



UNA REVISIÓN DE MODELOS DE MEMORIA DESDE EL GLOBAL HASTA EL BAYESIANO A REVIEW OF MEMORY MODELS FROM GLOBAL TO BAYESIAN

Pablo Gomez¹
DePaul University, Chicago

RESUMEN

En éste trabajo, se exploran los fundamentos tanto empíricos como teóricos de algunos de los modelos de memoria más importantes. Una situación importante en la investigación sobre memoria fue el desarrollo de "Modelos Globales de Memoria". En estos modelos, los juicios acerca de eventos únicos están basados sobre la comparación de todas las trazas de memoria en la prueba. Tres de los modelos de memoria global más importantes son SAM (Gillund y Shiffrin, 1984) TODAM (Murdock, 1982) y MINERVA2 (Hintzman, 1986, 1988). Este trabajo se organiza en forma cronológica. Se lleva a cabo una revisión de los primeros modelos sobre memoria global antes de la llegada de los modelos globales sobre memoria. Luego se discuten los modelos globales sobre memoria. Y finalmente, se discuten algunos de los datos y modelos que son inconsistentes con los modelos globales sobre memoria, junto con las direcciones contemporáneas en los modelos sobre memoria, tales como Los Modelos Bayesianos sobre Memoria.

¹ Pablo Gomez, PhD, Assistant Professor, Psychology Department, DePaul University, Chicago, IL
Correo Electrónico: pgomez1@depaul.edu





ABSTRAC

In this paper, the theoretical and empirical foundations of some of the most important memory models are explored. An important stage in memory research was the development of “Global Memory Models”. In these models, judgments about single events are based on comparing all memory traces in memory to the probe. The three most important global memory models are SAM (Gillund and Shiffrin, 1984), TODAM (Murdock, 1982), and MINERVA2 (Hintzman, 1986, 1988). This paper is organized chronologically: first data and models before the advent of global memory models are reviewed; then global memory models are discussed; and finally some of the data and models that are inconsistent with global memory models are discussed along with contemporary directions in memory modeling like Bayesian Memory Models.

Memory Research Before Global Memory Models.

Models

Before global memory models were developed, memory models assumed that judgments about events were based only on the contents in memory that were associated with the events. These early models used “localist” approaches to modeling memory. Murdock (1974) distinguished between seven different localist approaches: temporally-structured models; the buffer model of Atkinson and Shiffrin (1968); organization theory (Mandler, 1967, Tulving, 1968); the address register model (Broadbent, 1971); sampling models (Shiffrin, 1970); the levels of processing model (Craik and Lockhart, 1972); and the associative network model by Anderson and Bower (1972). Each of these approaches are defined and reviewed, in turn, below. It is important to recognize that these approaches are not necessarily exclusive. Many were formulated to explain specific memory phenomena and none provides comprehensive accounts of human memory.

Models that are considered temporally structured posit a number of separate memory stores. These stores are differentiated based on the forgetting rate of the





contents, the format of information and their size. For example, it is common to use two stores: a short-term store (contents decay within a minute) and a long-term store (contents decay over the course of days or weeks). The models that follow the temporal-structure approach (e.g., Murdock, 1967) are often depicted as flow charts where different boxes represent each memory store. Any single item can be stored in any of the stores or in several stores simultaneously. Examples of this type of approach include Bower (1967), Glanzer (1972), Kintsch (1968, 1970) and Wickelgren (1970).

One of the most influential temporally structured models was proposed by Atkinson and Shiffrin (1968, 1971). In this model there are both structural and control processes. The former are related to fixed properties of the system (such as the memory stores) while the latter are related to features or strategies used by the subject in processing information. This model specifies the properties of the short-term store and the operations performed in it. Within the short-term store there is a rehearsal buffer composed by a limited number of slots. Incoming items are transferred from the sensory register to the buffer with a given probability. If the buffer is full, a new item displaces an old item in the buffer. The probability of an item being displaced is a function of the amount of time that item has already been in the buffer. The probability of a correct transfer from short-term storage to long-term storage is also a function of the amount of time that the item has been in the buffer.

Another model that was a further development of the temporally structured approach was developed by Broadbent (1969, 1970, 1971). Murdock (1974) characterized this model as an address-register approach. The model assumes four components: the sensory store, the primary memory, the secondary memory, and between the primary and secondary memories an address-register. This address-register holds information about the items, but not the items themselves. The contents of the address-register are conceptualized in the model as retrieval cues necessary during performance in memory tasks.

Although these temporally structured models and their extensions appealing, they fail to account for a variety of important memory phenomena. An important set of





effects that these models are silent about are the effects of semantic information in memory tasks. Organization theory was an approach that attempted to account for semantic effects, and that Murdock (1974) described as a point of view, not as a real theory. The idea behind this approach was that the semantic organization of memory is reflected in subjects' performance in memory tasks, such as free recall. Bousfield (1953), Deese (1961), and Cofer (1967) among others, demonstrated that people recall in clusters of words. In their experiments, these clusters were based on categorical and associative information.

The approaches described so far were primarily concerned with response probabilities. Response latencies are also an important dependent variable in psychological research. Some theoreticians were interested in free recall models developed to account for inter-response times. It is a well-documented phenomenon that the first few items that are recalled have shorter inter-response time than those recalled later. Sampling models (Shiffrin, 1970 & Rundus, 1973) were proposed to explain this difference in inter-response times. The assumption behind these models is that there is a search with replacement in memory. According to this assumption, items that are already recalled can be resampled, slowing down the recall of new items. The probability of resampling increases as the number of items recalled increases.

Another influential approach in the study of memory was originally proposed by Craik and Lockhart (1972). They elaborated a conceptual framework that criticized the temporally partitioned stores approach. Their main critique was that temporal models do not pay enough attention to the effects of different types of encoding. Craik and Lockhart demonstrated that the persistency of a memory trace is a function of how deeply processed an item is. For example, an item that is processed at a structural level would have a weaker image than an item that is semantically encoded.

The final approach considered by Murdock (1974) in his review is called "Free Recall by an Associative Network" (FRAN, Anderson & Bower, 1972, Anderson, 1972). In this model there is a memory structure in which the entries are linked to one another based on dictionary definitions and a few random associations (to reflect human





idiosyncrasies). When a word is presented, it is stored in the short-term store, which works like a buffer. The long-term memory representation of the item and the associative paths from the item to other list items are tagged with a certain probability. These tagging processes are sensitive to the presentation rate of the item. The probability of an item being tagged is a function of the duration of that item in the buffer. Besides the buffer, there is a short-term store called the ENTRYSET, in which the words with the most associations are stored to be used as starters for associative retrieval. Retrieval is achieved by a sampling-search mechanism. First, items in the in the ENTRYSET and buffer are randomly sampled to be used as retrieval cues; then, the pathways from the cue are used to find tagged items; therefore, to retrieve an item it has to be reachable by a tagged path.

Before global memory models appeared, there was debate among theoreticians on the types of processes involved in recall and recognition. In this context, Anderson and Bower (1972) and Kintsch (1974) proposed their models, which are classified as “embedded two process models” by Gillund and Shiffrin (1984). These two models assume that the recognition mechanism is a part of the recall process. There are two stages in recall: during the first stage, long-term memory is searched for possible candidates to be recalled; in the second stage, those candidates are recognized based on familiarity (Kintsch, 1974) or contextual association (Anderson and Bower, 1973). Recognition is a simpler task according to these models, because only the final stage is necessary.

Tulving (1976) and Gillund & Shiffrin (1984) criticized these models because they were unable to explain how recognition is affected by some of the factors that affect recall, such as organizational effects, repetition, and list length. Following his idea of encoding specificity, Tulving (1976) suggested that recall and recognition have a common mechanism that is cue dependent and involves a search process and an implicit recognition. Gillund and Shiffrin (1984), on the other hand, noted that fast recognition at high accuracy levels challenges Tulving’s model because it predicts a large speed-accuracy trade-off given the need for an extended search.





To conclude this section, it should be noted that before global memory models appeared, although there were a lot of theories, few of them were explicitly stated as models. In many cases they were just “ways to think about” memory phenomena (i.e., hypotheses); however, they were influential on guiding the type of experiments performed, which is the subject of the next section of this paper.

Empirical Research

Since the 50's, a rich set of data has been produced by recall and recognition experiments. Successful memory models have to account for it in a parsimonious and reasonably comprehensive manner. What follows is a review some of the most relevant experimental results that motivated the appearance of global memory models.

Effects in Free recall

In recall there are three first order effects (Murdock, 1974). The first one is that in multitrial free recall the acquisition can be characterized as a negatively accelerated exponential function. This is because the probability of recall increases rapidly after a few trials, and very slowly when the number of trials is larger. The second effect is the serial position effect (e.g. Murdock, 1962). Serial position manifests in the following way: a “U” shaped function of probability of recall by serial position. There is better recall of the first and last items in a list (first three and last eight in Murdock's data), than for the items in the middle of the list. The third effect is related to the time between the recollection of each word. Inter-response time increases with ordinal position in the output (e.g., Murdock & Okada, 1970). Interresponse time can be thought of as a function of the number of items yet to be recalled.

Besides these first order effects in recall, several variables and manipulation have been know to affect recall and recognition and that I will present next.





List length and presentation time.

A robust effect in recall and recognition is that performance decreases as the number of studied items increases. In recognition experiments, response latency (e.g., Ratcliff and Murdock, 1976; experiment 3) and response probability (e.g., Strong, 1912) have been shown to be affected by list length. Presentation time has a similar effect as list length in recognition (e.g., Ratcliff, 1978; Ratcliff & Murdock, 1976; Murdock, 1974; Shiffrin, 1970). For example, Ratcliff and Murdock (1976) presented lists of 15 items at different presentation rates. All the words within a list had equal presentation times (presentation time was a between list variable). Accuracy was higher for longer presentation times. In a similar experiment, in which presentation time was a within list variable, Ratcliff (1978) presented words at four different study times. RT decreased as a function of presentation time at a faster rate than in the Ratcliff & Murdock & experiment. An explanation for the difference in results between the Ratcliff & Murdock & (1976), and the Ratcliff (1978) experiments, is that subjects in the Ratcliff & Murdock experiment might have been able to adjust their response criteria across the different lists.

List length and presentation time have similar effects in recall as in recognition. Murdock (1960, 1962) analyzed the effect of list length, presentation time and serial position in a recall experiment. He found a list length effect that was larger than the effect of presentation time across all serial positions (i.e., the position of the item in the list). Murdock interpreted this data with a simple hypothesis called the “total-time hypothesis”. This hypothesis explained list length effects in terms of the relative study time devoted to each item. According to this hypothesis, less study time is devoted for each item in longer lists than in shorter lists. Roberts (1972) found a non-linear effect of presentation time that contradicts Murdock’s total-time hypothesis. In Robert’s experiments, performance increased dramatically if total study time increased when the presentation times were short. However, for longer presentation times, increased study time did not improve performance.





Encoding Effects

Several studies have been performed to test the effects of different encoding variables in recall and recognition performance. Perhaps one of the best known encoding effects is the one reported Craik and Lockhart (1972) regarding levels-of-processing manipulations. The basic idea behind the levels-of-processing approach is that the quality of the initial encoding is reflected by measurements of performance in memory tasks. Craik and Lockhart also suggested that the incidental learning paradigms might be better than intentional learning paradigms because in the latter there might be covert idiosyncratic processes. An important manipulation following this approach was made by Woodward, Bjork and Jongeward (1973). They gave subjects a signal to either remember or forget a word after variable amounts of time. In a later surprise test, subjects were tested for all the words, including those that they were told to forget. The assumption was that if a subject received the signal to remember a word, this word would have had more maintenance rehearsal than if the subject had received the “forget” signal. Woodward et al’s results differentiate between recall and recognition: maintenance rehearsal benefited recognition but not recall. This result was accounted for within a “levels of processing” framework by Glenberg, Smith and Green (1978). They claimed that Woodward et al’s result were a consequence of a deep rehearsal (type II) for those words that were signaled to be remembered. On the other hand, shallow rehearsal (type I) was used in those words that were signaled to be forgotten.

Another type of encoding effect that received a lot of attention is “encoding specificity” (Tulving & Thomson, 1973). Encoding specificity refers to the fact that memory improves if the coding at test phase matches the coding used during the study phase. This effect has been shown using several manipulations; for example, Light and Carter-Sobell (1970) shown participants related pairs of words (e.g., traffic—jam) during study. Then, in the test phase, participants had to make a recognition judgment for “jam” in the presence of “traffic” or in the presence of another word (e.g., “strawberry”) which is related to a homophone of jam. In accordance with the encoding specificity principle, performance was better when the same sense of the word was used during study and





test. In a related experiment, Tulving and Watkins (1973) gave subjects pairs of words. Some of them were strong associates (bark-dog) and some rhymed words (worse-nurse). In a cued recall test, 74% of the associative target words were recalled as well as 56% of the rhyming targets. During the second test phase, 30% of those targets that were not recalled with their original rhyming cue were recalled with a new associative cue; and 22% of the targets that were not recalled with their original associative cue were recalled with a new rhyming cue.

Perhaps the best known example of the encoding specificity effect is an experiment by Tulving and Thomson (1973). In it, participants studied pairs of words, then were asked to produce associates of new words such that the targets studied before had a large probability of being produced (e.g., if the pair “head-light” was studied, “dark” was used as the cue for associates in the second phase). In the third phase of the experiment, subjects were asked to recognize, among the words that they generated as associates, the targets that they had studied before. In the final phase, a list of cues was given to subjects and they were asked to recall the target words. Performance in the recall phase (62.5% correct recall) was better than in recognition (25% hit rate). Tulving (1974) interpreted these results by arguing that memory has two sources. The first source is the information laid down as a result of the original perception of the event; and the second one is the retrieval cue, which is information present in the individual cognitive environment at the time retrieval occurs. Forgetting is a cue-dependent phenomenon consequence of a lack of appropriate information in the retrieval environment rather than a change in the specific stored information. Tulving’s claims challenged the dual process models of recall (e.g., Kintsch, 1970; Bahrck, 1970). These models assume that recall consists of an implicit generation of candidates and the recognition process performed in those candidates. According to these models, recall of an item that could not be recognized before is impossible because recall requires information whose existence is denied by the recognition failure.



Word frequency

When words are used in memory tasks, the participants already have a representation of those words in memory before the experiment starts. Not all of the words are equally accessible, and word frequency is a good estimate of their accessibility. There is an interesting interaction between word frequency and the type of memory task. Recall for high frequency words is better than for low frequency words (e.g., Deese, 1960; Sumbly, 1963). However, if high and low frequency words are presented in the same list the word frequency advantage decreases or disappears (Duncan, 1974; Gregg, 1976). On the other hand, recognition is better for low frequency words than for high frequency words (McCormack & Swenson, 1972; Shulman & Lovelace, 1970). This effect on recognition holds even if high and low frequency words are mixed in the same list (Gregg, 1976). There are also word frequency effects in the distractors in recognition. McCormack and Swenson (1972) showed that the false alarm rate is higher for high frequency words than for low frequency words. Given the complexity of these interactions, word frequency effects are an important benchmark for memory models that try to account for both recall and recognition.

Global Memory Models

The models discussed in the preceding review share a localist approach to decisions about memory. The effects of context during study and test, as well as the richness of the data described in the previous section, forced theoreticians to develop new models. A new breed of model, called “Global Memory Models” was developed under the assumption that decisions about memory are achieved by matching or comparing the global contents of memory to the stimuli or context in which the memory task is performed.

The SAM Model

The first global memory model to be reviewed is the “Search of Associative Memory” (SAM) model (Raaijmakers & Shiffrin, 1980, 1981; Gillund & Shiffrin, 1984).





This model assumes that long-term memory is a matrix of memory images. These images represent three types of information: contextual information, used to identify the setting in which an item occurred; item information, that can be used to name and identify a memory trace; and inter-item information, which represents the association between items. The strength of the associations between the memory images and cues (e.g., context, or an item presented during a recognition test) form a retrieval structure. This retrieval structure is constructed based on the encoding operations whose description follows.

Encoding

SAM shares with temporally structured models the assumption of a primary memory and a secondary memory. Encoding starts in the primary memory. The model assumes a buffer of limited size r ($r = 4$ in Gillund & Shiffrin, 1984) in primary memory. When a new item enters the buffer, the oldest item is dropped. While items are in the buffer, they are rehearsed and encoded. Items that share time in the buffer have larger inter-item associations, whose size is a function of the amount of time they spend together in the buffer. Inter-item associations between two items did not share time in the buffer are equal to their pre-experimental association. Note that all these inter-item association are stored in secondary memory and represented in SAM's retrieval structure

Retrieval Structure

The values in the retrieval structure are a function of the rehearsal and coding processes during study and pre-experimental associations. In the model, items in the buffer share rehearsal time with the other items in the buffer. So, for each unit of time an item is in the buffer, it gets $1/n$ of the rehearsal time t (where n is the number of items in the buffer).

In the retrieval structure as described by Gillund and Shiffrin (1984), the columns of the matrix represent memory images, while the rows represent the cues. The intersection between a memory image and a cue is the value of the retrieval strength





between that cue and that memory image. The value of this strength is obtained the following way:

The context cue is represented in the first row of the retrieval structure matrix; the strength of association between context and the memory images is equal to the parameter \underline{a}^* times the rehearsal time. The diagonal of the retrieval structure represents the self-strength associations between each cue and their own memory image. This self-association value is equal to the parameter \underline{c}^* times the rehearsal time. The between item strength is the value of parameter \underline{b}^* times the number of units of rehearsal time that the items were together in the short-term buffer. The strength of association between a cue and a memory image that were not in the short-term buffer together at any time is equal to the value of the parameter \underline{d} . Note that the strength of association between a distractor used as a cue and a memory image is also equal to the \underline{d} parameter. Variability is added to the strength values according to a 3-point distribution. There is a probability of 1/3 that a value remains the same; a probability of 1/3 that a value increases by a factor of \underline{v} ; and a probability of 1/3 that a value decreases by a factor of \underline{v} .

Recognition

During recognition, a direct-access familiarity process guides decisions. In a recognition task, when an item is presented, it combines with the context of the experiment to form a composite cue. The association of these composite cues to all memory images is summed across images according to:

$$F(C, I_j) = \sum_{k=1}^N S(I_j, I_k) \quad (1)$$

where $S(C, I_k)$ is the association between the context cue and memory image k ; $S(I_j, I_k)$ is the association between the memory image j (the test item) and image k ; and F is termed familiarity.

Because $S(I_j, I_k)$ values tend to be higher for items than for distractors, the sums for items $F(C, I_j)$ also tend to be higher than the sums for distractors. A criterion is set in



a manner similar to signal detection theory, and if the familiarity of a cue is larger than the criterion, an “old” response is produced. If the familiarity is lower than the criterion, a “new” response is produced. Signal detection measurements of performance, such as d' can be calculated in the usual manner. d' would therefore be an estimation of the distance between the items' and distractors' familiarity distributions. These distributions tend to normality by effect of the central limit theorem.

Recall

Recall in SAM (Raaijmakers & Shiffrin, 1981; Gillund & Shiffrin, 1984), assumes a repeated sample & recovery process. In the first stage -- the sampling stage, the associations between the cue and all the images are used together as a probe set. In the first search cycle, only the context is used as a cue. The probability of sampling any item is calculated by Luce's (1959) choice rule. The ratio of the activation of a given image to the summed activation across all items is given by:

$$P_S(I_i|C) = \frac{S(C, I_i)}{\sum_{k=1}^N S(C, I_k)} \quad (2)$$

Once an item is sampled, the probability of recalling the name of the item given that only context was used as a cue is given by:

$$P_R(I_i|C) = 1 - \exp\{-S(C, I_i)\} \quad (3)$$

When an item has already been recovered, it can be used together with context to form the probe set. In that case, the sampling probability of an image is given by:

$$P^S(I_i | C, I_j) = \frac{S(C, I_i) S(I_j, I_k)}{\sum_k S(C, I_k) S(I_j, I_k)} \quad (4)$$

An image sampled with Equation 4, in which the previously recovered item and context form a compound cue, has a recall probability given by:

$$P_R(I_i|C, I_j) = 1 - \exp\{-S(C, I_i) - S(I_j, I_i)\} \quad (5)$$





The composite cue is used until a new item is recalled or the number of failures in a row reaches the value of the L_{\max} parameter, in which case the next sample uses context alone (using Equation 2).

Besides the calculations of sampling and recall probability, an important part of the model is its resampling mechanism. For each recall failure and for each resampling of an already recalled item the counter k increases by one. When k reaches the value of the K_{\max} parameter, recall terminates. Note that the image of an item that has already been recovered can be resampled, just as in the resampling models (Shiffrin, 1970; Rundus, 1973) described before. This feature of the model allows it to make correct predictions about inter-response times as a function of output order.

List Length.

In the SAM model, the familiarity of an item is computed by adding the products of the strength of association between the cues and the memory images (Equation 1). If a longer study list is presented, the number of memory images increases. Because most of these images would not share rehearsal time with the image of the recognition item (i.e., they were not together in the buffer), they add values associated with the d parameter. When a new item is presented as a probe, all the strengths of association between memory images and this probe are also related to the d parameter. If the list length increases, the global familiarity increases by the same amount for both items and distractors. Performance is worse because longer lists add additional terms to the sum in Equations 1 and 3, and each of these terms adds independent variance components. The distractor and target distributions, although equally apart from each other as in shorter lists, overlap more in longer lists given that they have larger variances.

A different mechanism explains the effect of list length in recall. The effect is located in the sampling part of the model. The probability of sampling any item decreases as list length increases. This is because the probability of sampling a cue is given by Equations 2 and 4, and the denominator in this equation increases as list length increases. Recovery probabilities (Equations 3 and 5) do not change with list



length, but the probability of sampling an image highly associated with a cue decreases while the probability of sampling a distal image increases.

Presentation time

When items are studied for longer time, performance improves in recall and recognition. The SAM model captures this effect because increasing the presentation time increases the association strengths (context, inter-item and self-strength) that are built-up during rehearsal. Therefore, in recognition the mean of the familiarity distribution increases, making the difference between the mean of the distractor and the mean of the target distribution larger. In recall, sampling probabilities remain the same, because the effect of presentation time is on both the numerator and denominator of Equations 2 and 4. However, the probability of recovery increases with presentation time because the $S(C, I_j)$ component in Equations 4 and 5 increases.

Encoding effects

Intentionality of encoding.

Gillund and Shiffrin (1984) suggested that intentional encoding associated to increased expectancy of recall improves inter-item encoding at the expense of self encoding. This claim is supported by the model's ability to predict the correct interaction between intentionality of encoding and type of task (recall or recognition). Estes & DaPolito (1967) increased inter-item association by means of the instruction they gave to participants. They found that the instructions increased recall but did not increase recognition. In addition, when instructions were changed from incidental to intentional, recall performance improved, but recognition decreased.

Maintenance rehearsal

In the Woodward et al. (1973) experiment, maintenance rehearsal increased recognition but not recall. The assumption made by Gillund and Shiffrin was that maintenance rehearsal increases self-strength, but reduces inter-item and context





coding. This is because the item being rehearsed displaces other items from the buffer. The implementation of this assumption in SAM involves an increase in the \underline{c} parameter and a small decrease in the \underline{b} parameter. This increases the mean of the familiarity distribution for rehearsed items, improving recognition.

Because self-sampling is increased, in the recall model, the chances of sampling a new image are reduced. Each time an already recovered item is resampled, the \underline{k} and \underline{l} counters increase; therefore, the \underline{K}_{\max} and \underline{L}_{\max} criteria are reached faster.

Encoding specificity

Tulving & Thomson (1973) demonstrated that the match in semantic context between study and test is a predictor of performance. In a related experiment, Light and Carter-Sobell (1970) manipulated semantic context during the study phase introducing biasing words (e.g., strawberry or traffic) as adjectives to target words (e.g., jam). During a recognition test, the target words were tested in the presence of the same or a different biasing word. Performance was better when the biasing word was the same during study and test. Given that the SAM model has a cue-dependent retrieval mechanism (see Equation 2), it implements Tulving's cue-dependent retrieval view. SAM handles encoding specificity effects with two mechanisms: one is based on the coding of the test item and the other is based on the familiarity of the test pair. The first one is implemented by changing the values of the encoding parameters \underline{b} and \underline{c} . These parameters change because the encoding of an item is determined in part by the stimulus environment, and encoding during study and test time could be different. The second mechanism can be exemplified with the Light and Carter-Sobell experiment: if the biasing word is the same during test and study phases, the biasing term is strongly linked to itself and to the target, adding to the global familiarity.

Distractor similarity effects

When one or some of the new probes presented during a recognition test are semantically or orthographically similar to one or more of the targets, these distractors



are harder to identify as “new”. Gillund and Shiffrin (1984) reviewed a number of studies and concluded that physical similarity is more important than semantic similarity; however, SAM cannot differentiate between semantic and physical similarity. The SAM model assumes a different value for the d parameter when the similarity between the new and the old items increases. The d parameter is related to the strength of association between the image of two words that were not together in the study buffer. When the association between distractors and memory images of studied words increases, the variance and the mean of the familiarity distribution for distractors also increases, decreasing d' .

Word frequency effects

Performance for high frequency words is better in recall, but worse in recognition relative to low-frequency words. In the SAM model, high frequency words are thought to have higher associations than low frequency words. These associations are captured by the parameters b and d (note that self strength and context do not change as a function of word frequency). High frequency words have a d value of d_h and a b value of b_h ; on the other hand, low frequency words have a d value of d_l and a b value of b_l ($d_h > d_l$ and $b_h > b_l$). With these differential parameter values, SAM can account for the word frequency-task interaction. In recall, the superiority of high frequency words is related to the fact that the recovery probabilities are higher for high frequency words than for low frequency words. In recognition, the residual values of the items increases as d and b increase. This increases the spread of the signal distribution, decreasing d' .

The MINERVA2 MODEL

Hintzman (1986, 1988) proposed a different model he termed MINERVA2. This memory model shares the assumption of a global matching with SAM, but the two differ in other aspects of representation and process. MINERVA2 is primarily concerned with long-term memory (secondary memory). It does not provide an account of short-term, unlike SAM. MINERVA2 assumes that primary memory communicates with long-





term memory reading an image of the cue, or retrieving a memory image or “echo” to produce a response. In MINERVA2, memory traces are assumed to be composed of a vector of primitive features. These features can only take values of 1, -1 and 0. The values of 1 and -1 occur with equal probability and represent the relevant information of a stimulus. Information about a feature might be irrelevant or not encoded by the system; these features have a value of 0. Each memory experience leaves behind its own trace, even if it is a repeated one (i.e., when words are repeated in the study phase of a memory experiment, each presentation is stored as a separate memory trace). The memory traces in secondary memory are not perfect copies of the events in primary memory. Features are probabilistically encoded with probability \underline{L} of being encoded correctly per unit of time, and $1-\underline{L}$ of being encoded as a 0. The \underline{L} parameter is related to the quality of the encoding processes.

Recognition

When a probe is presented for recognition, it is matched in parallel to all memory traces. The activation of each trace is a function of its degree of match with the probe, which is computed by:

$$S(i) = \sum_{j=1}^N \frac{P(j)T(i, j)}{N}, \quad (6)$$

where $\underline{P}(j)$ is the value of feature j in the probe and $\underline{T}(i,j)$ is the value of the j th feature in the i th trace. The values that $\underline{S}(i)$ can take range between -1 and 1. When the probe and a trace are identical, the value of $\underline{S}(i)$ is equal to 1. On the other hand, when the trace and the probe are orthogonal, the value of $\underline{S}(i)$ is 0. Negative values have no special meaning in the model.

To preserve the sign of $\underline{S}(i)$, and to reduce the influence of several poorly matched traces that would be equivalent to one close match, $S(i)$ is transformed by

$$\underline{A}(i) = \underline{S}(i)^3. \quad (7)$$



This transformation enhances the strong matches, and reduces the activation values of weak matches, increasing the contrast between weak and strong items.

The sum of the activation across all traces can be thought of as the combined activation (or as Hintzman calls it: "echo intensity") and it is given by:

$$I = \sum_{i=1}^M A(i), \quad (8)$$

where M is the total number of traces in memory.

Two important properties of these equations should be pointed out. The first is that the expected value of I for a new probe is 0. The second is that this expected value is unrelated to M , while the standard deviation is.

Recall

MINERVA2 was formulated as a recognition model, however, based on Hintzman's (1988) suggestions a recall model can be envisioned. In the recall model, the representation of all the items in a list would share a set of "context features". During recall, the context features would be presented as a probe, and the traces activated to a threshold would be recalled. A slightly more sophisticated model would involve a sampling and retrieval mechanisms similar to those described in the SAM model for recall.

List Length.

MINERVA2 accounts for the effect of list length assuming that neither P (the probability of correctly encoding a feature) nor N (the number of features per probe) changes across different list lengths. However, the overlap between the signal and noise distributions does increase. This is not related to a smaller difference in the means of the distributions, because the expected value of $A(i)$ for a mismatch is 0. It is related to an increase in the unstudied to studied variance ratios (Gronlund & Elam, 1994). For each additional item in the list, there is an independent extra component in the variance of the familiarity distribution.



Presentation time

Presentation time is captured in MINERVA2 by modifying the value of the L parameter. Higher values are associated with longer presentation times. Changes in the L parameter produce a change in the mean of the I distribution for the old items. This is because there would be more non-zero values in Equation 7.

Encoding effects

MINERVA2 is a model of how information flows from primary to secondary memory, and how it is stored and retrieved from secondary memory; therefore, the model's ability to account for encoding effect can be evaluated. The model has only one parameter related to encoding: L . This limits the model's ability to account for interactions such as those found in intentionality of encoding manipulations (e.g., Estes & DaPolito, 1967). Furthermore, the recall model is not fully implemented, and testing the model's abilities to account for differential effects in recall and recognition cannot be done until a more elaborated recall model is available.

A few encoding effects can be explained adding assumptions to the recognition model. For example, to account for encoding specificity effects with MINERVA2, a set of "semantic context" features could be added to the studied items and the test probes. If the semantic context was the same during the study and test phases, the semantic context features of the test probe and the memory image would match, resulting in an increase in the value of L .

Another encoding phenomenon that MINERVA2 can account for, is the effect of similarity between items and distractors. If the items and distractors were similar, they would share some features (i.e., they would not be orthogonal). This would increase the expected value of the echo intensities for new probes. Note that for MINERVA2 there is no difference between semantic and physical features of the stimuli, and it would be impossible to model the difference between these two types of information.



Word frequency effects

MINERVA2 does not have a straightforward mechanism to account for word frequency effects. Hintzman (1986, 1988) suggests an ad hoc assumption to account for word frequency effects in recognition. Low frequency words have more salient properties than high frequency words, and are better encoded into secondary memory because they have higher \underline{L} values. Recent models like REM (Shiffrin & Steyvers, 1997), and McClelland and Chappell's (1998) use a representation mechanism similar to the one proposed in MINERVA2. These models account for word frequency effects in recognition through their assumption about how the values of the features are selected, and how this selection could be related to word frequency. Later in this paper, these new models will be reviewed in detail.

The TODAM MODEL

The third memory model I present is the TODAM model (Murdock, 1982, 1983). Most memory models, including those reviewed so far in this paper, assume that memory traces are stored separately from each other. TODAM, on the other hand, assumes that all items are stored in a single common memory vector. Another unique feature of the TODAM model is that it has a decision mechanism that has been further developed (Hockley & Murdock, 1987) to account for response times.

A single event of item is represented in TODAM as a vector of attributes. The value of each attribute is drawn from a normal distribution with mean of 0 and variance $1/\underline{N}$. The expected value of the dot product of two independent vectors equals to 0 ($\underline{E}[\underline{f} \bullet \underline{g}] = 0$). A vector of attributes representing an item is added to the global memory vector if the item is studied alone. When items are studied in pairs (associative information), the vectors representing the two words, plus their convolution are added to the global memory vector using:

$$\underline{M}_i = \alpha \underline{M}_{i-1} + \gamma_1 \underline{f}_i + \gamma_2 \underline{g}_i + \omega (\underline{f}_i * \underline{g}_i), \quad (9)$$

where \underline{f} and \underline{g} are the members of the pair, $*$ is the convolution operator, γ_t represents the weighting parameters for item information; ω is the weighting parameter for the





associative information and α is the forgetting parameter. Metcalfe & Murdock (1981) suggested that there are two "levels" of memory, and that the global memory vector is located in the M level, while the items are stored in the I level. The transfer between these two levels is not perfect, because encoding is assumed to be an error prone process. There is a probability p (where $0 < p < 1$) of encoding a particular element correctly, and a probability $(1-p)$ of encoding that element as 0.

Recognition

This model bases its recognition decisions on the dot product of the test probe and the memory vector. This dot product can be thought of as the "familiarity" dimension over which a signal detection (Swets & Green, 1961) type of mechanism is applied. The input to the decision mechanism is influenced by noise during the decision process. To simulate noise in the model, random variability is added to the dot product. This sum is compared to two response criteria (Atkinson & Juola, 1973, 1974). If the sum is smaller than the low criterion, a fast "new" response is produced; if it is larger than the high criterion, a fast "old" response is produced. If the value is between the two criteria, more independent random noise is added until the sum reaches either criterion. It is assumed that the difference of the values of the criteria decreases as a function of time; therefore, a response must eventually occur.

Each noise sampling cycle is assumed to take an increasingly longer time, this allows the model to account for the positive skew in RT distributions. The model provides a detailed explanation of how a reaction time is estimated. According to this explanation, there are two components that are convoluted to form a reaction time. The first component is the decision time that is derived from the following equation (Hockley & Murdock, 1987):

$$T_{(k)} = (k^2 + k + 2)BCT, \quad (10)$$

where I is the decision time, k denotes the number of previous cycles, and BCT is the base cycle time, which is the only parameter and is a scale factor. The second





component in the convolution is a normal distribution that represents the non decision processes, referred as TOS (time for other stages). Hockey & Murdock fitted this decision model to data from the convolution analysis presented by Ratcliff & Murdock (1976).

The decision stage outlined before has a limited effect on response probabilities and d' . In TODAM, as in any other model that has a signal detection type of decision mechanism, there are two factors related to recognition performance (as measured by d'). The first is one is the distance between the means of the signal and noise distributions. The second one is the ratio between the variances of the familiarity distributions for items and distractors. When the ratio increases, the overlap between the two distributions also increases, reducing d' . The parameters of TODAM affect these two factors in different ways. The parameter N (number of attributes) affects the size of the variances; large N s produce smaller ratio variances than small N s. On the other hand, the α and γ parameters affect the distance between the means and the variance ratio.

Recall

In the TODAM model, recall is achieved by a mechanism more complicated than the one proposed for recognition. The correlation between the memory vector and the cue is used during retrieval. Correlation is the inverse operation of convolution; the correlation between \underline{f} and \underline{f} convoluted with \underline{g} ($\underline{f} \# (\underline{f} * \underline{g})$) equals \underline{g}' , a vector similar to \underline{g} . This vector, \underline{g}' is the recalled item.

The way that item vectors are constructed make the correlation function as a retrieval mechanism: the expected value of the dot product of two unrelated unrelated items is 0 ($E[\underline{f} \bullet \underline{g}] = 0$), while the expected value of the dot-product of an item vector to itself is one ($E[\underline{f} \bullet \underline{f}] = 1$). When items are studied in pairs (e.g., \underline{f} and \underline{g}), the correlation of a member of the pair with the global memory vector results in vector similar to the other member of the pair ($\underline{g} \# \underline{M} = \underline{f}'$). Since the vector obtained with correlation is not exactly the vector that was studied, Metcalfe and Murdock (1981) suggested that the dot



product of this vector and all the vectors of the studied items should be obtained. If the highest value is that of $\underline{f} \bullet \underline{f}$, then a correct recollection is assumed to happen.

List length

According to closed form solutions particular of Weber (1988). In TODAM, both variance and mean familiarity are affected by list length. This is a problem according to Gronlund and Elam (1991), who point out that in order to fit the list length data, the parameters \underline{N} , α and γ have to be changed in rather arbitrary ways. For example, in some conditions the \underline{N} parameter (number of features) has to decrease for the short list condition to make to correct prediction. This assumption seems unreasonable. For this reason, TODAM does not provide an accurate description of the list length effects.

Presentation time

Murdock proposed a mechanism that can account for presentation time. Longer presentation times correspond to higher γ values. This increases the mean familiarity for both item and associative information. Although the variance also increases, it does so at a slower rate than the mean.

Encoding effects

TODAM's assumptions about encoding can be adjusted to try to account for encoding effects. This adjustments can be reflected in the values the features of the vectors, and in the values of the γ parameters.

Intentionality of encoding.

Estes and DaPolito (1967) showed that increased inter-item association, which is related to the γ_3 parameter in Equation (1), increases recall but not recognition. TODAM would have problems with this differential effect, given that a higher γ_3 value would increase both recall and recognition. This or any other effect that dissociates between recall and recognition, is hard to account for by TODAM.





Maintenance Rehearsal

Woodward et al (1973) have shown that maintenance rehearsal improves recognition but not recall. Mapping maintenance rehearsal into TODAM's parameters is not straightforward. The parameter BCT in Equation 10 is the only one associated only with recognition and not with recall; however, a reason why maintenance rehearsal should affect this parameter is not transparent. If maintenance rehearsal was related to the γ_i or ω parameters, the model would not predict differential effects on recall and recognition.

Encoding specificity

A similar mechanism as the one proposed during the discussion of MINERVA2, could be used by TODAM to account for encoding specificity effects. Some of the features or attributes of the memory vectors could contain contextual information; therefore, if these features match between the study and test phases, a stronger dot product between the probe and the memory vector would be found. This would increase performance in both recall and recognition, which is in accordance to the published data.

Tulving's cue-dependent assumptions are also implemented in TODAM (Ratcliff & McKoon, 1989). Failure to recognize recallable items is accounted by the model because the associative and item information are independent from each other.

Distractor similarity effects

A straightforward way to implement similarity between distractors and items is by having them share some features. This would increase the familiarity values of those distractors similar to the items and would increase the overlap between the item and distractor distributions of familiarity, decreasing d' .



Word frequency effects

The original formulation of TODAM does not have a way to account for word frequency effects. However, as described earlier, there are new models share with TODAM and MINERVA2 the assumption of memory images represented as a vector of features (Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998), and are able to account for word frequency effects. The values of the features are related to word frequency. A detail explanation of these models will be provided by the end of this paper.

Applications and Problems of Global Memory Models.

Although each of the models has particular strengths, the three models reviewed so far share the assumption of a global familiarity in memory and a response mechanism for recognition based on signal detection. Research that appeared after these model were developed was aimed at testing fundamental properties of the models. Some of the data collected proved to be hard to account for by the models. This challenged the core notions behind global memory models as well as specific assumptions of each model. In the next section, empirical effects that are particularly problematic for global memory models are reviewed: [a] time course of processing, [b] z-ROC curves, [c] the mirror effect, and [d] the null list-strength effect

Time Course of Processing

Speed-accuracy trade-off functions

One limitation of global memory models is that they are silent about time course of processing (excluding the update to TODAM, Hockley & Murdock, 1987). SAM, and MINERVA2 can not predict RT performance, although a random walk process (Link & Heath, 1972) or diffusion process (Ratcliff, 1978) could possibly be implemented as a decision process for these models. However, there is data that would be difficult to account for by global memory models even if a successful interface with a diffusion or





random walk process is accomplished. Some of this data was presented by Gronlund and Ratcliff (1989). They explored the time course of the recognition process of individual items and associative information. Their experiments used a response signal procedure (Doshier, 1981; Reed, 1973, 1976; Schouter & Bebbler, 1967; Wickelgren, 1977). In this procedure, subjects are instructed to respond within 200-300 ms after a signal is presented. The signal is presented after the onset of the stimulus at times determined by the experimenter. Performance (measured with d') can be plotted as a function of the lag between the stimulus and the signal + the RT after the signal. Accuracy increases monotonically and can be fitted using a negatively accelerated exponential function (Reed, 1976; Wickelgren, 1977; Doshier, 1981):

$$d'(t) = d'_a(1 - e^{-v(t - T_{er})}). \quad (10)$$

Also, an expression derived from the diffusion model (Ratcliff, 1978) could be used to fit the speed –accuracy trade-off curve:

$$d'(t) = \frac{d'_a}{\sqrt{1 + \frac{v}{t - T_{er}}}}, \quad (11)$$

where v is the rate of approach to the asymptote and also the ratio of the variance of the drift rate in the diffusion process, to the variance of the retrieval probe-memory relatedness distribution.

The parameters of these functions have psychological interpretations. T_{er} is the intercept, which is a measure of the minimum processing time necessary to for accuracy to begin to rise above chance; d'_a is the asymptotic performance; and v is the rate at which performance reaches asymptotic level.

Gronlund and Ratcliff's (1989) primary concern was the difference between the time course of processing of associative and item information. Global memory models assume that item information and associative information arise from a common mechanism, and should have similar intercept and rate of accumulation even if the asymptotic level of performance is different. In one of Gronlund and Ratcliff's



experiments, participants studied pairs of words. Then, participants were asked to recognize either pairs or single words. They were supposed to respond “old” if a pair of words consisted of two words from the study list even if they were not studied together (called AB’ pair). They were also instructed to respond “old” to pairs of word that were studied together (AB) as well as to individual items studied in the list. The response signal function of items studied together versus items not studied together had a larger intercept than the function of any other pair of stimuli type. This implies that associative information begins to be available later than item information (the T_{er} parameter is higher for the AB-AB’ function than for any other function). This result challenges global memory models.

In SAM, familiarity is given by the summed associations between the item (or pair of items in associative memory) and all the memory images. This mechanism cannot predict differences in the time of availability for item and associative information because both types of information are used at the same time. MINERVA2 models associative memory with the assumption that a probe consisting of two items makes a single vector of features twice as large as a single item’s vector. As Gronlund and Ratcliff (1989) point out, the contribution of item and associative are inseparable parts of the total activation.

The TODAM model has an important difference from MINERVA2 and SAM. In TODAM, memory is represented as a single vector to which each item’s trace is added during storage. For pair learning, the two items and the convolution of the items are added to the memory vector. Recognition is accomplished by comparing the value of the dot product between the memory vector and the probe to a high and low criterion. If the probe is a pair of items, they are also convoluted and the result of the convolution is treated as the probe; therefore, item and associative information go together into the convoluted probe, which is inconsistent with Gronlund and Ratcliff’s (1989) results. However, a simple modification could be envisioned. It would involve matching the two items separately for item information, and using the convolution only for associative



information. This way, it could be possible to predict differences in the intercept and rate of accumulation between item and associative information.

Non-monotonicity in the accumulation of information

There is experimental evidence that indicates a non-monotonic accumulation of information. This is at odds with the predictions of global memory models. Ratcliff & McKoon (1982, 1989) and Doshier (1984) found that in short lags, subjects make judgment errors when the information available early in the process is inconsistent with the information available later. Another result that seems to challenge the equal intercept and equal rate prediction made by global memory models is one presented by Doshier (1988), who found that associative and item information become available at the same time but have different rates.

In another experiment by Gronlund and Ratcliff, participants were instructed to respond “new” to AB’ pairs (two words studied but not in the same pair); therefore, they had to use associative information to discriminate between AB and AB’ pairs. Interestingly, the probability of responding old to AB’ pairs grows until 600 ms; then it goes down (note that the correct response to this type of pair is “new”). After 600 ms, the associative information becomes available; therefore, the drift of the diffusion process changes.

Global memory models were not designed to account for all time course effects, but the problems posed by Gronlund and Ratcliff’s results go beyond pure time course concerns. They challenge the notion of associative and item information being accessed at the same time and by the same mechanism.

Z-ROC Functions

Memory models should be able to accurately predict ROC curves. ROC curves can be constructed using payoffs, different ratios between old and new items, or by asking participants to rate the confidence of their responses. For each level of either of these manipulations, the hit and false alarm rates are calculated and serve as





coordinates of a point in the ROC space. These points form a ROC curve. The shape of the signal and noise distributions can be inferred z -transforming ROC curves. Given the assumption of normality of the signal and noise distributions, z -ROC curves are predicted to be straight lines that can be characterized by their slope and their y -intercept. The intercept of the best fitting line provides an estimate of d' and the slope is an estimator of the ratio between the standard deviations of the new (unstudied) and old (studied) distributions.

Ratcliff, Sheu and Gronlund (1992); Ratcliff, McKoon and Tindall (1994); and Gronlund and Elam (1994) examined empirical z -ROC functions using a variety of manipulations. Across several experiments, they found linear z -ROC functions, which is in accordance to the assumptions made by signal detection theory and by global memory models. However, the most important result for its implications for memory models is related to the slope of the z -ROC functions. Gronlund and Elam (1994) found that the slopes of the z -ROC functions were not different to 1, and remained constant across different list lengths. The constant slope is a problematic result for global memory models, because it implies that the ratio of signal and noise variances is unrelated to list length. As explained earlier in this paper, according to global memory models list length affects the ratio of signal to noise variance, of which the slope of a z -ROC curve is a measure. Ratcliff et al. (1992) manipulated list strength and found that when d' is larger than 0.5 the slopes of z -ROC functions remain constant at a 0.8 value. While the three models fail to predict accurately Ratcliff et al's results, they do it for different reasons. SAM and MINERVA2 predict that the slope of the z -ROC should decrease as a function of strength; while TODAM fails to predict the correct value for the slope. TODAM predicts a slope close to 1.0 (i.e., the familiarity of old and new items should have equal variances according to TODAM).

List-strength

List strength refers to the strengthening of certain items by studying them longer or by repeating them. Tulving & Hastie (1972) showed that recall of repeated items





improves at the expense of non-repeated items. As described in the first section of this paper, Ratcliff (1978) found that manipulating presentation times within a list had a larger effect in RT than if presentation time is a between list variable (as in Ratcliff & Murdock, 1976). Ratcliff, Clark & Shiffrin (1990) were interested in the effect of differential study time within a list and its implications for global memory models. Global memory models predict that memory for strong items should be greater in mixed lists than in a pure-strong list. Weak items, on the other hand, should be better in the pure condition. This prediction emerges from the decision mechanism assumed by these models. The explanation of this prediction was outlined by Shiffrin, Ratcliff and Clark (1990). Performance, as measured by d' is given by:

$$d' = \frac{E[F|i] - E[F|x]}{\sqrt{\text{Var}[F|x]}}, \quad (12)$$

where $E[F|i]$ is the expected value of familiarity when studied items are tested, $E[F|x]$ is the expected value of the familiarity when a distractors are tested, and $\text{Var}[F|x]$ is the standard deviation of the familiarity distribution when distractors are tested. Recognition performance in the mixed -pure manipulation can be denoted $\underline{d}'(\text{pure strong})$, $\underline{d}'(\text{pure weak})$, $\underline{d}'(\text{mixed strong})$, $\underline{d}'(\text{mixed weak})$. Consequently, the denominators in Equation 12 can be denoted as $\underline{\sigma}(\text{list type-strength})$, and the denominators as $\underline{\sigma}(\text{list type-strength})$. Global memory models predict that if the items are blocked with items of similar strength in the study list, then $\underline{u}(\text{pure strong}) = \underline{u}(\text{mixed strong})$ which are larger than $\underline{u}(\text{pure weak}) = \underline{u}(\text{mixed weak})$. In other words, the denominator is a function only of the target's strength, and it is unrelated to the composition of the list. The value of the denominator is unrelated to the strength of the items because it is the variance of the familiarity distribution when a distractor is tested. So, $\underline{\sigma}(\text{mixed-strong}) = \underline{\sigma}(\text{mixed-weak})$ because in both cases the list is the same.

The heart of the list strength effect prediction in global memory models comes from the following equations, which are just ratios of Equation 12:





$$\frac{d'(ps)}{d'(pw)} = \frac{u(ps)/\sigma(ps)}{u(pw)/\sigma(pw)} \quad (13 a)$$

and

$$\frac{d'(ms)}{d'(mw)} = \frac{u(ms)/\sigma(ms)}{u(mw)/\sigma(mw)}. \quad (13 b)$$

To estimate the size of the list strength effect the ratio of these two ratios can be obtained:

$$R_r = \frac{d'(ms)/d'(mw)}{d'(ps)/d'(pw)} = \frac{\sigma(ps)}{\sigma(pw)}. \quad (14)$$

R_r is greater than 1 if there is a list strength effect, and equal to 1 if there is a null list strength effect.

Models that predict a larger variance in the pure strong list than in the pure weak list will predict a list-strength effect. And since these variances can be decomposed into the sum of independent components (e.g., $\sigma(\underline{ps}) = \sqrt{N(\text{var}[i_s])}$, where i_s is the i^{th} item of the strong list),

$$R_r = \sqrt{\text{Var}(i_{STRONG}) / \text{Var}(i_{WEAK})}. \quad (15)$$

Models that decompose the variances in this way and predict larger variance associated with strong than with weak items, predict a list strength effect. For the reasons explained above and during the presentation of the models, SAM, TODAM, and MINERVA2 predict R_r to be larger than 1.

Ratcliff et al. (1990) found no evidence in favor of a list strength effect in recognition when the strength was manipulated using different study times. They found a list strength effect using repetition in free and cued recall but not in recognition. These



results are problematic for global memory models. The original formulation of these models predicted list-strength effects. Shiffrin et al. (1990) explored modifications to TODAM, MINERVA2 and SAM that could potentially help them account for the null list-strength effect. Only SAM could be modified to satisfactorily account for the data.

Among the global memory models, SAM is the only one that proposes a separate mechanism for recall and recognition. Two different cue weightings are used in recall and recognition. Ratcliff et al. (1990) argue that this is because in recall, context focuses more in the item trace than in recognition, which is based on global matching. The difference between adding items and adding strengths is that strong items are stored in a single memory image. Shiffrin, Ratcliff & Clark (1990) proposed a “differentiation” process within SAM. It consists in two complementary forces in the activation of strong traces by unrelated items: activation from the context cue increases while activation from the item cue decreases. This implies that when the association between a memory image and its list context is strengthened, the associations between the memory image and all other potential stimuli are weakened. This is maybe because the differences between a strong item and the unstudied items become more salient; therefore, when in recall only the context cue is used, there is an increase in activation. This differentiation process has been challenged by data from Ratcliff, McKoon & Tindall (1991). In their first two experiments they used fast presentation times, and their results are problematic for the differentiation approach. McClelland & Chappell, (1998) and Shiffrin & Steyvers (1997, 1998) have recently used a Bayesian decision approach to implement the notion of differentiation. In the next section, I will describe these models.

Rational Models

The Rational Approach

Over the last few years, models that share rational assumptions and a Bayesian decision mechanism have appeared. Anderson (1990) introduced the notion of a Bayesian decision mechanism for memory. An instantiation of Anderson’s ideas can be





found in a model proposed by Schooler and Anderson (1997). I will not go through all the details of the model, but I will mention some of its assumptions. It shares the Bayesian approach with other recent models. In its formulation, its authors revisited Estes's (1955) Stimulus Fluctuation Theory, in which the focus of the explanations shifted from internal process in the organism to the statistical properties of the environment. The other theoretical antecedent of Anderson & Schooler's model is Anderson's (1990) ACT-R framework. According to ACT-R, the function of memory is to solve a database retrieval problem. The rational approach claims that if we can define the problem that the cognitive system is trying to solve, as well as what the optimal solution to that problem is, the behavior of the system should correspond to that optimal solution. For example, the elements in context constitute an optimal query to long-term memory that retrieves the memories that are most likely to be relevant for the given situation.

The basic architecture assumed by the rational analysis is that information is stored in discrete memory structures or traces in long-term memory. Recall results from searching through the structures stored in long-term memory. The model assumes that there is unlimited capacity in long-term memory. The system tries to find the memory structure that is relevant for the current context; and evaluates the probability that each structure is needed. The order of the search processes of memory structures is related to their need probabilities. The system recognizes when the target structures are retrieved and correctly rejects structures that are not needed.

McClelland and Chappell Model

McClelland and Chappell (1998) proposed a rational model that is related to the differentiation process described in the "list strength" section of this paper, and might be interpreted as a specification of how that process occurs. The model rests on the assumption that as subjects become more familiar with an item they get a clearer sense of its features. This model shares with MINERVA2 and TODAM the assumption about memory images as collections of features. A memory experience consists in a pattern of activation \underline{S} across a set of feature units. Each feature of an item gets its values from





probabilistically generated 0's and 1's. Recognition can be characterized as the conditional probability of experiencing each feature, given that the "psychological experience" was produced by an item. The likelihood of experiencing a set of features given that the item is an instance of a given item μ is given by:

$$p(S|I^\mu) = \prod_i (\xi_i^\mu)^{S_i} (1 - \xi_i^\mu)^{(1-S_i)}. \quad (16)$$

S_i represents the presence (1) or absence (0) of feature i . ξ represents the probability of experiencing feature i in item μ . This equation means that if a feature is active, its conditional probability is multiplied; if it is not active, the complement of its probability is multiplied. Because the model assumes that subjects compute a subjective likelihood of the experience given an instance of item μ , Equation 16 changes to:

$$\pi(S|I^\mu) = \prod_i (w_i^\mu)^{S_i} (1 - w_i^\mu)^{(1-S_i)}, \quad (17)$$

where $\pi(S|I^\mu)$ is the subjective estimate, and w is the subjective conditional probability. Familiarization refines these subjective conditional probabilities. This increases the probability identifying familiar items because the subjective likelihood of the experience being produced by an item increases. It is important to note that this refinement also decreases the probability of incorrectly identifying familiar items; ergo it is not just a bias.

Using this assumption, McClelland and Chappell (1998) proposed a single-item recognition memory model that consists on a collection of item detectors (one for every studied item). Each detector estimates the conditional probabilities for item features during study and compares the to the estimates at test to make an old-new decision.

The structure of the model in which they implement this idea consists in a two-layer network in which the bottom layer is the feature layer in which the pattern S occurs. According to assumptions whose mechanism not already defined by the model,



there is a unique item detector in the top layer for each item. Each time a new item is encountered, a new item detector on the top layer is assigned to it. On the other hand, if an old item is repeated, the same unit used in previous presentations of the same item is selected.

The item detectors decide if a pattern during test was produced by the same item as the one corresponding to that detector. Two hypothesis can be considered: \underline{S} was generated by the conditional probabilities for the μ^{th} item, or it was not. The two hypotheses form the following odds' ratio:

$$\frac{\pi(I^\mu | S)}{\pi(\bar{I}^\mu | S)} = \frac{\pi(S | I^\mu)\pi(I^\mu)}{\pi(S | \bar{I}^\mu)\pi(\bar{I}^\mu)}. \quad (18)$$

The decision can be made setting a criterion, and if the ratio is above the criterion an “old” response is produced. I^μ is the hypothesis that \underline{S} was generated by the conditional probabilities for the μ^{th} item; and \bar{I}^μ is the hypothesis that \underline{S} was not generated by the conditional probabilities for the μ^{th} item. Note that this calculation is based on subjective estimates of the probabilities, and that each detector is blind to the probabilities computed by the other detectors.

Equation 16 can be rewritten as:

$$\pi(I^\mu) = \pi(I^\mu | list)\pi(list) + \pi(I^\mu | \bar{list})\pi(\bar{list}). \quad (19)$$

For a real item (i.e., not a distractor) $\pi(I^\mu | \bar{list}) = 0$ and $\pi(I^\mu | list) = 1/\Lambda$, where Λ is the subjective estimation of the number of items in the list; therefore,

$$\pi(I^\mu) = \frac{1}{\Lambda} \pi(list). \quad (20)$$

On the other hand $\pi(I^\mu | \bar{list})$ is equal to probability of the pattern being randomly generated. Because $\underline{w}_a = \underline{fp}_1 + (1 - \underline{f}) \underline{p}_0$ is the average probability of the units being active, the subjective conditional of \underline{S} not being generated by a real item is given by:



$$\pi(S|\bar{I}^\mu) = \prod_i w_a^{S_i} (1 - w_a)^{(1-S_i)}. \quad (21)$$

Since $\pi(\bar{I}^\mu) = \frac{\Lambda - 1}{\Lambda}$, the $\frac{\pi(I^\mu)}{\pi(\bar{I}^\mu)}$ component in Equation 18 can be reduced to

$\frac{\pi(list)}{\Lambda - \pi(list)}$; and in turn, Equation 18 can then be reduced to:

$$\frac{\pi(I^\mu | S)}{\pi(\bar{I}^\mu | S)} = \frac{\pi(list)}{\Lambda - \pi(list)} \prod_i \left(\frac{w_i^\mu}{w_a} \right)^{S_i} \left(\frac{1 - w_i^\mu}{1 - w_a} \right)^{(1-S_i)}. \quad (22)$$

The logarithmic of Equation 8 gives the estimated log-odds $[\lambda^\mu(S)]$, which is the basis for the decision in the model. It is approximately normally distributed; therefore, it can be analyzed with signal detection theory.

The model assumes a neural network setup for its implementation, and use Rumelhart and Zipser's (1986) learning rule:

$$w_i^\mu(t+1) = w_i^\mu(t) + \epsilon [S_i(t) - w_i^\mu(t)]. \quad (23)$$

$w_i^\mu(t)$ estimates the conditional probability of the i^{th} unit being active given that the μ^{th} item is active. In this learning rule the weight w will increase if the i^{th} unit is active (to a limit of 1) at a given time. Similarly, it will decrease to a limit of 0, if the i^{th} unit is not active. ϵ is the learning rate.

The weights between the feature nodes and the item recognition units have a belief distribution for the conditional probability, which McClelland and Chappell assume to be a Beta distributed with parameters \underline{a} and \underline{b} . The mean of this distribution is equal to $\underline{a} / (\underline{a} + \underline{b})$. During learning, after observing a pattern of activation \underline{S} the mean of the distribution is updated to $(\underline{a} + \underline{S}) / (\underline{a} + \underline{b} + 1)$, which can be transformed to:

$$w_i^\mu + \epsilon [S_i - w_i^\mu], \quad (24)$$



where $\frac{a}{(a+b)}$ is related to \underline{w}^{μ}_i and $1/(a+b)$ to ϵ , so that the learning rate for the s^{th} presentation is

$$\epsilon_s = \frac{1}{n_0 + S} \quad (25)$$

The mechanism just described yields a differentiation process like the one described by Ratcliff et al. (1990). The basic notion is that as item I^1 is learnt the probability of the patterns not corresponding to I^1 decreases. The log-odds distributions of studied and unstudied items separate because the mean of studied items increases at a much faster pace than its standard deviation; therefore, the response from the detector will be different to instances of I^1 than to instances of any other item. At the beginning, activation is based on a log normal distribution. To add variability to the model, some features are not encoded and keep their original random value.

List length effect

In this model, the decision is based in the maximum estimated log odds. For distractors this maximum is a maximum across identical distributions: one distribution for each detector (in longer lists there are more detectors). For items, the detector corresponding to that item, should have a larger mean than the rest of the detectors.

The list length effects operate on these distributions. One factor is the $\frac{\pi(list)}{\Lambda - \pi(list)}$ in

Equation 8. Recall that Λ is the estimated number of items; therefore, this term decreases as the number of items increases, which reduces the estimated log-odds from short to long lists. Another factor related to list length is the number of detectors. As the number of detectors grows, the maximum estimated log-odds also increases, inflating the false alarm rate. Since most of the items' maximum estimated log-odds comes from their item detector, the number of item detectors should not have a large impact on the hit rate.





List-strength and presentation time effects

To McClelland and Chappell, both d' and β are important data to fit in their simulation. β is generally neglected by researchers because it represents criterion setting, which is not related to memory per se. However, not only criterion setting can affect β . If the mean or standard deviation of either or both signal and noise distributions change, β will also change. This type of change was observed in Murnane and Shiffrin (1991) and replicated in McClelland and Chappell's simulations, in which $\ln[\beta]$ increases as strength increases even if the criterion has a fixed value.

The model is also consistent with the null list-strength effect. This is because of the model's differentiation mechanism. Strengthening the μ^{th} item reduces the estimated log-odds $\lambda^{\mu}(\underline{S})$ when \underline{S} is not generated by item μ , reducing the false alarm rate.

z-ROC curves

This model has two ways to produce z-ROC curves. One is related to the w_c criterion, and the other is by varying $\pi(\text{list})$, which could be thought of as an effect of changing the proportion of new and old items. Using either method, the slopes of the z-ROC curves are close to .8, which is the value found by a number of researchers before (Shiffrin & Murnane, 1991; Ratcliff, McKoon & Tindall, 1994). List strength and list length manipulations affect d' , but do not substantially affect the slope of the z-ROC. This model is able to keep a constant z-ROC slope because changes in d' are related to changes in the mean of the log-odd distribution, not to a change in the variance. There is currently some debate over the relationship of d' and the slope of the z-ROC curve.

Differentiation

Recent data reported by Hintzman and Curran (1995) might show evidence of a differentiation process consistent to the one described by McClelland and Chappell's model. In Hintzman and Curran's Experiment, nouns appeared once, three times, 8 times or 20 times during study. A judgment of frequency task followed the study phase.





When presented with similar words to those studied, false recognition was fairly high even if the related word was presented up to 20 times. A more detailed exploration of the data reveals that there is a U-shaped effect. False recognition of similar items increases as the number of presentations of studied items increases from one to three. However, if the studied item was repeated more than three times, false recognition of similar items slowly decreases. McClelland and Chappell (1998) claim that this U-shaped performance is a consequence of the differentiation process.

Word frequency effects

Low frequency words produce a higher hit rate and a lower false alarm rate (see Balota & Neely, 1980). This has been called the frequency mirror effect. In addition, $\ln[\beta]$ is larger for low frequency words, and the z -ROC slopes are smaller for low frequency words. McClelland and Chappell assume that word frequency is reflected on the feature content of their representation. Representations of low frequency words are assumed to be less variable than representations of high frequency words. This is because low frequency words have “more distinct definitions”. The model can capture this assumption by setting the p_0 parameter to a higher value for high frequency words.

This model can account for the frequency mirror effect and for the shallower z -ROC slopes for low frequency words. However, when fitting the model to a Ratcliff et al's (1994) experiment, in which a list-strength and a word frequency were manipulated, the model predicted a strength effect on the z -ROC slopes, which did not occur in the data.

REM

At the same time as McClelland and Chappell were developing their model, Shiffrin and Steyvers (1997) were developing a similar memory model called Retrieving Effectively from Memory (REM). The agenda for REM was similar to that for McClelland and Chappell: to account for the null list-strength effect, the z -ROC constant slope, and the mirror effect.





In REM memory consist of images; each image is composed by a vector of feature values. Information about a feature is represented by positive integers. These values differ in their environmental base rates. The distribution of these environmental base rates was chosen to be geometric with parameter g :

$$P[V = j] = (1 - g)^{j-1} g, \quad j = 1, \dots, \infty \quad (26)$$

The parameter of the geometric distribution is related to word frequency. High frequency words are assumed to have more common (smaller) feature values generated with a g_H value. Lower frequency words have less common feature values, generated with a g_L value ($g_H > g_L$). As in MINERVA2, a value of 0 in a feature means absence of knowledge about a feature.

During study, a memory image is stored for each word in the study list. However, the image is not an exact copy of the studied vector. For each unit of time an item is studied, there is a u^* probability of storing a nonzero value for each feature, given that nothing has been stored yet for that feature. If a nonzero value is going to be stored, the value is going to be correctly copied from the studied vector with probability c . With probability $1-c$, the feature value is going to be chosen randomly using Equation 26. Note that the correct value could be chosen by chance.

During test, when a probe is presented, it is compared in parallel with all the memory images. The value of matches between the probe and each of the memory images is noted ignoring those features in which the memory image has a value of 0.

For any given probe, there are two types of memory images. The first one, designated as d -image, include those images of items other than the probe. The second type of image, called s -image, is the image stored during a previous presentation of the probe.

The likelihood ratio λ_j is calculated for each probe. This likelihood is the probability of finding matches for the values in which the matches were found if the



image was an s-image, divided by the probability of finding those matches if the image was a d-image. This likelihood ratio can be calculated by:

$$\lambda_j = (1-c)^n \prod_{i=1}^{\infty} \left[\frac{c + (1-c)g(1-g)^{i-1}}{g(1-g)^{i-1}} \right]. \quad (27)$$

The recognition decision is given by the odds (Φ) in favor of an old item over a new item:

$$\Phi = \frac{1}{n} \sum_{j=1}^n \lambda_j. \quad (28)$$

If there are no biases, odds larger than 1 ($\Phi > 1.0$) produce an “old” response.

List length.

All the models reviewed in this paper can predict list length effects on d'. The real test for the models is whether they can predict the constant slope in z-ROCs across different list lengths. Those models that account for list length effects on d' modifying the variance of the signal or noise distributions predict large changes in the slope of the z-ROC. REM is not among these models. In REM, list length effects are predicted because each additional word in the study list increases the chance of matching distractor test probes. The Bayesian approach allows REM to make two important predictions: a constant z-ROC slope even if the variance of the log odd rises, and the mirror effect.

List-strength and presentation time effects

Presentation time effects are easily predicted by REM increasing the number of storage attempts for longer presentations. Storing more features increases the likelihood ratios for s-images, and decreases the ratios for d-images.

Throughout this paper, the null list-strength effect has been an important test for models. The null-list length effect refers to performance for weak items is not being hurt





when strong items are in the list, compared to a pure weak list. Shiffrin et al. (1990) proposed a differentiation mechanism for SAM. The structure of REM produces a differentiation mechanism automatically: stronger items have more features stored, making them less likely to be confused with the test word when the image of such strong item is a d' -image. As in list-length, the Bayesian approach allows REM to predict a mirror effect and a small variation in the slope of the z -ROC functions regardless of strength or presentation time.

Word frequency effects

Low frequency words are easier to recognize than high frequency words. REM is able to make the correct prediction about word frequency effects on recognition assuming that the distribution of feature values is different between low and high frequency words. High frequency words are assumed to have features with more probable values than low frequency words. This makes low frequency words harder to confuse or misidentify. Again as mentioned before, the Bayesian approach allows REM to predict a mirror effect, and a constant slope of the z -ROC curves.

A word on optimality

An important difference between global memory models and REM is that the former were mechanistic models that were not concerned with optimality. On the other hand, in REM optimality is assumed beforehand, and performance is based on knowledge by the subject of the distribution of psychological states. Subjects are supposed to use this knowledge in an optimal way when responding in a recognition task. A related approach to the one taken by REM is taken by McClelland and Chappell, with the difference that subjects do not have access to the real likelihood, but to flawed subjective estimates. This difference between the two models does not lead to differential predictions on the effects of the manipulations with which this paper is concerned. Future research could test these two competing models generating differential predictions.





Time course of processing

The two new models that have been described are able to account for some of the phenomena that global memory models could not predict. However, there is an issue that has not been addressed: the time course of memory processes. These new models are still silent about RT measurements of performance. An adequate memory model should be able to account for probabilities of response as well as mean RT and even RT distributions.

A result that is problematic for rational memory models is the one reported by Gronlund and Ratcliff (1989). They found that associative information is available after single item information is available. The two models, McClelland and Chappell, and REM, assume the same mechanism for associative and item information. Only a handful of memory experiments deals with time course issues; however, the timing of memory process could be another ground to test memory models.



References

- Anderson, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation (Vol. 5)*. New York: Academic Press.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97-123.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
- Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy in word recognition. In S. Kornblum (Ed.), *Attention and Performance (Vol. 4)*. New York: Academic Press.
- Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology (Vol. 1): Learning, memory, & thinking*. San Francisco: Freeman.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory (Vol. 2)*. New York: Academic Press.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, *August*, 82-90.
- Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576-587.
- Baird, H. P. (1970). Measurement of Memory for Prompted Recall. *Psychological Review*, *77*, 215-222.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229-240.





- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), **Coding processes in human memory**. Washington, DC: Winston.
- Bower, G. H. (1967). A descriptive theory of memory. In D. P. Kimble (Ed.), **Learning, Remembering and forgetting**. New York: New York Academy of Sciences.
- Broadbent, D. E. (1969). Communication Models for Memory. In G. A. Talland & N. C. Waugh (Eds.), **The pathology of Memory**. New York: Academic Press.
- Broadbent, D. E. (1970). Psychological aspects of short-term and long-term memory. **Proceedings of the Royal Society, Series B, 158**, 222-231.
- Broadbent, D.E. (1971). **Decision and Stress**. New York: Academic Press.
- Cofer, C. N. (1967). Does conceptual organization influence the amount retained in immediate free-recall? In B. Kleinmuntz (Ed.), **Concepts and the Structure of Memory**. New York: Wiley.
- Craik, F. I. M. & Lockhart, R. S. (1972) Levels of Processing: A framework for memory research. **Journal of Verbal Learning and Verbal Behavior, 11**, 671-684.
- Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. **Psychological Reports, 7**, 337-344.
- Dosher, B. A. (1981). The effects of delay and interference: A speed-accuracy study. **Cognitive Psychology, 13**, 551-582.
- Duncan, C. P. (1974). Retrieval of low-frequency words from mixed lists. **Bulletin of the Psychonomic Society, 4**, 137-138.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. **Psychological Review, 62**, 369-377.
- Estes, W. K. & DaPolito, F. J. (1967). Independent variation of information storage and retrieval process in paired-associate learning. **Journal of Experimental Psychology, 75**, 18-26.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. **Psychological Review, 91**, 1-65.



- Glanzer, M., & Adams, J. K (1990). The mirror effect in recognition memory: Data and theory. ***Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13**, 8-20.
- Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. ***Journal of Verbal Learning and Verbal Behavior*, 16**, 339-352.
- Gregg, V. H. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), ***Recall and recognition***. London: Wiley.
- Gronlund, S. D. & Elam, L. E. (1994). List-Length effect: Recognition Accuracy and Variance of Underlying Distributions. ***Journal of Experimental Psychology: Learning, Memory and Cognition*, 20**, 1355-1369.
- Gronlund, S. D. & Ratcliff, R. (1989). Time course of item and associative information: Implications for Global Memory Models. ***Journal of Experimental Psychology: Learning, Memory and Cognition*, 15**, 550-561.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. ***Psychological Review*, 93**, 411-428.
- }
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. ***Psychological Review*, 95**, 528-551.
- Hintzman, D. L. & Curran, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. ***Journal of Experimental Psychology: General*, 126**, 228-247.
- Hockley, W. & Murdock, B. (1987). A decision model for accuracy and response latency in recognition memory. ***Psychological Review*, 94**, 341-358.
- Kintsch, W. (1968). Recognition and free recall of organized lists. ***Journal of Experimental Psychology*, 78**, 481-487.
- Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Eds), ***Models of human memory***. New York: Academic Press.
- Kintsch, W. (1974). The representation of meaning in Memory. Hillsdale, NJ: Erlbaum.
- Light, L. L. & Catter-Sobell, I. (1970). Effects of changed semantic context on recognition memory. ***Journal of Verbal Learning and Verbal Behavior*, 9**, 1-11.



- Link, S. W. & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.
- Luce, R. D. (1959). *Individual Choice Behavior: A theoretical analysis*. New York: Wiley.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- Mandler, G., & Boeck, W. (1974). Retrieval processes in recognition. *Memory & Cognition*, **2**, 613-615.
- Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, **10**, 33-42.
- Mandler, G., Pearlstone, Z., & Koopmans, H. J. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, **8**, 410-423.
- McCormack, D. P. & Swenson, A. L. (1972). Recognition Memory for Common and rare words. *Journal of Experimental Psychology*, **95**, 72-77.
- McClelland, J. & Chapell, M. (1998). Familiarity Breeds Differentiation: A Subjective-Likelihood Approach to the Effects of Experience in Recognition Memory. *Psychological Review*, **105**, 724-760.
- Metcalfe, J. & Murdock, B. B., Jr. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, **20**, 161-189.
- Murdock, B. B. (1960). The immediate retention of unrelated words. *Journal of Experimental Psychology*, **60**, 222-234.
- Murdock, B. B. (1962). The serial position effect in free recall. *Journal of Experimental Psychology*, **64**, 482-488.
- Murdock, B. B. (1967). Recent development in short-term memory. *British Journal of Psychology*, **72**, 421-433.
- 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. (1983). A distributed memory model for serial-order information. ***Psychological Review*, 90**, 316-338.
- Murdock, B. B. (1970). Interresponse times in single-trial free recall. ***Journal of Experimental Psychology*, 86**, 263-267.
- 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.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), ***The psychology of learning and motivation (Vol. 14)***. San Diego, CA: Academic Press.
- 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., Clark, S., & Shiffrin, R. M. (1990). The list strength effect: I. Data and Discussion. ***Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16**, 163-178.
- Ratcliff, R. & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. ***Journal of Experimental Psychology: Learning, Memory, & Cognition*, 8**, 16-36.
- Ratcliff, R. & McKoon, G. (1989). Memory models, text processing, and cue-dependent retrieval. In Roediger, H. L. III & Craik, F. I. M. (Eds), ***Varieties of memory and consciousness: Essays in honour of Endel Tulving***. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ratcliff, R. & McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. ***Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20**, 763-785.
- Ratcliff, R. & Murdock, B. B. Jr. (1976). Retrieval Processes in recognition memory. ***Psychological Review*, 83**, 190-214.
- Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. ***Psychological Review*, 99**, 518-535.



- Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. **Science**, *181*, 574-576.
- Reed, A. V. (1976). List length and the time course of recognition in immediate memory. **Memory and Cognition**, *4*, 16-30.
- Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. **Journal of Experimental Psychology**, *92*, 365-372.
- Rumelhart, D. E. & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart & J. L. McClelland, & the PDP Research Group (Eds.), **Parallel distributed processing: Explorations in the microstructure of cognition, Vol1**. Foundations. Cambridge, MA:MIT Press.
- Rundus, D. (1973). Negative effects using list items as recall cues. **Journal of Verbal Learning and Verbal Behavior**, *12*, 43-50.
- Schooler, L. J. & Anderson, J. (1997). The role of process in the rational analysis of memory. **Cognitive Psychology**, *32*, 219-250.
- Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), **Models of human memory**. New York: Academic Press.
- Shiffrin, R. M. & Steyvers, M. (1997). A model for recognition memory: REM--retrieving effectively from memory. **Psychonomic Bulletin & Review**, *4*, 145-166.
- Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list strength effect: II. Theoretical Mechanisms. **Journal of Experimental Psychology: Learning, Memory, and Cognition**, *16*, 179-195.
- Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. **Psychological Review**, *19*, 447-462.
- Sumby, W. H. (1963). Word Frequency and Serial Position Effects. **Journal of Verbal Learning and Verbal Behavior**, *1*, 443-450.
- Tulving, E. (1968). When is recall higher than recognition? **Psychonomic Science**, *10*, 53-54.
- Tulving, E. (1974) Cue-dependent forgetting. **American Scientist**, *62*, 74-82.





- Tulving, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), **Recall and Recognition**. London: Wiley.
- Tulving, E. & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. **Journal of Experimental Psychology**, **92**, 297-304.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. **Psychological Review**, **80**, 352-373.
- Tulving, E., & Watkins, M. J. (1973). Continuity between recall and recognition. **American Journal of Psychology**, **86**, 739-748.
- Wickelgren, W. A. (1970). Multitrace strength theory. In D. A. Norman (Ed.), **Models of human memory**. New York: Academic Press.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. **Acta Psychologica**, **41**, 67-85.
- Woodward, A. E., Bjork, R. A., & Jongeward, R. H., Jr. (1973). Recall and recognition as a function of primary rehearsal. **Journal of Verbal Learning and Verbal Behavior**, **12**, 608-612.

[REGRESAR A ÍNDICE](#)

