Constraints on Models of Recognition and Recall imposed by Data on the Time Course of Retrieval

Peter A. Nobel and Richard M. Shiffrin
Indiana University
Bloomington, IN 47405
pnobel@ucs.indiana.edu
shiffrin@ucs.indiana.edu

Abstract

Reaction time distributions in recognition conditions were compared to those in cued recall to explore the time course of retrieval, to test current models, and to provide constraints for the development of new models (including, to take an example, the class of recurrent neural nets, since they naturally produce reaction time predictions). Two different experimental paradigms were used. Results from a free response procedure showed fundamental differences between the two test modes, both in mean reaction time and the general shape of the distributions. Analysis of data from a signal-to-respond procedure revealed large differences between recognition and recall in the rate of growth of performance. These results suggest the existence of different processes underlying retrieval in recognition and cued recall. One model posits parallel activation of separate memory traces; for recognition, the summed activation is used for a decision, but for recall a search is based on sequential probabilistic choices from the traces. Further constraining models was the observation of nearly identical reaction time distributions for positive and negative responses in recognition, suggesting a single process for recognition decisions for targets and distractors.

Introduction

Neural net and connectionist models have focused more on storage and representations than on retrieval, yet the number of retrieval modes, and the nature of each, is of crucial importance to modelers of memory. In the present research we explore the time course of retrieval in order to ask whether recognition and cued recall are carried out by similar mecha-nisms (in studies that match the response time requirements of the two tasks), to ask whether positive and negative recognition responses are the result of a single process, and to explore the dynamics of memory access. The issue is of con-temporary interest given that many neural net models, particularly of the recurrent variety, provide natural response time predictions.

In a typical long-term memory experiment, subjects are presented during a study phase with a list of items that has to be remembered. In a recall test phase, subjects have to generate the items of the previous studied list in either a random order, i.e. free recall, or a fixed order denoted by the presentation of cues, i.e. cued recall. In a long-term recognition test phase, subjects are presented with words that were either on the list (targets), or that are new (distractors). The subject's task is to identify a word as "old" or "new".

Recognition and recall are improved (for both reaction times and accuracy) by increased study time (see e.g. Ratcliff & Murdock, 1976), decreased list length (see e.g. Roberts, 1972), and lessened delay and/or shortened distractor task between study and test. However, the possibility of different retrieval mechanisms for the two tasks is heightened by several other findings: 1) The use of words having higher natural language frequency increases recall, but decreases recognition (see Hall, 1979). 2) With instructions for maintenance rehearsal, recognition improves (Glenberg & Adams, 1978), but recall is not much affected (see e.g. Dark & Loftus, 1976). 3) Strengthening some list items (by extra study time or extra repetitions) harms the free recall of other items, slightly reduces cued recall of other items, and has no effect on, or even slightly

This research was supported by grant AFOSR90-0215 to the Institute for the Study of Human Capabilities, and by grants NIMH 12717 and AFOSR 870089 to the second author.

helps, the recognition of other items (Ratcliff, Clark, & Shiffrin, 1990).

Memory models

Models that assume the same processes to underlie recall and recognition predict (in their simplest form) the same reaction time distributions, or at least the same shaped distributions, for the two conditions. Models that assume different processes, like the Search of Associative Memory (SAM) model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) can predict markedly differing distributions.

In SAM, each item is stored in memory as a separate image. The images contain different kinds of information that is rehearsed and coded together in short-term store (Raaijmakers & Shiffrin, 1981). Items are retrieved from long-term store through the weighted strength of association between retrieval cues and stored images. In particular, a given image's activation is determined by the multiplication of the weighted strengths between each cue and that image.

Recognition involves a global familiarity process, in which familiarity results from a single parallel process of activation of all images. Memory is probed with two cues: the context cue and the tested item. The familiarity of the probe is defined as the activation caused by the two probe cues, which is the sum of the activations of all the memory images. This value is compared to a criterion chosen by the subject, and "yes"-responses are made when the familiarity value is higher than this criterion. Such a model predicts sharply peaked response time distributions and similar distributions for 'old' and 'new' responses.

Recall involves an extended serial search, with two phases: sampling and recovery. Again, memory is probed with context and item cues. The probability of sampling a particular image is its activation strength divided by the sum of the activations of all images. After sampling, the information in the image, which is used for the decision and response, must be recovered. The key is that this process continues over and over until a response is found or the subject gives up. Such a model predicts response times spread out over long time periods, and different distributions for correct recalls, intrusions, and the time to 'give up'.

The Composite Holographic Associative Recall Model (CHARM) (Metcalfe Eich, 1982, 1985) is an example of a model that assumes the same retrieval processes underlying recall and recognition. In CHARM, items are represented as feature vectors and are stored in a convolution memory. If pair A - B is presented, the convolution of vectors A and B (A*B) is a vector; it is added to the convolution of A with itself (A*A), and the convolution of B with itself (B*B), and all three are added to the accumulated memory vector for all studied pairs (if not all pairs ever studied).

There is one retrieval process. It operates by correlating the probe vector with the memory vector. In a recall task, the output of this process, a vector itself, is compared to a separate list of words in memory and the response will be the best match above a certain cut-off of activation. In a recognition task, the dot product of the output of the correlation process with the probe is taken, and a positive response is made if the match is above a criterion. Because CHARM treats recognition the same as recall it does not predict differences in the latency distribution for the time to retrieve the trace; any differences would have to be differences in the post-retrieval processes of matching in recognition, or matching in recall.

Numerous memory models share this property that differences in retrieval time distributions for recognition and recall would have to be due to post retrieval operations; e.g. TODAM (Murdock, 1982), Matrix Model (Pike, 1984), MINERVA (Hintzman, 1988), and various connectionist and neural net models (e.g. McClelland & Rumelhart, 1985).

Reaction Times

The literature concerning reaction time (RT) in long-term memory research is mainly restricted to the recognition paradigm. For example, Ratcliff and Murdock (1976) found increasing RT for both hits and correct rejections as a function of output (test) position, decreasing RT as a function of input (study) position, increasing RT when presentation time increases, decreasing RT when the number of presentations increase, and increasing RT as a function of list length.

Some evidence supporting the notion of a sequential search in free recall was collected by Murdock and Okada (1970): Interresponse times increase in a positively accelerated function as recall progresses, interresponse times were shorter the more words were left to recall (for a fixed output position), and at any given output position the interresponse time is a good predictor of the number of words left to recall.

Thus there are data concerning reaction times in recognition and recall tasks separately; there do not seem to be reaction time data when both tasks are given to the same subjects in similar paradigms. In addition, Ratcliff (1978) has argued that testing of models requires closer looks at the reaction time distributions than their central tendencies. He suggests that at a minimum models should account for the shape of reaction time distributions (in particular their skewness), and specify the relationship between speed and accuracy. Ratcliff and Murdock (1976) in fact fit their observed RT distributions with a convolution of exponential and normal distributions. Ratcliff (1978) then fit his model to the parameters of these fitted distributions.

For these reasons a series of studies was designed, using several methodologies to measure reaction times, looking at the effect of several variables in recognition and cued recall conditions, and measuring the entire reaction time distribution.

Experiments

Ten subjects were presented in the study phase with a list of pairs of high frequency words that had to be remembered. The test phase consisted of either single item recognition or cued recall. In the recognition condition, the subject's task was to say whether the test item was on the list , and in the recall condition the subject's task was to recall the other word of the pair. Varied were list length (10 vs. 20 pairs) and presentation time (2 vs. 6 seconds). In order to equate the demand characteristics of the tasks as much as possible, subjects had to press one of two keys when they recognized or recalled, and press the other key if they did not; in the case of recall, a positive response had to be followed by the typing of the word recalled, allowing us to assess accuracy. Two different response procedures were employed. In the free response procedure, subjects were asked to respond as quickly and as accurately as possible after presentation of the test item. This procedure is commonly employed, but suffers from the possibility that subjects might adopt different strategies (e.g. differing biases to respond quickly) in recognition and recall. In the signal-to-respond procedure , which controls for these strategy differences, the subjects were told not to respond until a signal was given (a tone) and then to respond at once (within 300 ms). The delays until the signal ranged in ten steps from 100 ms to 4500 ms.

Results Bearing on Recognition/ Recall Differences

We give representative results because a complete accounting would literally require hundreds of figures. The demonstrated findings hold for the conditions not shown (unless otherwise stipulated). Figure 1 shows the reaction time distributions for correct recognitions of old words (hits) and for correct recalls: The recall distribution has a larger mean, larger variance, larger skewing, and extends over the entire time course of retrieval.

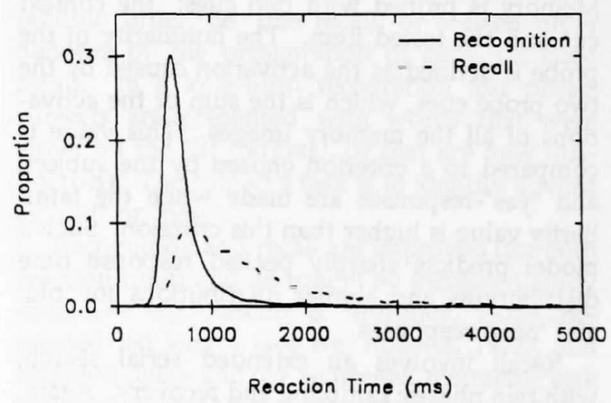

Figure 1. Reaction time distributions for correct recognitions (mean RT=710; st. dev.=299); median=630), and correct recalls (mean RT=1386; st. dev.=769; median=1163).

Figure 2 shows the reaction time distributions for incorrect recognitions of new items (false alarms), and for recalls of list items from other pairs, or, less commonly, non-list items (all termed intrusions): false alarms in recognition have a relatively low mean reaction time and variance, whereas intrusions in recall seem to have an almost uniform distribution.

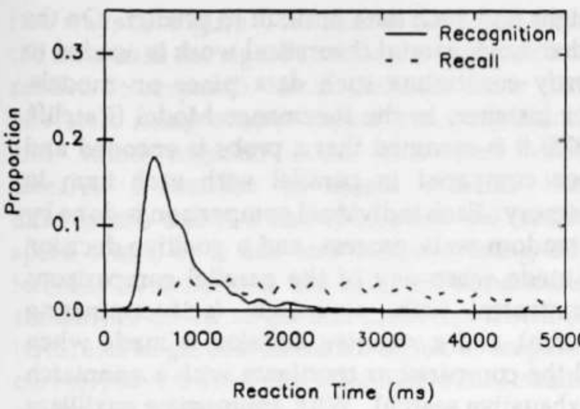

Figure 2. Reaction time distributions for false alarms (mean RT=818; st. dev.=439; median=689), and intrusions (mean RT=2381; st. dev.=1142; median=2320).

The signal-to-respond results can be used to assess the possibility that these large differences might be due to differing strategies or biases in recognition and recall. Presumably the subject will respond with whatever information is available at the time of the signal, whether recognition or recall is being tested. This procedure produces data of a somewhat different sort: The growth of accuracy is measured as a function of the signal delay.

Examination of typical retrieval functions for recognition memory shows an initial period of chance performance, followed by a period of rapid increases in accuracy, and finally, as retrieval time is further increased, accuracy reaches asymptote (see e.g. Dosher, 1984). These functions can be described by an exponential approach to an asymptote with 3 parameters: an asymptotic accuracy parameter that reflects memory information limitations, an intercept (at which point accuracy first rises above chance), and a rate of rise from chance to asymptote. The dynamics of retrieval is summarized by the intercept and the rate parameter. This results in a description of the level of performance, d' for recognition and $P(c)$ for recall, as a function of total processing time; i.e. delay-of-signal plus response time.

Figures 3 and 4 show performance (observed and predicted) as a function of total processing time for recognition and recall respectively, along with the best fitting exponential functions ($d'(t)$, or $P(c,t)=\lambda(1-\exp[-\beta(t-\delta)]$, for $t-\delta>0$, and 0 elsewhere; in which λ is the asymptote, β the rate, and δ the intercept). It is clear that processes underlying the dynamics of retrieval are quite different: Performance in recognition is

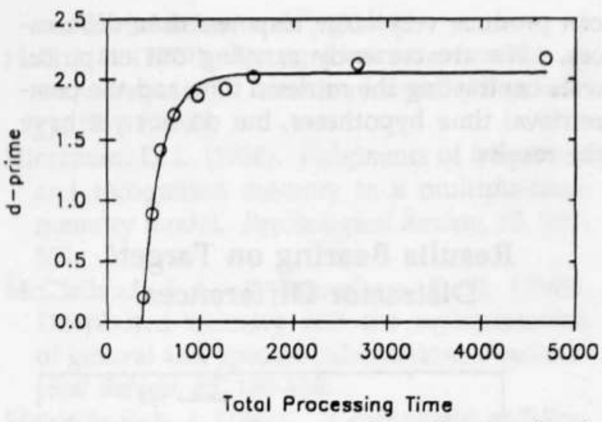

Figure 3. Recognition performance, d', as a function of total processing time in ms (λ=2.06; β=.00503; $1/\beta$=199; δ=373).

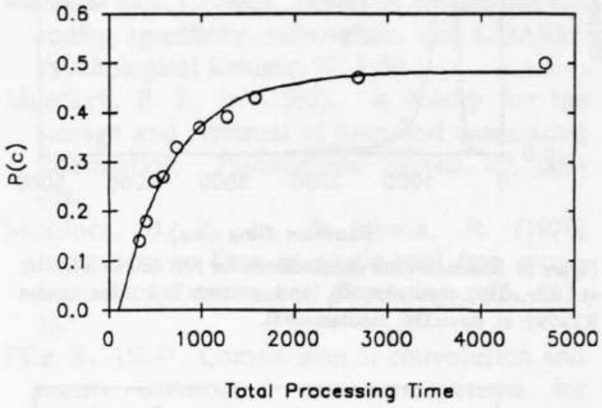

Figure 4. Recall performance, $P(c)$, as a function of total processing time in ms (λ=.48; β=.00161; $1/\beta$=621; δ=46).

characterized by a very fast rate of growth and asymptotic performance is reached fairly quickly, whereas recall performance shows a much more gradual approach to asymptote. These differences are reflected in the parameters of the best fitting functions.

Such results are generally consistent with a two process view of retrieval, such as the SAM model, in which the recall process is spread out in time. The unitary retrieval models would have to posit a difference in post-retrieval mechanisms to explain the recognition-recall differences. For example, in many models noisy information is retrieved from memory (in both recall and recognition). In recall, the process of generating a given word from the noisy information might take a highly variable amount of time, whereas in recognition the time might be relatively fixed (because only a match of the retrieved trace to the test item is needed). In such models it would be necessary to develop a model of post-retrieval response generation that can produce very large response time differences. We are currently carrying out empirical tests contrasting the retrieval time and the post-retrieval time hypotheses, but do not yet have the results.

Results Bearing on Target/ Distractor Differences

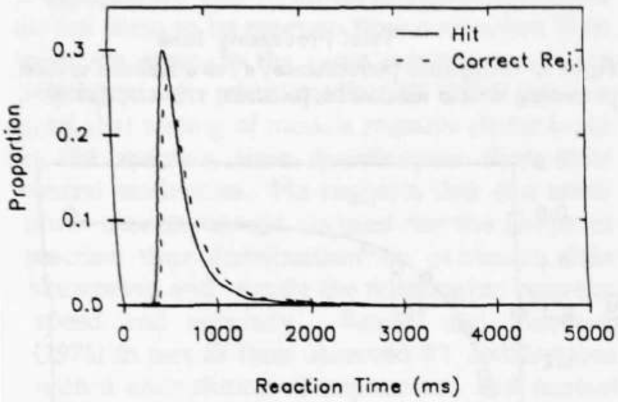

Figure 5. Reaction time distributions for hits (mean RT=710; st. dev.=299; median=630), and correct rejections (mean RT=792; st. dev=334; median=695).

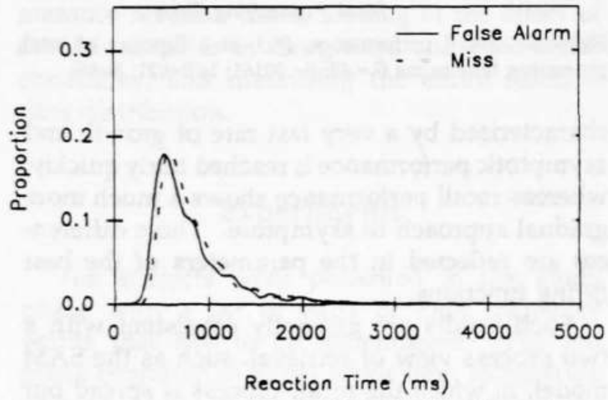

Figure 6. Reaction time distributions for false alarms (mean RT=818; st. dev.=439; median=689) and misses (mean RT=870; st. dev.=458; median=739).

Returning to the free response data, we consider the distributions for positive and negative responses in recognition (Figures 5 and 6). When the responses are correct (hits and correct rejections), the distributions show small differences in both the means and the shape. When these are incorrect (false alarms and misses), the distributions differ slightly in their means, but are identical in shape. Models that use quite different processes for targets and distractors might find such data difficult to predict. On the other hand, careful theoretical work is needed to verify constraints such data place on models. For instance, in the Resonance Model (Ratcliff, 1978) it is assumed that a probe is encoded and then compared in parallel with each item in memory. Each individual comparison is done by a random walk process, and a positive decision is made when any of the parallel comparisons terminates with a match (self-terminating search), and a negative decision is made when all the comparisons terminate with a nonmatch (exhaustive search). With appropriate auxilliary assumptions he was able to show that the model could predict hit and correct rejection distributions that are at least reasonably similar in form.

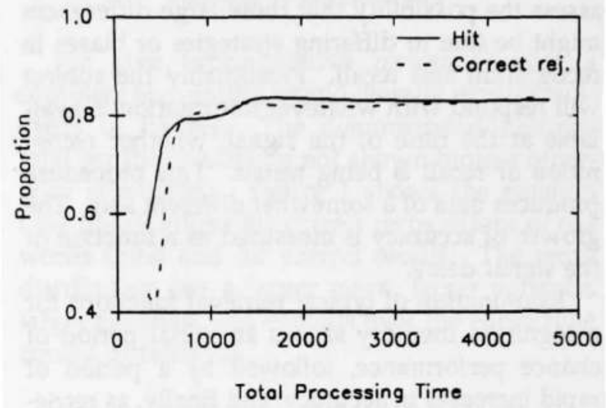

Figure 7. Accuracy growth curves for hits and correct rejections as a function of total processing time (in ms).

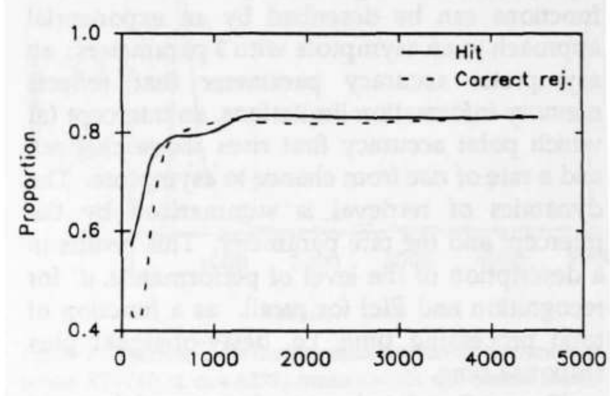

Figure 8. Accuracy growth curves for hits and correct rejections as a function of delay of signal (in ms).

A sharp eye actually reveals that the distributions are not quite identical. The point is revealed more clearly in the signal-to-respond data. Figures 7 and 8 give the accuracy growth curves for hits and correct rejections measured in two ways (in Figure 7, the abscissa includes both the time until the signal and the subsequent time needed to respond). Both methods show that hits start rising sooner than correct rejections, and come together soon thereafter. More research is needed to assess whether this difference is due to a bias to respond 'old' under speed stress, or is due to a real processing difference. Whichever is the case, the remarkable similarities of the target and distractor distributions, and target and distractor signal-to-respond curves, provide strict and informative constraints for models of retrieval.

General Conclusions

We have presented experimental data bearing on the time course of retrieval in both recognition and cued recall, using RT distributions for free response tasks, and accuracy growth curves in signal-to-respond tasks. The large differences between recognition and recall suggest the existence of distinct processes underlying retrieval in the two paradigms. However, we are carrying out further experiments to see whether the differences can be explained in terms of a post-retrieval "clean-up" process in recall (e.g. Metcalfe Eich, 1982). In addition, targets and distractors have nearly identical RT distributions, and fairly similar accuracy growth curves. This suggests a single process for recognition judgments for targets and distractors (such as summation of activation in SAM), and provides general constraints for future model development.

References

Dark, V. J., & Loftus, G. R. (1976). The role of rehearsal in long-term memory performance. *Journal of Verbal Learning and Verbal Behavior, 15*, 479-490.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 541-574.

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

Glenberg, A., & Adams, F. (1978). Type I rehearsal and recognition. *Journal of Verbal Learning and Verbal Behavior, 17*, 455-463.

Hall, J. F. (1979). Recognition as a function of word frequency. *American Journal of Psychology, 92*, 497-505.

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

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Psychological Review, 85*, 159-188.

Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 627-661.

Metcalfe Eich, J. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. Psychological Review, 92, 1-38.

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

Murdock, B. B., Jr., & Okada, R. (1970) Interresponse time in single-trial free recall. *Journal of Experimental Psychology, 86*, 263-267.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91*, 281-294.

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., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190-214.

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.

Tulving, E., & Watkins, M. J. (1973). Continuity between recall and recognition. *American Journal of Psychology, 86*, 739-748.