton, 1969). Schooler, Shiffrin, and Raaijmakers (1998) have also proposed a model for bias that does not make use of a separate memory system; their model uses the mechanisms of REM. In both these cases, the models' primary aim is to explain standard processing, and priming is only a by-product of the standard processes (Morton, 1970). In addition, each of these models is capable of dealing with other criterial tests that have been said to identify separate memory systems: dissociations and stochastic independence (see also Ratcliff & McKoon, 1996).

At this point, the domains of implicit and explicit memory are related only by contrasts (this is implicit memory or it is explicit memory). But as models for implicit memory are developed, it is hoped that relationships between the two will become apparent (as in Schooler et al., 1998) and that theoretical progress will be made.

Connectionist Memory Models

Connectionist models might seem natural for the storage of information. Much of the early work in connectionist/neural network modeling did have a connection to memory. But there have been few recent attempts to take standard connectionist architectures and build new memory models. For a simplistic example, one might think that a multilayer connectionist network would be ideal for storing information. Vectors of features could be entered into the model and the system trained to respond positively when a learned item was presented for test. The problem is that the model will learn to respond positively for everything because it has not been presented with negative instances (people do not have to be given negative instances because they know that any item not from the study list should be given a negative response). Another possibility might be to have every encounter with a study item add to its strength in memory (e.g., something like Anderson's, 1991, matrix model). But if this growth leads to an increase in variability, then a list strength effect is predicted, contrary to data (see Ratcliff, 1990). The key thing to keep in mind is that data constrain and rule out many simple-minded translations of many connectionist representational schemes and learning algorithms as a basis for memory models.

However, because the current memory models are distributed, they offer the possibility of translation into connectionist terms. For example, McClelland and Chappell's model (1999) is couched in connectionist terms, but is almost identical in structure to the REM model. The REM model assumes that items are stored in independent vectors in memory and these are accessed in parallel at test. The McClelland and Chappell model assumes that each item stored has a node and weights from it to vector elements. This means, for each item, there is a separate weight for each element in the vector. The correspondence with REM can be seen if the weights are equated with features in the separate memory vectors in REM. The conclusion is that there is considerable overlap in representational assumptions between the connectionist model and both the newer and the older global memory models; it follows that the insights from one of the domains should be used in theoretical development in the other.

Conclusions

In many fields of sciences, theory and experimental work go hand in glove. In cognitive psychology, experimental work and theoretical work seem to have much less interaction. Often this is because as new experimental paradigms are introduced, a lot of fruitful experimental work can be performed to examine and test verbal hypotheses. Then, often after a great deal of experimental work has been performed, the questions become more about details of methods and design than about the larger questions that started the investigation. This results in a reduction in the amount of research in the domain, especially if some other interesting empirical domain has come to prominence. The questions that generate the most interest in the empirical approach are often not critical for testing or evaluating models. This means that the models may have little to say about the empirical phenomena; they may be able to easily explain the results without adding additional insight. This is one source of an often expressed sentiment among experimentalists, that it is impossible to falsify models. But, as described above, the last 15 or 20 years of development and testing of models present a different picture and show a progression in theory development and evaluation.

Models provide a means of going beyond the more traditional empirical approach in several ways. Models are needed when we want to address issues across different experiments, across different experimental paradigms, or across different subject populations. Also, models are needed to relate different dependent variables such as reaction time, accuracy, confidence judgments, recall accuracy, forced choice accuracy, and so on.

Models should aspire to the following properties: a model should be fairly comprehensive, covering data from a range of experimental tasks and a range of manipulations of independent variables. A model should not be a restatement of the experimental data—that is, it should produce a coherent explanation of the data and classify or organize the data differently from a simple empirically based classification, perhaps by showing invariances in parameter values, performance characteristics, or structures that cannot be seen in the experimental data. Finally, a model can gain considerable power if it can deal with more than one dependent variable at a time.

The models of memory of the 1980s were designed to achieve these aims. They were comprehensive and they were able to handle most of the experimental data within their domains. They were falsifiable in that their basic assumptions could be put to empirical test. Their failures led to the new models (McClelland & Chappell, 1999; Shiffrin & Steyvers, 1997, 1998), which are now, in their turn, ripe for evaluation and testing.

Acknowledgments This research was supported by National Institute of Mental Health grant HD MH44640 and National Institute for Deafness and other Communication Disorders grant DC01240.

References

Anderson, J. A. (1991). Why, having so many neurons, do we have so few thoughts? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 477–507). Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Washington, DC: Winston.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341–380.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing the speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York: Academic Press.

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2; pp. 90–195). New York: Academic Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory and Cognition, 13,* 8–20.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94,* 341–358.

McClelland, J. L., & Chappell, M. (1999). Familiarity breeds differentiation: A Bayesian approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724–760.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 128–148.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1155–1172.

McNamara, T. P. (1992a). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99*, 650–662.

McNamara, T. P. (1992b). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1173–1190.

McNamara, T. P. (1994a). Priming and theories of memory: A reply to Ratcliff and McKoon. *Psychological Review, 101*, 185–187.

McNamara, T. P. (1994b). Theories of priming: II. Types of primes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 507–520.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76*, 165–178.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Murdock, B. B. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609–626.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*, 839–862.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 855–874.

Neath, I. (1998). *Human memory: An introduction to research, data, and theory.* Pacific Grove, CA: Brooks/Cole.

Norman, D. A. (1970). *Models of human memory.* New York: Academic Press.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 3–27.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review, 97*, 285–308.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*, 385–408.

Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: Spreading activation theories versus compound cue theories. *Psychological Review, 101*, 177–184.

Ratcliff, R., & McKoon, G. (1995). Sequential effects in lexical decision: Tests of compound cue retrieval theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1380–1388.

Ratcliff, R., & McKoon, G. (1996). Biases in implicit memory tasks. *Journal of Experimental Psychology: General, 125*, 403–421.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104*, 319–343.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 163–178.

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 1–38). Cambridge, MA: MIT Press.

Schooler, L., Shiffrin, R. M., & Raaijmakers, J. (1998). *Theoretical note: A model for implicit effects in perceptual identification.* Manuscript submitted for publication.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). The list strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 179–195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*, 145–166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational*

Models of Cognition (pp. 73–95), Oxford University Press: Oxford, England.

Slamecka, N. J. (1991). The analysis of recognition. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock*. Hillsdale, NJ: Erlbaum.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195–231.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology, 92*, 297–304.